

Оценки параметров распределения

Курс лекций «Метрология»

Математическая статистика

Задачи математической статистики:

- 1) определение способов сбора статистических данных,
- 2) разработка методов анализа статистических данных:
 - а) расчет оценок (оценка вероятности, оценки параметров известного распределения и др.),
 - б) проверка статистических гипотез.

Выборочный метод

Пусть требуется изучить совокупность однородных объектов относительно некоторого качественного или количественного признака.

1) *Сплошное обследование всех элементов совокупности*

Может оказаться невозможным из-за большого или бесконечного объема исследуемой совокупности, физического уничтожения объектов при изучении, экономических причин.

2) *Выборочный метод*

Изучение ограниченного числа объектов, случайно отобранных из совокупности.

Выборочный метод

Генеральная совокупность – совокупность объектов, из которой производится отбор.

Выборочная совокупность (выборка) - совокупность случайно отобранных объектов.

Объем совокупности (генеральной или выборочной) – количество объектов в этой совокупности.

Выборка должна быть **репрезентативной** (представительной), т. е. должна хорошо отражать пропорции генеральной совокупности.

Для этого выборка должна быть **случайной**: все объекты генеральной совокупности должны иметь одинаковую вероятность попадания в выборку.

Оценки параметров распределения

Задача: нахождение оценок параметров распределения случайной величины на основании выборки.

Случайная величина X представляет собой генеральную совокупность бесконечного объема.

Выборка $x_1 \dots x_n$ образована n значениями случайной величины, полученными в результате независимых наблюдений.

Распределение случайной величины X (генеральная совокупность) описывается параметром a .

Путем обработки n значений случайной величины X (выборка) можно получить **оценку** данного параметра.

Оценки параметров распределения

Виды оценок:

- 1) **Точечная оценка** \tilde{a} параметра a определяется одним числом, наиболее близким к параметру a .
- 2) **Интервальная оценка** $[\tilde{a}_1; \tilde{a}_2]$ параметра a определяется в виде доверительного интервала, задаваемого своими границами \tilde{a}_1 и \tilde{a}_2 , между которыми с заданной доверительной вероятностью p находится параметр a .

Оценки вычисляются на основании выборки $x_1 \dots x_n$.

Точечные оценки

Точечная оценка \tilde{a} параметра a представляет собой случайную величину на множестве выборок из одной и той же генеральной совокупности.

Требования к точечным оценкам:

- 1) **Несмещенность.** Несмещенная оценка - оценка \tilde{a} , математическое ожидание которой равно значению оцениваемого параметра a при любом объеме выборки.
- 2) **Эффективность.** Эффективная оценка - оценка \tilde{a} , которая при заданном объеме выборки имеет наименьшую возможную дисперсию.
- 3) **Состоятельность.** Состоятельная оценка - оценка \tilde{a} , которая при объеме выборки $n \rightarrow \infty$ стремится по вероятности к оцениваемому параметру. Используется для выборок большого объема.

Точечная оценка математического ожидания

Точечной оценкой математического ожидания $M(X)$ случайной величины X является ***выборочное среднее***:

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

Выборочное среднее является несмещенной, эффективной и состоятельной оценкой.

Точечная оценка дисперсии

Точечной оценкой дисперсии $D(X)$ случайной величины X является **выборочная дисперсия**:

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

Выборочная дисперсия является состоятельной, но смещенной оценкой.

$$M(\tilde{D}) = \frac{n-1}{n} D(X)$$

Точечная оценка дисперсии

Точечной оценкой дисперсии $D(X)$ случайной величины X является ***исправленная дисперсия***:

$$S_X^2 = \frac{n}{n-1} \tilde{D}$$
$$S_X^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Исправленная дисперсия является состоятельной и несмещенной оценкой.

Точечная оценка среднего квадратического отклонения

Точечной оценкой среднего квадратического отклонения $\sigma(X)$ случайной величины X является **исправленное среднее квадратическое отклонение**:

$$S_X = +\sqrt{S_X^2} = +\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Исправленное среднее квадратическое отклонение является состоятельной, но смещенной оценкой.

Точечная оценка среднего квадратического отклонения

Несмещенной точечной оценкой среднего квадратического отклонения $\sigma(X)$ случайной величины X является **исправленное среднее квадратическое отклонение**, умноженное на поправочный коэффициент $k(n)$:

$$S_X = +k(n) \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Коэффициент $k(n)$ зависит от объема выборки и изменяется: от $k(n) = 1,13$ при $n = 3$ до $k(n) \approx 1,03$ при $n \rightarrow \infty$.

Данная оценка является состоятельной и несмещенной.

Точечная оценка ско среднего арифметического

Точечная оценка среднего квадратического отклонения среднего арифметического одинаково распределенных независимых случайных величин $X_1 \dots X_n$ связана с оценкой среднего квадратического отклонения S_X отдельной случайной величины X_i ($i = 1 \dots n$) соотношением:

$$S_{\bar{X}} = \frac{S_X}{\sqrt{n}}$$

Интервальные оценки

Интервальная оценка параметра a задается в виде доверительного интервала $[\tilde{a}_1; \tilde{a}_2]$ и доверительной вероятности p :

$$P(\tilde{a}_1 \leq a \leq \tilde{a}_2) = p$$

Положение границ интервала \tilde{a}_1 и \tilde{a}_2 зависит от доверительной вероятности p , значение которой выбирают близким к единице: 0,95; 0,99.

$q = 1 - p$ — уровень значимости.

Интервал для значений случайной величины

При нормальном распределении случайной величины X с математическим ожиданием $M(X)$ и ско $\sigma(X)$ с доверительной вероятностью p значение случайной величины X принадлежит интервалу:

$$X \in \left[\bar{x} - u_{\frac{1+p}{2}} \cdot S_X ; \bar{x} + u_{\frac{1+p}{2}} \cdot S_X \right]$$

\bar{x} – оценка математического ожидания $M(X)$;

S_X – оценка ско $\sigma(X)$;

$u_{\frac{1+p}{2}}$ – квантиль стандартного нормального распределения порядка $(1+p)/2$.

$$X = \bar{x} \pm u_{\frac{1+p}{2}} \cdot S_X, \quad p = \dots$$

Доверительный интервал для математического ожидания

При нормальном распределении случайной величины X с математическим ожиданием $M(X)$ и неизвестным ско $\sigma(X)$ с доверительной вероятностью p математическое ожидание $M(X)$ принадлежит интервалу:

$$M(X) \in \left[\bar{x} - t_{\frac{1+p}{2}}(n-1) \cdot S_{\bar{X}}; \bar{x} + t_{\frac{1+p}{2}}(n-1) \cdot S_{\bar{X}} \right]$$

\bar{x} – оценка математического ожидания $M(X)$;

$S_{\bar{X}} = S_X / \sqrt{n}$ – оценка ско выборочного среднего;

$t_{\frac{1+p}{2}}(n-1)$ – квантиль распределения Стьюдента порядка $(1+p)/2$ с $(n-1)$ степенями свободы.

$$M(X) = \bar{x} \pm t_{\frac{1+p}{2}}(n-1) \cdot S_{\bar{X}}, \quad p = \dots$$

Распределение Стьюдента

Случайная величина T имеет распределение Стьюдента с k ($k > 0$) степенями свободы, если

$$T = \frac{Z}{\sqrt{Y/k}}$$

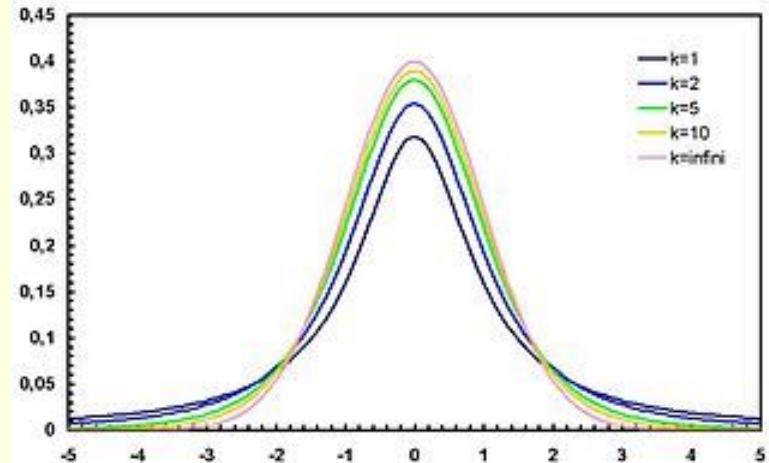
Z – случайная величина, имеющая стандартное нормальное распределение: $M(Z) = 0$, $\sigma(Z) = 1$;

Y – случайная величина, имеющая распределение χ^2 с k степенями свободы;

Z и Y – независимые случайные величины.

$$f(t) = \frac{1}{\sqrt{\pi \cdot k}} \cdot \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma(k/2)} \cdot \left(1 + \frac{t^2}{k}\right)^{-\frac{k+1}{2}}$$

С возрастанием числа степеней свободы k распределение Стьюдента быстро приближается к нормальному распределению.



Доверительный интервал для среднего квадратического отклонения

При нормальном распределении случайной величины X с $\text{ско } \sigma(X)$ с доверительной вероятностью p $\text{ско } \sigma(X)$ принадлежит интервалу:

$$\sigma(X) \in \left[S_X \cdot \sqrt{\frac{n-1}{\chi_{\frac{1+p}{2}}^2(n-1)}}; S_X \cdot \sqrt{\frac{n-1}{\chi_{\frac{1-p}{2}}^2(n-1)}} \right]$$

S_X – оценка $\text{ско } \sigma(X)$;

$\chi_{\frac{1-p}{2}}^2(n-1)$, $\chi_{\frac{1+p}{2}}^2(n-1)$ – квантили распределения χ^2 порядка $(1-p)/2$ и $(1+p)/2$ с $(n-1)$ степенями свободы.

Распределение χ^2

Случайная величина Y имеет распределение χ^2 с k ($k > 0$) степенями свободы, если

$$Y = \sum_{i=1}^k X_i^2$$

$X_1 \dots X_k$ – независимые случайные величины, имеющие стандартное нормальное распределение: $M(X_1) = \dots = M(X_k) = 0$, $\sigma(X_1) = \dots = \sigma(X_k) = 1$.

$$f(y) = \begin{cases} 0 & \text{при } y \leq 0, \\ \frac{y^{\frac{k}{2}-1}}{2^{\frac{k}{2}} \cdot \Gamma\left(\frac{k}{2}\right)} \cdot e^{-\frac{y}{2}} & \text{при } y > 0. \end{cases}$$

С возрастанием числа степеней свободы k распределение χ^2 медленно приближается к нормальному распределению.

