ТОЧЕЧНАЯ И ИНТЕРВАЛЬНАЯ ОЦЕНКА СЛУЧАЙНОЙ ВЕЛИЧИНЫ

Одной из задач математической статистики, является определение параметров большого массива по исследованию его части.

Два вида оценки:

* ***Точечная оценка случайной величины;***

* ***Интервальная оценка случайной величины.***

ОЦЕНКИ ПАРАМЕТРОВ В ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ

Рассмотрим изучение некоторого количественного признака X. Его распределение в генеральной совокупности характеризуется параметрами, называемыми числовыми характеристиками генеральной совокупности. К ним относятся генеральная средняя, генеральная дисперсия и генеральное среднее квадратическое отклонение.

Определение. *Генеральной средней хг называется среднее арифметическое всех значений изучаемого признака в генеральной совокупности:*

$$\bar{x}_{г} = \frac{1}{N}\sum_{i=1}^{N} x_i ,$$

где N - объем генеральной совокупности; xi - значения признака для различных объектов генеральной совокупности с присвоенными им номерами i = 1, 2, N.

Можно показать, что генеральная средняя равна математическому ожиданию, но для определенности дальше мы будем пользоваться генеральной средней.

Как отмечалось ранее, рассчитать генеральную среднюю практически сложно или невозможно из-за большого объема генеральной совокупности, сложности измерений и проч. Поэтому для изучения генеральной совокупности из нее извлекают выборку относительно небольшого объема.

Определение. *Выборочной средней хв называется среднее арифметическое всех значений изучаемого признака в выборке:*

$$\bar{x}_{\text{в}} = \frac{1}{n}\sum_{i=1}^{n} x_i ,$$

где n - объем выборки; x_i - значения признака. Если некоторые значения x_i встречаются несколько раз (т. е. соответствующие частоты $m_i > 1$), то в сумму их следует подставлять столько же раз.

В математической статистике показывается, что в случае репрезентативной выборки математическое ожидание выборочной средней (для разных выборок) равно генеральной средней:

$$M(\overline{x}_{\text{в}}) = \overline{x}_{\text{г}}.$$

Для оценки генеральной средней, считается, что неизвестная генеральная средняя приблизительно равна рассчитанной выборочной средней:

$$\overline{x}_{\text{г}} \approx \overline{x}_{\text{в}}.$$

Это **<u>точечная оценка</u>** генеральной средней. Точечной она называется потому, что характеризуется одним числом - точкой на числовой оси.

Определение. *Генеральной дисперсией* $\sigma_{\text{г}}^2$ *называется среднее арифметическое квадратов отклонений всех значений изучаемого признака X в генеральной совокупности от генеральной средней:*

$$\sigma_{\text{г}}^2 = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \bar{x}_{\text{г}}\right)^2,$$

где x_i - значения признака; $x_{\text{г}}$ - генеральная средняя.

Генеральная дисперсия характеризует среднее отклонение значений признака в генеральной совокупности относительно генеральной средней.

Рассчитать генеральную дисперсию практически очень сложно. Поэтому для ее оценки вычисляют соответствующие характеристики в выборке.

Определение. *Выборочной дисперсией* $\sigma_{\text{в}}^2$ *называется среднее арифметическое квадратов отклонений всех значений изучаемого признака в выборке относительно выборочной средней:*

$$\sigma_{\text{в}}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}_{\text{в}}\right)^2,$$

где x_i - значение признака; $x_{\text{в}}$ - выборочная средняя. Если некоторые значения x_t встречаются несколько раз (т. е. соответствующие частоты $m_i > 1$), то в сумму их следует подставлять столько же раз.

Выборочная дисперсия характеризует среднее отклонение значений признака в выборке относительно выборочной средней.

Если данные представлены в виде интервального ряда распределения, то выборочную среднюю и выборочную дисперсию можно приближенно вычислить по следующим формулам:

$$\overline{x}_{\text{В}} = \frac{1}{n}\sum_{i=1}^{k} x_i^* m_i ;$$

$$\sigma_{\text{В}}^2 = \frac{1}{n}\sum_{i=1}^{k}\left(x_i^* - \overline{x}_{\text{В}}\right)^2 m_i ,$$

Определение. *Генеральным средним квадратическим отклонением* $\sigma_г$*, называется квадратный корень из генеральной дисперсии:*

$$\sigma_{\text{Г}} = \sqrt{\sigma_{\text{Г}}^2} .$$

Определение. *Выборочным средним квадратическим отклонением* $\sigma_в$ *называется квадратный корень из выборочной дисперсии:*

$$\sigma_{\text{В}} = \sqrt{\sigma_{\text{В}}^2} .$$

ИНТЕРВАЛЬНАЯ ОЦЕНКА. ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ. ДОВЕРИТЕЛЬНАЯ ВЕРОЯТНОСТЬ.

При достаточно большом объеме выборки можно сделать вполне надежные заключения о параметрах генеральной совокупности. Однако на практике часто имеют дело с выборками небольшого объема (n<30). При небольшом объеме выборки пользуются ***интервальными оценками***. В этом случае указывается интервал (*доверительный интервал*).

Доверительный интервал – интервал, в котором с заданной *доверительной вероятностью* находится истинное значение случайной величины (среднее значение генеральной совокупности).

Доверительная вероятность – вероятность, с которой в заданном интервале (доверительном интервале) находится истинное значение случайной величины (среднее значение генеральной совокупности). Обычно в медико-биологических исследованиях доверительную вероятность принимают равной 0,95.

ТАБЛИЦА КОЭФФИЦИЕНТОВ СТЬЮДЕНТА

$f = n - 1$	Значение коэффициента $t_{P,f}$			
	$P = 0.9$	$P = 0.95$	$P = 0.99$	$P = 0.999$
1	6.31	12.71	63.66	636.6
2	2.92	4.30	9.93	31.60
3	2.35	3.18	5.84	12.90
4	2.13	2.78	4.60	8.60
5	2.02	2.57	4.03	6.90
6	1.94	2.45	3.71	5.96
7	1.90	2.37	3.50	5.40
8	1.86	2.31	3.36	5.04
9	1.83	2.26	3.25	4.78
10	1.81	2.23	3.17	4.60
20	1.73	2.09	2.85	3.85
120	1.66	1.98	2.62	3.37
∞	1.65	1.96	2.58	3.29

Доверительный интервал математически записывают так:

$$\bar{x}_в - mt_{\alpha;n} \le x_г \le \bar{x}_в + mt_{\alpha;n}, \text{ или}$$

$$\bar{x}_в \pm mt_{\alpha;n}, \text{ где}$$

$t_{\alpha;n}$ - коэффициент Стьюдента (величина табличная, размерности не имеет),

α - доверительная вероятность,

n – объем выборки;

m – ошибка среднего.

Чтобы определить доверительный интервал, необходимо:

1. Вычислить среднее значение выборки $\bar{x}_в$;
2. Вычислить дисперсию для выборки $\sigma_в^2$;
3. Вычислить исправленную выборочную дисперсию:

$$S^2 = \frac{n}{n-1}\sigma_в^2$$

4. Вычислить ошибку среднего $m = \frac{S}{\sqrt{n}}$.

ЗАДАЧИ

Задача № 1 . При оценке жизненной емкости легких спринтеров был получен следующий вариационный ряд: (Xi, л) - 5, 6, 7, 8, 9; mi – 1, 2, 4, 3, 2. Найдите дисперсию данной случайной величины, если ее выборочное среднее равно 7,25 л.

Решение:

1. Составим статистическое распределение:

X_i	5	6	7	8	9
m_i	1	2	4	3	2

Объем выборки составляет: **n=12**

Выборочное среднее равно: $X_{cp} = 7{,}25$

Выборочная дисперсия рассчитываются по формуле:

$$\overline{D\text{в}} = \frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{x_B})^2 \cdot m_i$$

$\overline{D\text{в}} = \frac{1}{12}[(5\text{-}7{,}25)^2+(6\text{-}7{,}25)^2\cdot 2+(7\text{-}7{,}25)^2\cdot 4+(8\text{-}7{,}25)^2\cdot 3+(9\text{-}7{,}25)^2\cdot 2]=$

$=\frac{1}{12}[5{,}0625+1{,}5625\cdot 2+0{,}0625\cdot 4+0{,}5625\cdot 3+3{,}0625\cdot 2]=\frac{1}{12}[5{,}0625+3{,}125+0{,}25+1{,}6875+6{,}125]=\frac{16{,}25}{12}=1{,}35.$

Задача № 2. При исследования частоты дыхания гимнастов было установлено, что среднее квадратическое отклонение равно 1,1. Найдите выборочную дисперсию данного показателя.

Решение:

$\sigma_B = \sqrt{D_B}$, следовательно $(\sigma_B)^2 = D_B$.

$\sigma_B = 1{,}1$;

$D_B = (1{,}1)^2 = 1{,}21$

Ответ: $D_B = 1{,}21$.

Задача № 3. При исследования двигательных функциий каратистов было установлено, что выборочная дисперсия показателя теппинг-теста равна 1,69. Найдите среднее квадратическое отклонение данного показателя.

Решение:

$$\sigma_B = \sqrt{D_B},$$

$$D_B = 1{,}69;$$

$$\sigma_B = \sqrt{1{,}69} = 1{,}3 \text{ уд/с}$$

Ответ: $\sigma_B = 1{,}3$ уд/с

Задача № 4. При исследовании скорости распространения механической волны на поражённых участках кожи у больных псориазом в регрессирующей стадии были получены следующие результаты, м/с: 38, 39, 41,41, 38, 43, 40, 40, 42, 38, 38, 39, 38, 41, 42, 41, 42, 41, 39, 43, 42, 43, 40, 39, 40, 38, 43, 42, 39, 42.

Дать точечную оценку распределения (рассчитать математическое ожидание, дисперсию, среднее квадратическое отклонение);

Дать интервальную оценку данного показателя на уровне доверительной вероятности $\alpha = 0{,}95$.

Решение:

1. Построение вариационного ряда.

Объем выборки n=30 элементов.

X_i	38	39	40	41	42	43
m_i	6	5	4	5	6	4

Точечная оценка распределения

Выборочная средняя - $\overline{\text{Хв}} = \frac{1}{n}\sum_{i=1}^{k} x_i \cdot m_i$

$\overline{\text{Хв}} = \frac{1}{30}(38\cdot 6+39\cdot 5+40\cdot 4+41\cdot 5+42\cdot 6+43\cdot 4)=40{,}4$ м/с.

Выборочная дисперсия- $\overline{D\text{в}} = \frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{x_B})^2 \cdot m_i$

$\overline{D\text{в}} = \frac{1}{30}[(38-40{,}4)^2\cdot 6+(39-40{,}4)^2\cdot 5+(40-40{,}4)^2\cdot 4+(41-40{,}4)^2\cdot 5+(42-40{,}4)^2\cdot 6+$
$+(43-40{,}4)^2\cdot 4]= 2{,}97$

Выборочное среднее квадратическое отклонение

$\sigma_B = \sqrt{D_B}$; $\sigma_B = \sqrt{2{,}97} = 1{,}72$

Интервальная оценка: P=0,95, $t_{\alpha,n}$=2,04.

Доверительный интервал:

$$\overline{x_B} - mt_{\alpha,n} \leq x_\text{г} \leq \overline{x_B} + mt_{\alpha,n}$$

m- -ошибка среднего:

$m = \frac{\sigma_B}{\sqrt{n}}$; $m = \frac{1{,}72}{\sqrt{30}} = 0{,}31$

Доверительный интервал:

$$40{,}4 - 0{,}6324 \leq x_\text{г} \leq 40{,}4 + 0{,}6324$$

$$39{,}77 \leq x_\text{г} \leq 41{,}03$$

Ответ: $39{,}77 \leq x_\text{г} \leq 41{,}03$

ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

1. Измерения вязкости крови у 10 пациентов с полицитемией дали следующие результаты: (Xi, мПа*с) – 14,9; 15,1; 16,2; 17,3; 18,2; 18,5, mi –1, 2, 3, 2, 1, 1. Найдите выборочную дисперсию данного показателя, если его выборочное среднее равно 16,5 мПа*с. (1,49).

2. При исследования частоты дыхания гимнастов было установлено, что среднее квадратическое отклонение равно 1,5. Найдите выборочную дисперсию данного показателя.

3. При оценке жизненной емкости легких (X,л) спринтеров было установлено, что выборочная дисперсия данного показателя составляет 1,35. Найдите выборочное среднее квадратическое отклонение.

4.При исследовании скорости распространения механической волны в коже щеки после процедуры криомассажа у пациенток с сухим типом кожи были получены следующие результаты, м/с: 38, 58, 46, 39, 49, 62, 62, 49, 43, 44, 68, 41, 54, 64, 64, 38, 58, 46, 39, 49, 62, 62, 49, 43, 44, 68, 41, 54, 64, 64. Дать точечную оценку распределения (рассчитать математическое ожидание, дисперсию, среднее квадратическое отклонение); дать интервальную оценку данного показателя на уровне доверительной вероятности $\alpha = 0,95$.