Статистические оценки параметров распределения

Доверительные интервалы

1. **Виды статистических ошибок**
2. **Интервальные оценки**
3. **Доверительные интервалы**

Виды статистических ошибок

Def:

Статистической оценкой неизвестного параметра теоретического распределения называют функцию от наблюдаемых случайных величин.

Для того, чтобы статистические оценки давали «хорошие» приближения оцениваемых параметров, они должны

Def:

Несмещенной называют статистическую оценку Θ^*, математическое ожидание которой равно оцениваемому параметру Θ при любом объеме выборки, т.е.

$\mathcal{M}(\Theta^*) = \Theta$.

Смещенной, если $\mathcal{M}(\Theta^*) \neq \Theta$.

Def:

Эффективной называют статистическую оценку, которая

(при заданном объеме выборки

Def:

Состоятельной называют статистическую оценку, которая при $n \to \infty$ стремится по вероятности к оцениваемому параметру.

Оценки бывают **точечными**, которые определяются одним числом. Все оценки, рассмотренные выше –

Точечные оценки

$$m_{\bar{x}} = \frac{S_x}{\sqrt{n-1}};$$

$$m_{S_x} = \frac{S_x}{\sqrt{2(n-1)}};$$

$$M_{C_v} = \frac{C_v}{\sqrt{2n}}$$

При выборке малого объема точечная оценка может разительно отличаться от оцениваемого параметра, т.е. приводить к грубым ошибкам.

По этой причине при небольшом объеме выборки следует пользоваться интервальными оценками.

Интервальные оценки

Интервальные оценки позволяют установить точность и надежность оценок.

Пусть найденная по данным выборки статистическая характеристика Θ^* служит оценкой неизвестного параметра Θ.

Если $\delta > 0$ и $|\Theta - \Theta^*| < \delta$, то чем меньше δ, тем оценка точнее.

Т.о., положительное число δ характеризует точность оценки.

Однако статистические методы не позволяют категорически утверждать, что оценка Θ^* удовлетворяет неравенству $|\Theta - \Theta^*| < \delta$; можно лишь говорить о вероятности γ, с которой это неравенство осуществляется.

Def:

Надежностью (доверительной вероятностью) оценки Θ по Θ^* называют вероятность γ, с которой осуществляется неравенство $|\Theta - \Theta^*| < \delta$. $\gamma = 0{,}95;\ 0{,}99;\ 0{,}999$.

$$\boxed{P\left[\left|\Theta - \Theta^*\right| < \delta\right] = \gamma}$$

Заменив неравенство $|\Theta - \Theta^*| < \delta$ равносильным уме двойным неравенством

$$-\delta < \Theta - \Theta^* < \delta$$

$$\Theta^* - \delta < \Theta < \Theta^* + \delta$$

$$P\left[\Theta^* - \delta < \Theta < \Theta^* + \delta\right] = \gamma$$

Вероятность того, что интервал $(\Theta^* - \delta; \Theta^* + \delta)$ заключает в себе (покрывает) неизвестный параметр Θ, равна γ.

Доверительный интервал

Def:

Доверительным интервалом называется случайный интервал $(Q^* - \delta;\ O^* + \delta)$, в пределах которого с вероятностью γ находится неизвестный оцениваемый параметр.

Доверительные интервалы находят по различным формулам, в зависимости от исходных данных

Доверительный интервал для математического ожидания нормально распределенного признака с известным средним квадратическим отклонением находят по формуле:

$$\gamma = p(\overline{x}_в - \frac{t\sigma(X)}{\sqrt{n}} < M(X) < \overline{x}_в + \frac{t\sigma(X)}{\sqrt{n}}) = 2\Phi(t),$$

$$\frac{\sigma(X)}{\sqrt{n}}$$

где - среднее квадратическое отклонение,

Приведенная формула позволяет решать следующие задачи:

1) **По заданным надежности γ и объеме выборки** n **находить точность δ и доверительный интервал.**

2) **По заданным надежности γ и точности δ находить объем выборки** n.

3) **По заданным точности δ и объеме выборки** n **находить надежность γ**

В случае большой выборки при $n > 30$ и неизвестном среднем квадратическом отклонении $\sigma(X)$ доверительный интервал находят по формуле:

$$\bar{x}_в - \frac{tS}{\sqrt{n}} < M(X) < \bar{x}_в + \frac{tS}{\sqrt{n}}$$

где S – исправленное выборочное среднее квадратическое отклонение, то есть оценка $\sigma(X)$.

Исследование большой выборки может оказаться невозможным по различным признакам. Кроме этого, с уменьшением n доверительный интервал увеличивается.

При определении доверительного интервала в случае нормального распределения при неизвестном σ признака X в генеральной

$$T = \frac{\bar{x}_в - M(X)}{S}\sqrt{n}$$

Эта величина соответствует закону t **– распределения Стьюдента.**

Дифференциальная функция распределения T **обозначается** $S(t;\ n)$ **и зависит только от объема выборки** n**.**

Вероятность попадания случайной величины в соответствующий интервал равна:

$$\boxed{P(|T| < t_\gamma) = 2\int_0^t S(t_\gamma; n)\,dt = \gamma(t_\gamma; n).}$$

Доверительный интервал для оценки математического ожидания при неизвестном σ.

$$\boxed{\bar{x}_в - t_\gamma \frac{S}{\sqrt{n}} < M(X) < \bar{x}_в + t_\gamma \frac{S}{\sqrt{n}},}$$

где $t_\gamma = t(\gamma; n)$ – числа, приведенные в специальных таблицах.

Примечание: **при большом объеме выборки** $(n \geq 30)$ **значения** $t\gamma$ **таблицы Стьюдента и** t **таблицы Лапласа практически равны. Поэтому выбор формулы, по которой определяют доверительный интервал, диктуется исходными данными.**

Пример

Для определения средней живой массы трехмесячного теленка определенной породы были взвешены *100* животных и

Масса, **кг**	*23-25*	*25-27*	*27-29*	*29-31*	*31-33*	*33-35*	*35-37*
Число телят, **гол**	*3*	*10*	*6*	*16*	*15*	*30*	*20*

Найти:

1) величины, которые следует принять за среднюю массу и среднее квадратическое отклонение;

2) ошибку средней и коэффициетнт вариаций;

3) доверительный интервал, в котором с вероятностью 0,95

Решение

1) В качестве приближенного значения средней массы принимаем выборочную среднюю, а за значение признака – середины интервалов

$$\bar{x}_{в} = \frac{24\cdot 3 + 26\cdot 10 + 28\cdot 6 + 30\cdot 16 + 32\cdot 15 + 34\cdot 30 + 36\cdot 20}{100} = 32 \quad .$$

Вычисляем выборочную исправленную дисперсию

$$S^2 = \frac{1}{99}(3(24-32)^2 + 10(26-32)^2 + 6(28-32)^2 + 16(30-32)^2$$

$$+15(32-32)^2 + 30(34-32)^2 + 20(36-32)^2 = 11,62.$$

Находим исправленное выборочное среднее квадратичное отклонение

$$S = \sqrt{\frac{100}{99}\cdot 11{,}62} = 3{,}45 \quad .$$

2) **Ошибка средней равна**

$$\boxed{\delta_{\bar{x}} = \frac{3,45}{\sqrt{100}} = 0,345 кг}$$

Коэффициент вариации

$$\boxed{V = \frac{3,45}{32}100 = 10,78\%}$$

показывает, что изменчивость признака средняя.

3) **Поскольку** $n = 100 > 30$ **и у нас случай нормального распределения, то доверительный интервал находим по формуле**

$$\overline{x}_{в} - \frac{tS}{\sqrt{n}} < M(X) < \overline{x}_{в} + \frac{tS}{\sqrt{n}}$$

Из условия $2\Phi(t\gamma) = 0.95$ определяем $\Phi(t\gamma) = 0,475$, а по таблице приложений находим $t\gamma = 1,96$.

Поэтому

$$\boxed{32 - \frac{1,96 \cdot 3,45}{\sqrt{100}} < x < 32 + \frac{1,96 \cdot 3,45}{\sqrt{100}}}$$

или $31,32 < x < 32,68$ кг – доверительный интервал для заданной вероятности.

Замечание: если требуется оценить математическое ожидание с наперед заданной точностью **δ** и надежностью **γ**, то максимальный объем выборки, который обеспечит эту точность, находится по формуле

$$\boxed{n = t_{\gamma}^{2}\frac{S^{2}}{\delta^{2}}\left(n \geq \ldots\right)}$$

Объем **выборочной совокупности при повторном способе отбора** **находят по формуле**:

$$\boxed{n = \frac{t^2 S^2(X)}{\delta^2},}$$

где параметр t **определяют из**

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

по таблицам Лапласа,

$$\boxed{\delta = \frac{t S_{\bar{x}}}{\sqrt{n}}.}$$

Доверительный интервал для оценки среднего квадратического отклонения σ нормального распределения.

$$P\left(\left|\sigma - S\right| < \delta\right) = \gamma$$

$$S - \delta < \sigma < S + \delta$$

$$S\left(1 - \frac{\delta}{S}\right) < \sigma < S\left(1 + \frac{\delta}{S}\right)$$

$$\boxed{\frac{\delta}{S} = q}$$

$$\boxed{S(1-q) < \sigma < S(1+q)}$$

q **находят по приложению №**4 **руководства Гмурмана В.С.**