§2. Точечные оценки параметров ГС

Пусть распределение СВ X (некоторого признака ГС) задаётся вероятностями $p_i=p_i(x_i, \theta)$ (для дискретной СВ) или плотностью вероятности $f=f(x,\theta)$ (для непрерывной СВ), которые зависят от неизвестного параметра θ.

Этим параметром может быть, например, параметр λ закона Пуассона или параметры a и σ нормального распределения.

- Оценки параметров ГС, полученные на основании выборки, называются *статистическими*.

- Если статистическая оценка характеризуется одним числом, то она называется *точечной*.

Понятие несмещённости, состоятельности и эффективности

Статистическая оценка $\tilde{\theta}$ является случайной величиной и меняется в зависимости от выборки. Сам параметр θ является некоторым постоянным (неслучайным) числом, которое представляет истинное значение параметра ГС.

Ясно, что оценку следует выбирать т.о., чтобы её значения как можно точнее оценивали значение неизвестного параметра θ.

- Оценка $\tilde{\theta}$ называется *несмещённой*, если

$$\boxed{M\left(\tilde{\theta}\right)=\theta.}$$

Если это требование не выполняется, то в среднем оценка $\tilde{\theta}$ будет всегда давать значение θ с некоторым отклонением.

- Несмещённая оценка $\tilde{\theta}$ называется *эффективной*, если $D\left(\tilde{\theta}\right)$ является наименьшей среди дисперсий всех возможных оценок параметра θ, вычисленных по одному и тому же объёму выборки n.

§3. Интервальные оценки параметров ГС

- Оценка $\tilde{\theta}$ называется *состоятельной*, если

для $\forall\, \varepsilon > 0$

$$\boxed{\lim_{n\to\infty} P\left(\left|\tilde{\theta} - \theta\right| < \varepsilon\right) = 1.} \qquad \textbf{(1)}$$

Условие (1) означает, что $\tilde{\theta}$ стремится к θ по вероятности (пишут $\tilde{\theta} \xrightarrow{P} \theta$), так что при больших *n* отклонение $\tilde{\theta}$ от θ становится сколь угодно малым.

При выборке малого объёма точечная оценка может существенно отличаться от истинного значения неизвестного параметра. Поэтому пользуются интервальными оценками.

- Оценка, определённая двумя числами – концами интервала, называется *интервальной*.

Доверительная вероятность. Доверительный интервал

Пусть $\tilde{\theta}$ – оценка неизвестного параметра θ, полученная по данным выборки.

Очевидно, оценка тем точнее, чем меньше $|\theta - \tilde{\theta}|$.

Пусть $\delta > 0$ и $|\theta - \tilde{\theta}| < \delta$. Тогда, чем меньше δ, тем точнее оценка $\tilde{\theta}$.

- Число δ называется *точностью* оценки.
- *Доверительной вероятностью* (*надёжностью*) оценки $\tilde{\theta}$ параметра θ называется число γ, равное вероятности

$$\boxed{P\left(\left|\theta - \tilde{\theta}\right| < \delta\right) = \gamma} \qquad \textbf{(1)}$$

Формулу (1) можно записать так:

$$\boxed{P\left(\tilde{\theta}-\delta<\theta<\tilde{\theta}+\delta\right)=\gamma} \qquad \textbf{(2)}$$

Формула (2) означает: вероятность того, что интервал $\left(\tilde{\theta}-\delta;\tilde{\theta}+\delta\right)$ заключает в себе неизвестный параметр θ, равна γ.

- Интервал $\left(\tilde{\theta}-\delta;\tilde{\theta}+\delta\right)$ называется *доверительным*. Концы интервала – *доверительными границами*.

Замечание 1. Обычно надёжность γ задаётся заранее. В качестве γ берут число, близкое к 1.

Замечание 2. Доверительные границы являются случайными величинами (они изменяются от выборки к выборке).

Доверительный интервал для оценки параметра a нормального распределения

Пусть признак X генеральной совокупности имеет нормальное распределение с заданным σ и неизвестным a.

Доверительный интервал, заключающий в себе параметр a с надёжностью (вероятностью) γ , определяется двойным неравенством:

$$\boxed{\bar{x} - \frac{t\sigma}{\sqrt{n}} < a < \bar{x} + \frac{t\sigma}{\sqrt{n}}}$$

При этом точность оценки

$$\delta = \frac{t\sigma}{\sqrt{n}}.$$

Замечание. Число t определяется из равенства

$$2\Phi(t) = \gamma.$$

Значение t находится с помощью таблиц функции Лапласа.

Пусть σ неизвестно. Тогда вместо него используют исправленное ско S, которое является оценкой σ.

Доверительный интервал для оценки параметра σ нормального распределения

Доверительный интервал, заключающий в себе параметр σ с надёжностью (вероятностью) γ, определяется двойным неравенством:

$$\boxed{\frac{S}{1+q} < \sigma < \frac{S}{1-q},}$$

где S – несмещённое ско, q – параметр, который определяется из таблиц по известным значениям γ и n.

- **Задание 1 (24.21 (Ерм.)).**

Найдите доверительный интервал с надёжностью 0.8 для оценки математического ожидания нормально распределённой СВ X с ско, равным 5, выборочным средним 20 и объёмом выборки 25.

?

$(18.72; 21.28)$

- **Задание 2 (24.22 (Ерм.)).**

На овцеводческой ферме из стада произведена выборка для взвешивания 36 овец. Их средний вес оказался равным 50 кг. Предположив распределение веса нормальным и определив несмещённую оценку дисперсии $S^2 = 16$, найдите доверительный интервал для оценки математического ожидания с надёжностью 0.9.

?

$(48.9; 51.5)$

- **Задание 3 (24.25 (Ерм.)).**

По данным выборки объёма 20 найдено несмещённое ско, равное 2, нормально распределённой СВ. Найдите с надёжностью 0.95 доверительный интервал для оценки ско этой СВ.

?