Статистические оценки параметров распределения Доверительные интервалы

1. **Требования к статистическим оценкам**
2. **Точечные оценки**
3. **Интервальные оценки.**

Доверительные интервалы

Виды статистических оценок

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

Для того, чтобы статистические оценки давали «хорошие» приближения оцениваемых параметров, они должны удовлетворять определенным требованиям.

Несмещенной называют статистическую оценку Θ^*, математическое ожидание которой равно оцениваемому параметру Θ при любом объеме выборки, т.е. $M(\Theta^*) = \Theta$.

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

Эффективной называют статистическую оценку, которая (при заданном объеме выборки n) имеет наименьшую возможную дисперсию.

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

Оценки бывают **точечными**, которые определяются одним числом.

Выборочная средняя

- Выборочной средней называют среднее арифметическое значение признака выборочной совокупности.

$$\bar{x}_в = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

- $$\bar{x}_в = \frac{\sum_{i=1}^{k} x_i f_i}{n}$$ взвешенная средняя

Выборочная дисперсия

- Выборочной дисперсией называют среднее арифметическое квадратов отклонения наблюдаемых значений признака от их среднего значения .

$$D_в = \frac{\sum_{i=1}^{n}(x_i - \bar{x}_в)^2}{n} \qquad D_в = \frac{\sum_{i=1}^{k}(x_i - \bar{x}_в)^2 f_i}{n}$$

- Легко «исправить» выборочную дисперсию так, чтобы её математическое ожидание было равно генеральной дисперсии. Достаточно для этого умножить $D_в$ на дробь $\frac{n}{n-1}$.
- Сделав это, получим исправленную дисперсию, которую обычно обозначают через s^2 :

$$s^2 = \frac{n}{n-1} D_в = \frac{n}{n-1} \frac{\sum_{i=1}^{k} (x_i - \overline{x}_в)^2 f_i}{n} = \frac{\sum_{i=1}^{k} (x_i - \overline{x}_в)^2 f_i}{n-1}$$

- Исправленная дисперсия является, несмещённой оценкой генеральной дисперсии.
- Итак, в качестве оценки генеральной дисперсии принимают исправленную дисперсию

$$s^2 = \frac{\sum_{i=1}^{k} (x_i - \overline{x}_k)^2 f_i}{n-1}$$

- Кроме дисперсии для характеристики рассеяния значений признака выборочной совокупности вокруг своего среднего значения пользуются сводной характеристикой – средним квадратическим отклонением.
- **Выборочным средним квадратическим отклонением (стандартным)** называют квадратный корень из выборочной дисперсии: $\sigma_в = \sqrt{D_в}$.
- Для оценки же среднего квадратического отклонения генеральной совокупности используют «исправленное» среднее квадратическое отклонение, которое равно квадратному корню из исправленной дисперсии

$$s = \sqrt{\frac{\sum_{i=1}^{k}(x_i - \overline{x}_в)^2 f_i}{n-1}}$$

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.

Доверительный интервал

Доверительным интервалом называется случайный интервал $(\Theta^* - \delta; \Theta^* + \delta)$, который покрывает неизвестный параметр с заданной надежностью γ.

Метод доверительных интервалов разработал американский статистик Ю. Нейман, исходя из идей английского статистика Р.Фишера.

Доверительный интервал для оценки математического ожидания нормального распределения при известном среднем квадратическом отклонении

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

t – параметр, величину которого находят по таблицам Лапласа из соотношения $\gamma=2\Phi(t)$.

- **Смысл полученного соотношения таков: с надёжностью γ можно утверждать, что доверительный интервал $\left(\bar{x} - \frac{t\sigma}{\sqrt{n}}, \bar{x} + \frac{t\sigma}{\sqrt{n}}\right)$ покрывает неизвестный параметр ; точность оценки $\delta = \frac{t\sigma}{\sqrt{n}}$.**

- **Укажем ещё, что число *t* определяется из равенства $2\Phi(t) = \gamma$, или $\Phi(t) = \frac{\gamma}{2}$; по таблице функции Лапласа находят аргумент *t*, которому соответствует значение функции Лапласа, равное $\frac{\gamma}{2}$.**

- **Доверительным вероятностям, как это видно из таблицы функции Лапласа, соответствуют следующие величины нормированных отклонений:**
- **вероятности γ =0,95 соответствует t_1 = 1,96; вероятности γ = 0,99 соответствует t_2 = 2,58; вероятности γ = 0,999 соответствует t_3= 3,29.**
- **Выбор того или иного порога доверительной вероятности исследователь осуществляет исходя из практических соображений той ответственности, с какой делаются выводы о генеральных параметрах.**

Примечание: **при большом объеме выборки**
($n \geq 30$) значения t_{γ} 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 результаты сведены в таблицу

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 коэффициент вариации;
3) доверительный интервал, в котором с вероятностью 0,95 заключена средняя масса.

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

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

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

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

Из условия **2Ф(t γ) = 0.95** определяем **Ф(t γ) = 0,475**, а по таблице приложений находим **t γ = 1,96**.

Поэтому

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

или **31,3 < a < 32,7 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

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

Доверительный интервал для оценки среднего квадратического отклонения σ нормального распределения.

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

$$\boxed{S - \delta < \sigma < S + \delta}$$

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

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

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

q находят по приложению №4 руководства Гмурмана В.Е.