



СТАТИСТИЧЕСКИЕ ОЦЕНКИ

Тема 9



План лекции:

- 9.1. Статистическая оценка
- 9.2. Точечные оценки
- 9.3. Интервальные оценки
- 9.4. Доверительный интервал для оценки математического ожидания нормального распределения при известной дисперсии
- 9.5. Доверительный интервал для оценки для оценки математического ожидания при неизвестной дисперсии
- 9.6. Доверительный интервал для оценки среднего квадратического отклонения нормального распределения математического ожидания при неизвестной дисперсии

9.1. Статистическая оценка

Одной из центральных задач математической статистики является задача оценивания теоретического распределения случайной величины на основе выборочных данных.

При этом часто предполагается, что вид закона распределения генеральной совокупности известен, но неизвестны параметры этого распределения, такие как математическое ожидание, дисперсия. Требуется найти приближенные значения этих параметров, то есть получить статистические оценки указанных параметров.



Статистической оценкой $\bar{\theta}$ параметра θ теоретического распределения называют его приближенное значение, зависящее от данных выбора.

Рассматривая выборочные значения $x_1, x_2, \dots, x_n$ как реализации случайных величин $X_1, X_2, \dots, X_n$, получивших конкретные значения в результате опытов, можно представить оценку $\bar{\theta}$ как функцию этих случайных величин: $\bar{\theta} = \varphi(X_1, X_2, \dots, X_n)$. Это означает, что оценка тоже является случайной величиной.



Если для оценки θ взять несколько (k) выборок, то получим столько же случайных оценок $\bar{\theta}_1, \bar{\theta}_2, \dots, \bar{\theta}_k$.

Если число наблюдений невелико, то замена неизвестного параметра θ оценкой $\bar{\theta}$ приводит к ошибке, которая тем больше, чем меньше число опытов.

9.2. Точечные оценки

Статистические оценки могут быть *точечными* и *интервальными*.

Точечные оценки представляют собой число или точку на числовой оси. Чтобы оценка $\bar{\theta}$ была близка к значению параметра θ , она должна обладать свойствами состоятельности, несмещенности и эффективности.

Оценка $\bar{\theta}$ параметра θ называется *состоятельной*, если она сходится по вероятности к оцениваемому параметру, то есть для любого $\varepsilon > 0$:

$$\lim_{n \rightarrow \infty} P(|\bar{\theta} - \theta| < \varepsilon) = 1.$$

Свойство состоятельности нужно проверять в первую очередь. Оно *обязательно* для любого правила оценивания. Несостоятельные оценки не используются

Оценка $\bar{\theta}$ параметра θ называется *несмещенной*, если $M(\bar{\theta}) = \theta$, то есть математическое ожидание оценки равно оцениваемому параметру. Если $M(\bar{\theta}) \neq \theta$, то оценка $\bar{\theta}$ называется *смещенной*.

Это свойство оценки желательно, но не обязательно. Часто полученная оценка бывает смещенной, но ее можно поправить так, чтобы она стала несмещенной.

Требования несмещенности особенно важно при малом числе опытов.

Несмещенная оценка $\bar{\theta}$ параметра θ называется *эффективной*, если она среди всех несмещенных оценок, в определенном классе оценок данного параметра, обладает наименьшей дисперсией.

Можно показать, что:

- $\bar{x}_e$ является состоятельной, несмещенной и эффективной оценкой $M(X)$ в классе линейных оценок;
- D_e является состоятельной, смещенной оценкой $D(X)$;

- $S^2 = \frac{n}{n-1} D_e$ является состоятельной, несмещенной оценкой $D(X)$; (при больших n разница между S^2 и D_e мала. S^2 используется при малых выборках, обычно при $n \leq 30$);

- относительная частота $\frac{n_A}{n}$ появления события

A в n независимых испытаниях является состоятельной, несмещенной и эффективной оценкой, в классе линейных оценок, неизвестной вероятности $p = P(A)$ (p - вероятность появления события A в каждом испытании);



- эмпирическая функция распределения выборки $F^*(x)$ является состоятельной, несмещенной оценкой функции распределения $F(x)$ случайной величины X .

Для нахождения оценок неизвестных параметров используют различные методы. Наиболее распространенными являются: метод моментов, метод максимального правдоподобия (ММП), метод наименьших квадратов (МНК).

9.3. Интервальные оценки

При выборке малого объема точечная оценка может существенно отличаться от оцениваемого параметра. В этом случае целесообразно использовать интервальные оценки.

Интервальной называют оценку, которая определяется двумя числами – концами интервала.

Пусть найденная по данным выборки величина $\bar{\theta}$ служит оценкой неизвестного параметра θ . Оценка $\bar{\theta}$ определяет θ тем точнее, чем меньше $|\theta - \bar{\theta}|$, то есть чем меньше δ в неравенстве $|\theta - \bar{\theta}| < \delta$ ($\delta > 0$).

Поскольку $\bar{\theta}$ - случайная величина, то и разность $|\theta - \bar{\theta}|$ - случайная величина. Поэтому неравенство $|\theta - \bar{\theta}| < \delta$, при заданном δ может выполняться только с некоторой вероятностью.

Доверительной вероятностью (надежностью) оценки $\bar{\theta}$ параметра θ называется вероятность γ , с которой выполняется неравенство $|\theta - \bar{\theta}| < \delta$.

Обычно задается надежность γ и определяется δ .
Чаще всего надежность задается значениями от 0,95 и выше, в зависимости от конкретно решаемой задачи.

Неравенство $|\theta - \bar{\theta}| < \delta$ можно записать
 $\bar{\theta} - \delta < \theta < \bar{\theta} + \delta$.

Доверительным интервалом называется интервал $(\bar{\theta} - \delta; \bar{\theta} + \delta)$, который покрывает неизвестный параметр с заданной надежностью γ .

9.4. Доверительный интервал для оценки математического ожидания нормального распределения при известной дисперсии

Пусть случайная величина X имеет нормальное распределение. Известно значение σ и задана доверительная вероятность (надежность) γ . Требуется построить доверительный интервал для параметра a по выборочному среднему $\overline{X_e}$, $\gamma = 2\Phi(t)$

Доверительный интервал

$$\left(\overline{X_e} - t \cdot \frac{\sigma}{\sqrt{n}}; \overline{X_e} + t \cdot \frac{\sigma}{\sqrt{n}} \right) \text{ покрывает с}$$

надежностью γ математическое ожидание a .

Пример. Случайная величина имеет нормальное распределение с известным средним квадратическим отклонением $\sigma=3$. Найти доверительный интервал оценки неизвестного математического ожидания по выборочной средней $\bar{x}_e$, если объем выборки $n=36$, а надежность оценки $\gamma=0,95$.

1. Находим t : $2\Phi(t) = 0,95 \quad \Phi(t) = 0,475$

По таблице значений функции Лапласа $t = 1,96$.

2. Определяем $t \cdot \frac{\sigma}{\sqrt{n}} = 1,96 \cdot \frac{3}{\sqrt{36}} = 0,98$.

Доверительный интервал запишется в виде:
 $(\bar{x}_e - 0,98; \bar{x}_e + 0,98)$.

9.5. Доверительный интервал для оценки математического ожидания при неизвестной дисперсии

Пусть случайная величина X имеет нормальное распределение: $N(a; \sigma)$, причем σ - неизвестно, γ - задана. Если $D(X)$ неизвестна, то пользуются оценкой S^2 . Введем случайную величину $T = \frac{\overline{X}_e - a}{\frac{S}{\sqrt{n}}}$,

Случайная величина T имеет распределение Стьюдента с $(n - 1)$ степенью свободы.

Тогда доверительный интервал для оценки $a = M(X)$ имеет вид:

$$\left(\overline{X}_e - t_j \cdot \frac{S}{\sqrt{n}}; \overline{X}_e + t_j \cdot \frac{S}{\sqrt{n}} \right),$$

где $\overline{X}_e$ - выборочное среднее;

S - исправленное среднее квадратическое отклонение;

t_j - находим по таблице квантилей распределения Стьюдента зависимости от числа степеней свободы и доверительной вероятности γ .

Пример. Произведено пять независимых наблюдений над случайной величиной $X \sim N(a; \sigma)$. Результаты наблюдений таковы:

$$x_1 = -35, \quad x_2 = 20, \quad x_3 = 15, \quad x_4 = -12, \quad x_5 = 42.$$

Построить для неизвестного $M(x) = a$ доверительный интервал, если $\gamma = 0,95$.

1. Находим $\overline{x}_e$:
$$\overline{x}_e = \frac{1}{5}(-35 + 20 + 15 - 12 + 42) = \frac{1}{5} \cdot 30 = 6$$

2. Находим S^2 :
$$S^2 = \frac{1}{4}((-35 - 6)^2 + (20 - 6)^2 + (15 - 6)^2 + (-12 - 6)^2 + (42 - 6)^2) =$$
$$= \frac{1}{4}3638 = 909,5$$

3. По таблице квантилей распределения Стьюдента для $\gamma = 0,95$ и $n - 1 = 4$ находим t_j .

Доверительный интервал:

$$\left(6 - 2,78 \frac{30,2}{2,24}; 6 + 2,78 \frac{30,2}{2,24} \right) \quad \text{или} \quad | (-31,5; 43,5).$$

9.6. Доверительный интервал для оценки среднего квадратического отклонения нормального распределения

Если $M(X) = a$ неизвестно, то доверительный интервал для оценки $\sigma(X)$ имеет вид:

$$\left(\frac{\sqrt{n-1} \cdot S}{\chi_2}, \frac{\sqrt{n-1} \cdot S}{\chi_1} \right)$$

где n - объем выборки; S - исправленное среднее квадратическое отклонение:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_e)^2,$$

$$\chi_1^2 = \chi_{\frac{1+\gamma}{2}; n-1}^2, \quad \chi_2^2 = \chi_{\frac{1-\gamma}{2}; n-1}^2 \quad - \quad \text{квантили } \chi^2 - \text{распределения,}$$

определяемые по таблице $\chi_{\alpha, k}^2$ при $k = n-1$ и $\alpha = \frac{1+\gamma}{2}$, $\alpha = \frac{1-\gamma}{2}$.

Пример. Для оценки параметра $\sigma(X)$ нормально распределенной случайной величины была сделана выборка объема в 25 единиц и вычислено $S = 0,8$. Найти доверительный интервал, покрывающий σ с вероятностью $\gamma = 0,95$. Имеем $n = 25$, $\gamma = 0,95$.

$$\chi_1^2 = \chi_{\frac{1+0,95}{2}; 25-1}^2 = \chi^2(0,975; 24) = 12,4$$

$$\chi_2^2 = \chi_{\frac{1-0,95}{2}; -1}^2 = \chi^2(0,025; 24) = 39,4$$

Доверительный интервал имеет вид:

$$\left(\frac{\sqrt{24} \cdot 0,8}{\sqrt{39,4}}; \frac{\sqrt{24} \cdot 0,8}{\sqrt{12,4}} \right) \quad \text{или} \quad (0,79; 1,4) .$$