Интервальные оценки.

Точечные оценки параметров генеральной совокупности могут быть приняты в качестве ориентировочных, первоначальных результатов обработки выборочных данных. Их недостаток заключается в том, что неизвестно, с какой точностью оценивается параметр. Если для выборок большого объема точность обычно бывает достаточной (при условии несмещенности, эффективности и состоятельности оценок), то для выборок небольшого объема вопрос точности оценок становится очень важным.

Введем понятие интервальной оценки неизвестного параметра генеральной совокупности (или случайной величины ξ, определенной на множестве объектов этой генеральной совокупности). Обозначим этот параметр через Δ. По сделанной выборке по определенным правилам найдем числа Δ_1 и Δ_2, так чтобы выполнялось условие:

$$P(\Delta_1 < \Delta < \Delta_2) = P(\Delta \in (\Delta_1; \Delta_2)) = \gamma$$

Числа Δ_1 и Δ_2 называются **доверительными границами**, интервал (Δ_1, Δ_2) — **доверительным интервалом** для параметра Δ. Число γ называется **доверительной вероятностью** или **надежностью** сделанной оценки.

Сначала задается надежность. Обычно ее выбирают равной 0.95, 0.99 или 0.999. Тогда вероятность того, что интересующий нас параметр попал в интервал (Δ_1, Δ_2) достаточно высока. Число $(\Delta_1 + \Delta_2) / 2$ – середина доверительного интервала – будет давать значение параметра Δ с **точностью** $(\Delta_2 - \Delta_1) / 2$, которая представляет собой половину длины доверительного интервала.

Границы Δ_1 и Δ_2 определяются из выборочных данных и являются функциями от случайных величин $x_1, x_2, \dots, x_n$, а следовательно – сами случайные величины. Отсюда – доверительный интервал (Δ_1, Δ_2) тоже случаен. Он может покрывать параметр Δ или нет. Именно в таком смысле нужно понимать случайное событие, заключающееся в том, что доверительный интервал покрывает число Δ.

Доверительный интервал для математического ожидания нормального распределения при известной дисперсии.

Пусть случайная величина ξ (можно говорить о генеральной совокупности) распределена по нормальному закону, для которого известна дисперсия $D\xi = \sigma^2$ ($\sigma > 0$). Из генеральной совокупности (на множестве объектов которой определена случайная величина) делается выборка объема n. Выборка $x_1, x_2, ..., x_n$ рассматривается как совокупность n независимых случайных величин, распределенных так же как ξ (подход, которому дано объяснение выше по тексту).

Ранее также обсуждались и доказаны следующие равенства:

$Mx_1 = Mx_2 = ... = Mx_n = M\xi$;

$Dx_1 = Dx_2 = ... = Dx_n = D\xi$;

$M\bar{x} = M\xi$;

$D\bar{x} = D\xi / n$;

Достаточно просто доказать (мы доказательство опускаем), что случайная величина $\bar{x}$ в данном случае также распределена по нормальному закону.

Обозначим неизвестную величину $M\xi$ через a и подберем по заданной надежности γ число $d > 0$ так, чтобы выполнялось условие:

$$P(|\bar{x} - a| < d) = \gamma \tag{1}$$

Так как случайная величина $\bar{x}$ распределена по нормальному закону с математическим ожиданием $M\bar{x} = M\xi = a$ и дисперсией $D\bar{x} = D\xi / n = \sigma^2/n$, получаем:

$$P(|\bar{x} - a| < d) = P(a - d < \bar{x} < a + d) =$$

$$= \Phi\left(\frac{a + d - a}{\frac{\sigma}{\sqrt{n}}}\right) - \Phi\left(\frac{a - d - a}{\frac{\square}{\sqrt{n}}}\right) = 2\Phi\left(\frac{d\sqrt{n}}{\sigma}\right)$$

Осталось подобрать d таким, чтобы выполнялось равенство $2\Phi\left(\frac{d\sqrt{n}}{\sigma}\right)=\gamma$ или $\Phi\left(\frac{d\sqrt{n}}{\sigma}\right)=\frac{\gamma}{2}$.

Для любого $\gamma\in[0;1]$ можно по таблице найти такое число t, что $\Phi(t)=\gamma/2$. Это число t иногда называют **квантилем**.

Теперь из равенства

$$\frac{d\sqrt{n}}{\sigma}=t$$

определим значение d: $d=\frac{\sigma t}{\sqrt{n}}$.

Окончательный результат получим, представив формулу (1) в виде:

$$P\left(\bar{x}-\frac{\sigma t}{\sqrt{n}}<a<\bar{x}+\frac{\sigma t}{\sqrt{n}}\right)=\gamma\ .$$

Смысл последней формулы состоит в следующем: с надежностью γ доверительный интервал

$$\left(\bar{x}-\frac{\sigma t}{\sqrt{n}};\bar{x}+\frac{\sigma t}{\sqrt{n}}\right)$$

покрывает неизвестный параметр $a = M\xi$ генеральной совокупности. Можно сказать иначе: точечная оценка $\bar{x}$ определяет значение параметра $M\xi$ с точностью $d=\sigma t/\sqrt{n}$ и надежностью γ.

Задача. Пусть имеется генеральная совокупность с некоторой характеристикой, распределенной по нормальному закону с дисперсией, равной 6,25. Произведена выборка объема $n = 27$ и получено средневыборочное значение характеристики $\bar{x}= 12$. Найти доверительный интервал, покрывающий неизвестное математическое ожидание исследуемой характеристики генеральной совокупности с надежностью γ=0,99.

Решение. Сначала по таблице для функции Лапласа найдем значение t из равенства $\Phi(t) = \gamma / 2 = 0{,}495$. По полученному значению $t = 2{,}58$ определим точность оценки (или половину длины доверительного интервала) d: d = 2,5×2,58 / $\sqrt{27} \approx$ 1,24. Отсюда получаем искомый доверительный интервал: (10,76; 13,24).

Доверительный интервал для математического ожидания нормального распределения при неизвестной дисперсии.

Пусть ξ – случайная величина, распределенная по нормальному закону с неизвестным математическим ожиданием $M\xi$, которое обозначим буквой a . Произведем выборку объема n. Определим среднюю выборочную $\bar{x}$ и исправленную выборочную дисперсию s^2 по известным формулам.

Случайная величина

$$t = \frac{(\bar{x} - a)\sqrt{n}}{s}$$

распределена по закону Стьюдента с n – 1 степенями свободы.

Задача заключается в том, чтобы по заданной надежности γ и по числу степеней свободы $n - 1$ найти такое число t_γ , чтобы выполнялось равенство

$$P\left(\left|\frac{(\bar{x} - a)\sqrt{n}}{s}\right| < t_\gamma\right) = \gamma \qquad (2)$$

или эквивалентное равенство

$$P\left(\bar{x} - t_\gamma \frac{s}{\sqrt{n}} < a < \bar{x} + t_\gamma \frac{s}{\sqrt{n}}\right) = \gamma. \qquad (3)$$

Здесь в скобках написано условие того, что значение неизвестного параметра a принадлежит некоторому промежутку, который и является доверительным интервалом. Его границы зависят от надежности γ , а также от параметров выборки $\bar{x}$ и s.

Чтобы определить значение t_γ по величине γ, равенство (2) преобразуем к виду:

$$P\left(\left|\frac{(\bar{x}-a)\sqrt{n}}{s}\right| \geq t_\gamma\right) = 1-\gamma$$

Теперь по таблице для случайной величины *t*, распределенной по закону Стьюдента, по вероятности $1 - \gamma$ и числу степеней свободы $n - 1$ находим t_γ. Формула (3) дает ответ поставленной задачи.

Задача. На контрольных испытаниях 20-ти электроламп средняя продолжительность их работы оказалась равной 2000 часов при среднем квадратическом отклонении (рассчитанном как корень квадратный из исправленной выборочной дисперсии), равном 11-ти часам. Известно, что продолжительность работы лампы является нормально распределенной случайной величиной. Определить с надежностью 0,95 доверительный интервал для математического ожидания этой случайной величины.

Решение. Величина $1 - \gamma$ в данном случае равна 0,05. По таблице распределения Стьюдента, при числе степеней свободы, равном 19, находим: $t_\gamma = 2{,}093$. Вычислим теперь точность оценки: $2{,}093 \times 121/\sqrt{20}$ = 56,6. Отсюда получаем искомый доверительный интервал:

(1943,4; 2056,6).

Доверительный интервал для дисперсии нормального распределения.

Пусть случайная величина ξ распределена по нормальному закону, для которого дисперсия $D\xi$ неизвестна. Делается выборка объема n . Из нее определяется исправленная выборочная дисперсия s^2. Случайная величина

$$\chi^2 = \frac{(n-1)s^2}{D\xi}$$

распределена по закону χ^2 с $n - 1$ степенями свободы. По заданной надежности γ можно найти сколько угодно границ ${\chi_1}^2$ и ${\chi_2}^2$ интервалов, таких, что

$$P\left({\chi_1}^2 < \chi^2 < {\chi_2}^2\right) = \gamma \qquad (*)$$

Найдем $\chi_1{}^2$ и $\chi_2{}^2$ из следующих условий:

$$P(\chi^2 \le \chi_1{}^2) = (1-\gamma)/\,2 \qquad (**)$$

$$P(\chi^2 \ge \chi_2{}^2) = (1-\gamma)/\,2 \qquad (***)$$

Очевидно, что при выполнении двух последних условий справедливо равенство (*).

В таблицах для случайной величины χ^2 обычно дается решение уравнения $P(\chi^2 \ge \chi_q{}^2) = q$. Из такой таблицы по заданной величине q и по числу степеней свободы $n - 1$ можно определить значение $\chi_q{}^2$. Таким образом, сразу находится значение $\chi_2{}^2$ в формуле (***).

Для определения $\chi_1{}^2$ преобразуем (**):

$$P(\chi^2 \ge \chi_1{}^2) = 1 - (1-\gamma)/\,2 = (1+\gamma)/\,2$$

Полученное равенство позволяет определить по таблице значение $\chi_1{}^2$.

Теперь, когда найдены значения $\chi_1{}^2$ и $\chi_2{}^2$, представим равенство (*) в виде

$$P\left(\chi_1{}^2 < \frac{(n-1)s^2}{D\xi} < \chi_2{}^2\right) = \gamma\,.$$

Последнее равенство перепишем в такой форме, чтобы были определены границы доверительного интервала для неизвестной величины $D\xi$:

$$P\left(\frac{(n-1)s^2}{\chi_2{}^2} < D\xi < \frac{(n-1)s^2}{\chi_1{}^2}\right) = \gamma\,.$$

Отсюда легко получить формулу, по которой находится доверительный интервал для стандартного отклонения:

$$P\left(\frac{\sqrt{(n-1)}s}{\sqrt{\chi_2{}^2}} < \sqrt{D\xi} < \frac{\sqrt{(n-1)}s}{\sqrt{\chi_1{}^2}}\right) = \gamma \qquad (****)$$

Задача. Будем считать, что шум в кабинах вертолетов одного и того же типа при работающих в определенном режиме двигателях — случайная величина, распределенная по нормальному закону. Было случайным образом выбрано 20 вертолетов, и произведены замеры уровня шума (в децибелах) в каждом из них. Исправленная выборочная дисперсия измерений оказалась равной 22,5. Найти доверительный интервал, накрывающий неизвестное стандартное отклонение величины шума в кабинах вертолетов данного типа с надежностью 98%.

Решение. По числу степеней свободы, равному 19, и по вероятности (1 – 0,98)/2 = 0,01 находим из таблицы распределения χ^2 величину $\chi_2^2 = 36{,}2$. Аналогичным образом при вероятности (1 + 0,98)/2 = 0,99 получаем $\chi_1^2 = 7{,}63$. Используя формулу (****), получаем искомый доверительный интервал: (3,44; 7,49).