Лекция 19

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

СТАТИСТИЧЕСКИЕ ОЦЕНКИ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ

Точечные оценки

Любое значение неизвестного параметра, от которого зависит закон распределения случайной величины, вычисленное по опытным данным, всегда является приближенным.

Оценкой параметра и называется в статистике его приближенное случайное значение, вычисленное на основе ограниченного числа опытов. Если оценка параметра характеризуется одним числом, то она называется **точечной**.

Пусть из генеральной совокупности произведена выборка объема n для изучения некоторого признака X. Обозначим неизвестный параметр теоретического распределения интересующего нас признака объектов генеральной совокупности через θ. Требуется по данным выборки найти "подходящую" оценку θ^* для параметра θ.

Очевидно, для некоторой другой выборки оценка θ^* будет принимать иное значение, то есть θ^* - случайная величина, зависящая от данных опытов и их числа n.

Чтобы оценка θ^* давала близкое приближение к оцениваемому параметру, она должна удовлетворять определенным требованиям.

1. При увеличении n оценка θ^* должна сходиться по вероятности к параметру θ, то есть должно выполняться равенство $\lim\limits_{n\to\infty} P\left(\left|\theta^* - \theta\right| < \varepsilon\right) = 1$.

Оценка, обладающая таким свойством, называется **состоятельной**.

2. Необходимо, чтобы пользуясь θ^* вместо θ, мы не допускали систематической (неслучайной) ошибки в сторону занижения или завышения действительного значения оцениваемого параметра, то есть, чтобы $M(\theta^*) = \theta$.

Оценка θ^*, математическое ожидание которой равна оцениваемому параметру, называется **несмещенной.**

3. Оценка θ^* должна обладать по сравнению с другими возможными оценками наименьшей дисперсией: $D(\theta^*) = \min$.

Оценка, обладающая таким свойством, называется **эффективной.**

Ниже рассмотрены повторные и бесповторные выборки и точечные оценки генеральной средней и генеральной дисперсии, удовлетворяющие указанным требованиям.

Оценка генеральной средней повторной выборки

Пусть распределение признака X в генеральной совокупности характе-ризуется таблицей 3. Для оценки неизвестной генеральной средней X производится повторная выборка объема n. При этом каждый отобранный объект вновь возвращается в генеральную совокупность и, следовательно, состав ее восстанавливается. Поэтому результат отбора любого объекта в выборку не будет влиять на результаты следующих отборов, то есть справедлива схема независимых повторяющихся испытаний.

Пусть X_i - случайная величина, значение которой совпадает со значением интересующего нас признака x_i при i-ом наблюдении ($i = 1,2 \ldots n$). Величины X_i можно рассматривать как n независимых и одинаково распределенных случайных величин с параметрами

$$M(X_i) = \bar{x}_0, \quad D(X_i) = \sigma_0^2.$$

Если все X_i различны, то для определения генеральной средней используется в качестве оценки средняя арифметическая наблюдавшихся значений x_i:

$$\bar{x}_в = \frac{1}{n}\sum_{i=1}^{n} x_i ,$$

называемая **выборочной средней**.

Если среди значений x_i $(i=1,2,...,m)$ есть повторяющиеся значения с частотами n_i, причем $\sum_{i=1}^{m} n_i = n$, то

$$\overline{x}_в = \frac{1}{n}\sum_{i=1}^{m} x_i n_i .$$

Выборочная средняя является статистическим аналогом математического ожидания случайной величины в теории вероятностей.

1. Так как величины X_i удовлетворяют условиям теоремы Чебышёва (они независимы, их дисперсии ограничены одной и той же постоянной $C=\sigma_0^2$), то для выборки достаточно большого объема

$$\lim_{n\to\infty} P(|\overline{x}_в - \overline{x}_0| < \varepsilon) = 1,$$

то есть оценка $\overline{x}_в$ является состоятельной. Поэтому для различных выборок достаточно большого объема из одной и той же генеральной совокупности выборочные средние будут практически совпадать между собой. В этом проявляется, так называемое, **свойство устойчивости выборочных средних**.

2. Выборочная средняя $\bar{x}_в$ является несмещенной оценкой $\bar{x}_0$, так как

$$M(\bar{x}_в) = M\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} M(x_i) = \bar{x}_0.$$

Дисперсия $\bar{x}_в$ равна

$$D(\bar{x}_в) = D\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} D(x_i) = \frac{n\sigma_0^2}{n^2} = \frac{\sigma_0^2}{n}.$$

Следовательно, с увеличением числа наблюдений n $D(\bar{x}_в) \to 0$, то есть разброс значений $\bar{x}_в$ относительно $\bar{x}_0$ уменьшается.

3. Можно показать также, что $\bar{x}_в$ является эффективной оценкой, и при увеличении объема выборки закон распределения величины $\bar{x}_в$, как суммы одинаково распределенных независимых случайных величин, приближается к нормальному закону распределения с параметрами

$$M(X) = a = \bar{x}_0, \qquad \sigma = \frac{\sigma_0}{\sqrt{n}}.$$

$$P(|\bar{x}_в - \bar{x}_0| < \varepsilon) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right),$$

где $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{-t^2/2}dt$ - функция Лапласа.

Оценка генеральной средней бесповторной выборки

Пусть, как и выше, Y_i – случайная величина, значение которой совпадает со значением интересующего нас признака y_i при отборе i- го элемента.

Так как выбор каждого отдельного элемента по схеме бесповторной выборки будет влиять на исход последующих выборов, то $Y_1, Y_2, \dots, Y_n$ - зависимые случайные величины. Можно показать, что Y_i распределены по одному и тому же закону с параметрами

$$M(Y_i) = \overline{y}_0, \qquad D(Y_i) = \sigma_0^2.$$

Выборочной средней бесповторной выборки называется средняя арифметическая значений y_i :

$$\overline{y}_в = \frac{1}{n}\sum_{i=1}^{n} y_i,$$

Оценка $\overline{y}_в$ - несмещенная оценка генеральной средней $\overline{y}_0$, так как

$$M(\overline{y}_в) = M\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) = \frac{1}{n}\sum_{i=1}^{n} M(y_i) = \overline{y}_0.$$

Можно показать, что дисперсия выборочной средней равна

$$D(\overline{y}_в) = \frac{\sigma_0^2}{n} \cdot \frac{N-n}{N-1}.$$

Так как величины Y_i зависимые, то условия применимости теоремы Чебышёва к последовательности случайных величин $\{Y_i\}$ не соблюдаются. Применяя поэтому к выборочной средней неравенство Чебышёва, можно записать

$$P(|\overline{y}_в - \overline{y}_0| < \varepsilon) \geq 1 - \frac{D(\overline{y}_в)}{\varepsilon^2} = 1 - \frac{\sigma_0^2}{n\varepsilon^2} \cdot \frac{N-n}{N-1},$$

следовательно $\lim\limits_{n\to\infty} P(|\overline{y}_в - \overline{y}_0| < \varepsilon) = 1$, то есть $\overline{y}_в$ - состоятельная оценка $\overline{y}_0$.

Оценка $\overline{y}_в$ является также эффективной оценкой генеральной средней. С увеличением объема выборки закон распределения $\overline{y}_в$ приближается к нормальному. Поэтому $P(|\overline{y}_в - \overline{y}_0| < \varepsilon) = 2\Phi(\varepsilon / \sigma)$, где

$$\sigma = \sqrt{D(\overline{y}_в)} = \sqrt{\frac{\sigma_0^2}{n} \frac{N-n}{N-1}} = \sqrt{\frac{\sigma_0^2}{n}\left(1 - \frac{n-1}{N-1}\right)}.$$

Так как объем генеральной совокупности N, как правило, весьма большой, то

$$\sigma \approx \sqrt{\frac{\sigma_0^2}{n}\left(1 - \frac{n}{N}\right)}.$$

Определение генеральной дисперсии
Повторная выборка

Если в выражениях для дисперсии случайной величины X

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

заменить математическое ожидание его аналогом в статистике - выборочной средней $\bar{x}_в$, то получим статистический аналог дисперсии:

$$D^*(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_в)^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \bar{x}_в^2$$

Величина $D^*(X)$ называется **выборочной дисперсией**.

Если значения признака $x_1, x_2, \ldots, x_m$ повторяющиеся с частотами соответственно $n_1, n_2, \ldots, n_m$, причем $\sum_{i=1}^{m} n_i = n$, то выборочная дисперсия вычисляется по формуле

$$D^*(X) = \frac{\sum_{i=1}^{m}(x_i - \bar{x}_в)^2 n_i}{n}$$

или

$$D^*(X) = \overline{x^2} - \bar{x}_в^2,$$ где $\overline{x^2} = \frac{\sum x_i^2 n_i}{n}$ - средняя квадратов значений признака.

Оценка $D^*(X)$ является состоятельной. Действительно, первый член в правой части формулы (7.5) – среднее арифметическое n значений x_i^2 и, следовательно, сходится по вероятности к $M(X^2)$, поэтому

$$\lim_{n\to\infty} P(\left|D^*(X) - D(X)\right| < \varepsilon) = 1.$$

Можно показать, что

$$M\left(D^*\right) = \sigma_0^2 \frac{n-1}{n}.$$

Следовательно, D^* - смещенная оценка параметра σ_0^2. Ее использование приводит к систематической ошибке в определении генеральной дисперсии, давая заниженное значение σ_0^2.

Умножая $D^*(X)$ на поправочный множитель $\frac{n}{n-1}$, получим так называемую, **исправленную дисперсию**

$$\overline{D}(X) = \frac{n}{n-1} D^*(X) = \frac{\sum_{i=1}^{n} (x_i - \overline{x}_в)^2}{n-1}$$

или для случая, когда имеются повторяющиеся значения признака $\overline{D}(X) = \dfrac{\sum_{i=1}^{m} (x_i - \overline{x}_в)^2 n_i}{n-1}$.

Очевидно, исправленная дисперсия является несмещенной оценкой дисперсии в генеральной совокупности. На практике ею пользуются, если $n < 30$. При бóльших n, естественно, обе оценки
Оценка $\overline{D}$ в общем случае не является эффективной. Однако для наиболее распространенного на практике нормального закона она оказывается асимптотически эффективной (то есть при больших n отношение ее дисперсии к минимально возможной дисперсии неограниченно приближается к единице).
отличаются друг от друга очень мало.

Бесповторная выборка

Как и для повторной выборки, можно показать, что величина

$$D_1^*(Y) = \frac{\sum_{i=1}^{n}(y_i - \overline{y}_в)^2}{n}$$

является смещенной оценкой генеральной дисперсии при бесповторной выборке:

$$M(D_1^*) = \frac{n-1}{n}\frac{N}{N-1}\sigma_0^2.$$

Несмещенной оценкой генеральной дисперсии при этом будет исправленная дисперсия

$$\overline{D}_1 = \frac{n}{n-1}\cdot\frac{N-1}{N}D_1^*.$$

Замечание. Дисперсии оценок генеральных средних $\bar{x}_0, \bar{y}_0$ зависят от $\bar{x}_0, \bar{y}_0$ (так как σ_0 выражается через $\bar{x}_0$ или $\bar{y}_0$). Но $\bar{x}_0, \bar{y}_0$ являются неизвестными параметрами, иначе бы отпала необходимость в применении выборочного метода. Чтобы преодолеть это противоречие на практике, в формулах для дисперсий величин $\bar{x}_в, \bar{y}_в$ генеральную дисперсию σ_0^2 заменяют выборочной (или исправленной) дисперсией.

<u>Пример.</u> При обработке наружного диаметра 15 карданных валов были получены следующие размеры в мм (см. таблицу 6). Определить несмещенные оценки математического ожидания и среднего квадратического отклонения диаметров, полагая, что обработанные диаметры имеют нормальное распределение.

Результаты вычислений представлены в таблице 6.

Таблица 6

№	x_i [мм]	$\lvert x_i - \overline{x}_в \rvert$	$\lvert x_i - \overline{x}_в \rvert^2$
1	42,22	0,047	0,0022
2	41,87	0,397	0,1576
3	42,56	0,293	0,0858
4	42,03	0,237	0,0562
5	42,48	0,213	0,0454
6	42,31	0,043	0,0018
7	40,15	2,117	4,4817
8	42,82	0,553	0,3058
9	43,83	1,563	2,4430
10	43,40	1,133	1,2837
11	41,13	1,137	1,2928
12	41,72	0,547	0,2992
13	41,35	0,917	0,8409
14	44,13	1,863	3,4708
15	42,00	0,267	0,0712
	$\sum_{i=1}^{15} x_i = 634$		$\sum_{i=1}^{15} (x_i - \overline{x}_в)^2 = 14,8381$

По данным таблицы находим

$$\overline{x}_в = \frac{1}{15} \cdot \sum_{i=1}^{15} x_i = 42{,}27, \quad \overline{D}(X) = \frac{\sum_{i=1}^{15} (x_i - \overline{x}_в)^2}{n-1} = 1{,}06, \quad \overline{\sigma} = \sqrt{\overline{D}(X)} = 1{,}03.$$

Пример. Объем генеральной совокупности $N = 10000$, объем выборки n=1000. В результате измерения интересующего нас признака X получено: $\bar{x}_в = 15{,}5$, $\overline{D}(X) = 3{,}15$. Найти вероятность того, что среднее значение признака X отличается от своей оценки на величину $\varepsilon \leq 0{,}1$, если выборка повторная; бесповторная.

Выборка повторная:

По формуле (7.3) имеем

$$P\left(\left|\bar{x}_в - \bar{x}_0\right| < \varepsilon\right) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right),$$

где $\sigma = \dfrac{\sigma_0}{\sqrt{n}}$.

Заменяя неизвестное σ, его оценкой, получим

$$\sigma_0 \approx \bar{\sigma} = \sqrt{\bar{D}(X)} = \sqrt{3{,}15} = 1{,}775,$$

следовательно,

$$\sigma = \frac{\bar{\sigma}}{\sqrt{n}} = \frac{1{,}775}{\sqrt{1000}} = 0{,}056 \quad и \quad P\left(\left|15{,}5 - \bar{x}_0\right| \le 0{,}1\right) = 2\Phi\left(\frac{0{,}1}{0{,}056}\right) = 0{,}9265.$$

Выборка бесповторная.

При этом

$$\sigma = \sqrt{\frac{\sigma_0^{*}}{n}\left(1 - \frac{n}{N}\right)} \approx \sqrt{\frac{\sigma}{n}\left(1 - \frac{n}{N}\right)} = \sqrt{\frac{3{,}15}{1000}\left(1 - \frac{1000}{10000}\right)} = 0{,}053 \quad и$$

$$P\left(\left|15{,}5 - \bar{x}_0\right| \le 0{,}1\right) = 2\,\Phi\left(\frac{0{,}1}{0{,}053}\right) = 0{,}9412.$$

Метод максимального правдоподобия

Для построения точечных оценок в статистике применяют различные методы: метод максимального правдоподобия, метод моментов, метод наименьших квадратов. Ограничимся первым из них.

Пусть X – случайная величина, которая в результате n независимых опытов приняла значения: $x_1, x_2, ..., x_n$, и пусть закон распределения X известен, но с точностью до некоторого параметра a, от которого он зависит. Требуется найти подходящую точечную оценку $\overline{a}$ параметра a.

Введем обозначение: $P(X = x_i) = P(x_i, a)$ и составим функцию L, равную произведению вероятностей независимых событий $X = x_1, ..., X = x_n$, то есть вероятности их совместного осуществления:

$$L(x_1, x_2, ..., x_n, a) = P(x_1, a) \cdot P(x_2, a) \cdot ... \cdot P(x_n, a).$$

Функция $L(x_1, x_2, ..., x_n, a)$ аргумента a (x_i – фиксированные числа) называется **функцией правдоподобия** дискретной случайной величины X.

Идея метода заключается в том, что в качестве точечной оценки параметра a принимается такое значение $\overline{a}$, при котором функция правдоподобия принимает максимальное значение. Действительно, в экспериментах реализуются обычно именно те значения $x_1, x_2, ..., x_n$ случайной величины X, вероятность которых максимальна.

Оценку $\overline{a}$ называют **оценкой максимального правдоподобия**.

Для отыскания максимума функции правдоподобия применяются обычные правила отыскания экстремума функции:

Решается уравнение $\frac{dL}{da} = 0$ (его называют **уравнением правдоподобия**), затем вычисляется вторая производная $\frac{d^2L}{da^2}$. Если она при $a = \overline{a}$ отрицательна, то $\overline{a}$ - точка максимума. Найденную точку максимума $\overline{a}$ и принимают за оценку наибольшего правдоподобия параметра a.

Замечания.

1. Функции L и lnL достигают максимума при одном и том же значении параметра a Поэтому вместо отыскания максимума функции L часто ищут максимум функции lnL – логарифмической функции правдоподобия, что оказывается удобнее.
2. Для непрерывной случайной величины X функцией правдоподобия называется функция параметра a вида:

$$L(x_1, x_2, ..., x_n, a) = f(x_1, a) \cdot f(x_2, a) \cdot ... \cdot f(x_n, a),$$

где $f(x_i, a)$ - плотность вероятностей.

Оценка максимального правдоподобия неизвестного параметра распределения непрерывной случайной величины строится так же, как и для дискретной случайной величины.

Пример. Найти методом максимального правдоподобия оценки параметров a и σ нормального закона распределения:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}},$$

если значения, принятые случайной величиной X в результате n испытаний равны: $x_1, x_2, ..., x_n$.

Так как нормальный закон распределения характеризуется двумя параметрами $a_1 = a$ и $a_2 = \sigma$, то функция правдоподобия будет функцией двух переменных:

$$L = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_1-a)^2}{2\sigma^2}} \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_2-a)^2}{2\sigma^2}} \cdot ... \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_n-a)^2}{2\sigma^2}} = \frac{1}{\sigma^n(\sqrt{2\pi})^n} e^{-\frac{\sum(x_i-a)^2}{2\sigma^2}}.$$

Логарифмируя это выражение, получим:

$$\ln L = \ln\left[\frac{1}{\sigma^n(\sqrt{2\pi})^n}\right] + \ln e^{-\frac{\sum(x_i-a)^2}{2\sigma^2}} = -n\ln\sigma + \ln\left(\sqrt{2\pi}\right)^{-n} - \frac{\sum(x_i-a)^2}{2\sigma^2}.$$

Частные производные от логарифмической функции правдоподобия по a и σ равны:

$$\frac{\partial \ln L}{\partial a} = -\frac{1}{2\sigma^2}\left[-2(x_1-a) - 2(x_2-a) - ... - 2(x_n-a)\right] = \frac{1}{\sigma^2}\sum(x_i - na),$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{\sum(x_i-a)^2}{\sigma^3}.$$

Поэтому система уравнений правдоподобия примет вид

$$\begin{cases} \dfrac{\partial \ln L}{\partial a} = 0, \\ \dfrac{\partial \ln L}{\partial \sigma} = 0 \end{cases}$$

$$\begin{cases} \frac{1}{\sigma^2}\sum(x_i - na) = 0, \\ -\frac{n}{\sigma} + \frac{\sum(x_i - a)^2}{\sigma^3} = 0. \end{cases}$$

Решая эту систему, получим:

$$a = \frac{\sum x_i}{n}, \qquad \sigma^2 = \frac{\sum(x_i - a)^2}{n}.$$

Следовательно, искомые оценки максимального правдоподобия будут:

$$\overline{a} = \frac{\sum x_i}{n} = \overline{x}_в, \qquad \sigma_* = \sqrt{D^*} = \sqrt{\frac{\sum(x_i - \overline{a})^2}{n}}.$$

Точность оценки. Доверительный интервал и доверительная вероятность

Точечные оценки параметров распределения при выборках малого объема могут существенно отличаться от действительных значений оцениваемых параметров. Поэтому в статистике часто пользуются интервальными оценками (особенно при небольшом числе наблюдений), которые служат для оценки точности и надежности точечных оценок. **Интервальной называется оценка,** которая определяется двумя числами – концами интервала, в котором заключено неизвестное значение параметра.

Пусть для неизвестного параметра θ найдена по данным выборки несмещенная оценка θ^*. Чтобы оценить возможную при этом ошибку, назначим некоторую достаточно большую вероятность γ такую, что любое событие, происходящее с вероятностью γ, можно считать практически достоверным. Найдем далее такое $\varepsilon > 0$, при котором с вероятностью γ можно утверждать, что отклонение θ^* от θ по модулю не будет превосходить ε, то есть

$$P\left(\left|\theta^* - \theta\right| < \varepsilon\right) = \gamma \qquad \text{или} \qquad P\left(\theta^* - \varepsilon < \theta < \theta^* + \varepsilon\right) = \gamma.$$

Величина ε называется **точностью оценки**. Вероятность γ, с которой осуществляется неравенство $\left|\theta^* - \theta\right| < \varepsilon$, называется **доверительной вероятностью** или надежностью оценки. Обычно γ задается равным 0,95; 0,99; 0,999.

Интервал $I_\gamma = \left|\theta^* - \varepsilon,\ \theta^* + \varepsilon\right|$, в котором с надежностью γ заключено неизвестное значение параметра θ, называется **доверительным интервалом.** Так как длина интервала и положение его на оси абсцисс, определяемое центром θ^*, случайны, то говорят, что доверительный интервал I_γ накрывает неизвестный параметр θ с заданной надежностью γ.

Доверительный интервал для оценки генеральной средней при известном среднем квадратическом отклонении

Пусть для случайной величины X с неизвестной генеральной средней $\overline{x}_0$ по данным выборки объема n найдена точечная оценка $\overline{x}_в = \frac{1}{n}\sum_{i=1}^{n} x_i$. Будем предполагать для простоты, что $\sigma(\overline{x}_в)$ известно. Для построения доверительного интервала необходимо найти такое $\varepsilon > 0$, чтобы выполнялось неравенство

$$P(|\overline{x}_в - \overline{x}_0| < \varepsilon) = \gamma .$$

Воспользуемся тем, что $\overline{x}_в$ как сумма независимых одинаково распределенных случайных величин X_i при достаточно большом n (а практически уже при n >10-20) согласно теореме Ляпунова имеет закон распределения, близкий к нормальному. Итак, считая, что X распределена по нормальному закону с параметрами $M(\overline{x}_в) = \overline{x}_0$ (так как $\overline{x}_в$-несмещенная оценка) и $\sigma(\overline{x}_в)$, можно записать

$$P(|\overline{x}_в - \overline{x}_0| < \varepsilon) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right).$$

Учитывая, что вероятность P задана и равна γ, получим

$$P(\overline{x}_в - t\sigma < \overline{x}_0 < \overline{x}_в + t\sigma) = 2\Phi(t) = \gamma,$$

где число $t=\frac{\varepsilon}{\sigma}$ называется **квантилем нормального распределения** и определяется из условия

$$\Phi(t)=\frac{\gamma}{2}.$$

Следовательно, с надежностью γ можно утверждать, что довер и-тельный интервал $I_\gamma=[\overline{x}_в - t\sigma;\ \overline{x}_в + t\sigma]$ накрывает генеральную среднюю $\overline{x}_0$. Точность оценки при этом $\varepsilon = t\sigma$.

Выражение для $\sigma(\overline{x}_в)$ зависит от вида выборки. Так, для повторной выборки

$$\sigma(\overline{x}_в)=\frac{\sigma_0}{\sqrt{n}},$$

для бесповторной выборки

$$\sigma(\overline{x}_в)=\sigma_0\sqrt{\frac{1}{n}\left(1-\frac{n}{N}\right)}.$$

Замечания.

1. По условию генеральная дисперсия σ_0^2 предполагается известной, но если это не так, то дляёиспользуется соответ ствующая точечная оценка.

2. Из формулы $t=\frac{\varepsilon}{\sigma}$ с учетом выражений для σ получаем:

для повторной выборки

$$n = \frac{t^2 \sigma_0^2}{\varepsilon^2}, \tag{7.7}$$

для бесповторной выборки

$$n = \frac{t^2 \sigma_0^2}{\varepsilon^2 + \frac{t^2 \sigma_0^2}{N}}. \tag{7.8}$$

Следовательно, если требуется оценить генеральную среднюю с наперёд заданной точностью ε и надежностью γ, то потребный объем выборки определяется по формулам (7.7), (7.8) соответственно для повторной и бесповторной выборок.

Пример. В условиях примера (стр. 106) построить доверительный интервал для математического ожидания, соответствующий доверительный вероятности γ=0,95 в предположении, что выборка является повто р-ной.

Так как $\overline{x}_в = 42{,}27$ мм, $\overline{\sigma} = 1{,}03$ мм, то заменяя σ_0 его оценкой $\overline{\sigma}$, получим

$$\sigma(\overline{x}_в) = \sqrt{D(\overline{x}_в)} = \sqrt{\frac{\sigma_0^2}{n}} \approx \frac{\overline{\sigma}}{\sqrt{n}} = \frac{1{,}03}{\sqrt{15}} = 0{,}266$$

По таблице функции Лапласа (см. Приложение 2) находим t при заданном γ=0,95:

$$\Phi(t) = \frac{\gamma}{2} = 0{,}475 \quad \text{и} \quad t = 1{,}96.$$

Точность оценки $\varepsilon = t\sigma = 1{,}96 \cdot 0{,}266 = 0{,}52.$

Границы доверительного интервала

$$\overline{x}_в - \varepsilon = 42{,}27 - 0{,}52 = 41{,}75\,мм, \qquad \overline{x}_в + \varepsilon = 42{,}27 + 0{,}52 = 42{,}79\,мм.$$

Следовательно, с надежностью 0,95 можно утверждать, что генеральная средняя $\overline{x}_0$ заключена в пределах $41{,}75 < \overline{x}_0 < 42{,}79.$

Отметим , что повышение надежности оценкиγ приводит к возра с-танию $\Phi(t)$, ε и доверительного интервала, то есть к уменьшению точности определения действительного значения параметра.

Пример. В условиях примера

15,23	15,37	15,48	15,48	15,43	15,35	15,36	15,40	15,45	15,29
15,48	15,58	15,44	15,56	15,28	15,59	15,47	15,41	15,54	15,20
15,38	15,43	15,35	15,56	15,51	15,47	15,40	15,29	15,20	15,46
15,42	15,44	15,41	15,29	15,48	15,39	15,50	15,38	15,45	15,50
15,45	15,42	15,29	15,53	15,34	15,55	15,33	15,32	15,44	15,46
15,32	15,46	15,32	15,48	15,38	15,43	15,51	15,43	15,60	15,44
15,55	15,29	15,31	15,44	15,43	15,44	15,31	15,58	15,28	15,24
15,34	15,49	15,50	15,38	15,48	15,43	15,37	15,29	15,54	15,33
15,36	15,46	15,23	15,44	15,38	15,27	15,52	15,40	15,26	15,37
15,59	15,48	15,46	15,40	15,24	15,41	15,34	15,43	15,38	15,50

найти с надежностью 0,95 точность γ, с которой выборочная средняя оценивает математическое ожидание диаметров изготовляемых валиков и доверительный интервал для математического ожидания диаметров. Предполагается, что диаметры распределены нормально; выборка повторная. Для рассматриваемой выборки выборочная средняя

Для рассматриваемой выборки выборочная средняя

$$\overline{x}_в = \frac{\sum x_i}{100} = 15{,}411 \text{ мм}.$$

$$\overline{\sigma} = \sqrt{\overline{D}} = \sqrt{\frac{\sum (x_i - \overline{x}_в)^2}{99}} = \sqrt{0{,}00907} = 0{,}095.$$

Поэтому

$$\varepsilon = \frac{t\sigma}{\sqrt{n}} \approx \frac{1{,}96 \cdot 0{,}095}{\sqrt{100}} = 0{,}021, \quad \overline{x}_в - \varepsilon = 15{,}392, \qquad \overline{x}_в + \varepsilon = 15{,}430 \quad \text{и}$$

$$15{,}392 < \overline{x}_0 < 15{,}430.$$

Пример. Определить необходимый объем повторной и бесповторной выборок для определения средней продолжительности горения электрических лампочек, чтобы с вероятностью 0,99 предельная ошибка выборки не превышала 50 часов. Объем всей партии лампочек – 5000 шт. Генеральное среднее квадратическое отклонение принять равным 150 часов.

По условию $\varepsilon = 50$, $\sigma = 150$, $N = 5000$, $\gamma = 0{,}99$, следовательно,

$$\Phi(t) = \frac{\gamma}{2} = 0{,}495 \quad \text{и} \quad t = 2{,}58.$$

$$n_{повт.} = \frac{t^2\sigma_o^2}{\varepsilon^2} = 59{,}9; \qquad n_{бесп.} = \frac{t^2\sigma_o^2}{\varepsilon^2 + t^2\sigma_o^2 / N} \approx 59{,}2.$$

Итак, выборки должны содержать не менее 60 лампочек.