Введение в экономико-математическое моделирование

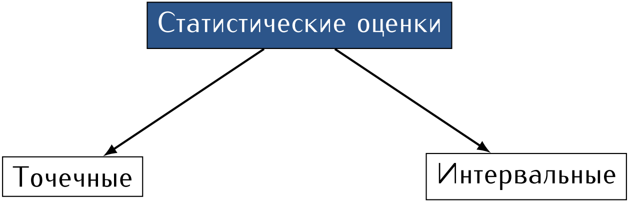
Лекция 16. Статистические оценки

канд. физ.-матем. наук, доцент Д. В. Чупраков

usr10381@vyatsu.ru

Статистическая оценка

Приближенное значение θ^* характеристики θ случайной величины X, вычисленное по выборке V называется статистической оценкой характеристики θ.



- Оценка определяемая одним числом, называется точечной.
- Промежуток, который с некоторой вероятностью накрывает значение характеристики θ называется интервальной оценкой.

Три орешка для Золушки

- Несмещенность — среднее значение оценок $\theta^*(V)$ по всевозможным выборкам равно оцениваемой характеристике θ:
$$M\big(\theta^*(V)\big) = \theta$$

- Состоятельность — с увеличением числа опытов n случайная величина θ^* приближается к θ
$$|\theta^*(V_n) - \theta| \xrightarrow{n \to N} 0$$

- Эффективность — оценка θ^* обладает наименьшей дисперсией по сравнению с другими оценками параметра θ:
$$D(\theta^*) \xrightarrow{\theta^*} \min$$

Каждая несмещенная оценка характеризуется своей дисперсией.

Генеральная доля

Пусть признаком A обладают M элементов генеральной совокупности, содержащей N элементов.

Рассмотрим в качестве оценки выборочную долю $w = m/n$ элементов обладающих признаком À в выборке.

Теорема

Выборочная доля $\omega = \frac{m}{n}$ является несмещенной и состоятельной оценкой генеральной доли $p = \frac{M}{N}$.
Дисперсия оценки равна σ_w

Выборка

Повторная: $$\sigma_w = \frac{pq}{n}$$

Бесповторная: $$\sigma_w = \frac{pq}{n}\left(\frac{N-n}{N-1}\right)$$

Генеральное среднее

Теорема

Выборочная средняя $\bar{x}_в$ есть несмещенная и состоятельная оценка генеральной совокупности $\bar{x}$.
Дисперсия оценки равна $\sigma_{\bar{x}}$

Выборка

Повторная

$$\sigma_{\bar{x}} = \frac{\sigma^2}{n}$$

Бесповторная

$$\sigma_{\bar{x}} = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

При нормальном распределении величины X среднее выборочное $\bar{x}_{\text{в}}$ будет эффективной оценкой генеральной средней

Генеральная дисперсия

Генеральную дисперсию σ^2 оценим выборочной дисперсией s^2?

Генеральная дисперсия

Генеральную дисперсию σ^2 оценим выборочной дисперсией s^2?

Это не лучший выбор:

Теорема

*Выборочная дисперсия s^2 является смещенной,
но состоятельная оценкой генеральной дисперсии σ^2.*

Генеральная дисперсия

Генеральную дисперсию σ^2 оценим выборочной дисперсией s^2?

Это не лучший выбор:

Теорема

*Выборочная дисперсия s^2 является <u>смещенной</u>,
но состоятельная оценкой генеральной дисперсии σ^2.*

Рассмотрим исправленную выборочную дисперсию:

$$\hat{s}^2 = \frac{n}{n-1} s^2$$

Теорема

*Исправленная выборочная дисперсия $\hat{s}^2$ является <u>несмещенной</u>
и состоятельная оценкой генеральной дисперсии σ^2.*

Теорема

Эмпирическая функция распределения выборки $F^*(x)$, *построенная по статистическому дискретному ряду, является несмещенной состоятельной оценкой функции распределения* $F(x)$ *случайной величины* X.

Проблема точечных оценок

Допустим, мы нашли способ по выборке V оценить неизвестный параметр θ.

Точечная оценка зависит от элементов в выборке, поэтому она случайна.

Использование точечной оценки может привести к существенным ошибкам.

Как оценить надёжность замены реального параметра θ его точечной оценкой θ^*?

Рассмотрим промежуток $[\theta^* - \varepsilon; \theta^* + \varepsilon]$

Доверительная вероятность γ — вероятность накрыть интервалом $[\theta^* - \varepsilon; \theta^* + \varepsilon]$ величину характеристики θ:

$$\gamma = P(\theta^* - \varepsilon \leqslant \theta \leqslant \theta^* + \varepsilon)$$

Доверительный интервал:

$$I_\gamma = [\theta^* - \varepsilon; \theta^* + \varepsilon]$$

Свойства интервальной оценки

- Параметр ε — точность доверительного интервала

  Чем меньше ε, тем точнее интервальная оценка.
- Параметр γ — надежность доверительного интервала

  Чем больше γ, тем надежнее интервал.

Противоречивость параметров

При фиксированном размере выборки выборке увеличение точности ведет к снижению надежности и наоборот.

Доверительные интервалы для параметров нормального распределения

Интервал для генеральной выборочной I

$$X \sim N(\mu, \sigma)$$

Найдем доверительный интервал для $\bar{x}$ с надежностью γ
Выборка
$$X_1, X_2, \ldots, \ X_n$$

- С. В. X_i — независимы,
- $X_i \sim N(\mu, \sigma)$

Значит,

$$M(X_1) = M(X_2) = \ldots = M(X_n) = M(X) = \mu,$$
$$D(X_1) = D(X_2) = \ldots = D(X_n) = D(X) = \sigma^2.$$

Выборочное среднее $\overline{X}_{\text{в}}$ также будет распределено по нормальному закону.
Найдем параметры распределения:
Математическое ожидание:

$$M(\overline{X}_{\text{в}}) = M\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} M(X_i) = \frac{1}{n}\sum_{i=1}^{n} M(X) =$$
$$= \frac{1}{n}\cdot n\cdot M(X) = M(X) = \bar{x}$$

Дисперсия:

$$D(\overline{X}_{\text{в}}) = D\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i) = \frac{1}{n^2}\sum_{i=1}^{n} D(X) =$$

$$= \frac{1}{n^2}\cdot n\cdot D(X) = \frac{D(X)}{n} = \frac{\sigma^2}{n} = \sigma_{\bar{x}}^2$$

Итак,

$$M(\overline{X}_{\text{в}}) = \bar{x} \qquad D(\overline{X}_{\text{в}}) = \sigma_{\bar{x}}^2$$

Воспользуемся формулой вероятности отклонения С.В. $X \sim N(\mu, \sigma)$ от своего математического ожидания на величину, не превосходящую Δ:

$$\gamma = P(|\bar{x}_{\text{в}} - \bar{x}| < \Delta) = 2\Phi\left(\frac{\Delta}{\sigma_{\bar{x}}}\right) = 2\Phi(t),$$

$$t = \frac{\Delta}{\sigma_{\bar{x}}}, \quad \sigma_{\bar{x}} = \sqrt{\frac{D(X)}{n}}$$

Теорема

Доверительный интервал для $\bar{x}$ *есть*

$$(\bar{x} - t \cdot \sigma_{\bar{x}}; \bar{x} + t \cdot \sigma_{\bar{x}})$$

где t *определяется из уравнения* $\Phi(t) = \gamma/2$.

Выборка

Повторная:

$$\sigma_{\bar{x}} = \frac{\sigma^2}{n} \approx \frac{s^2}{n}$$

Бесповторная:

$$\sigma_{\bar{x}} = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) \approx \frac{s^2}{n}\left(1 - \frac{n}{N}\right)$$

Интервальные оценки

Параметр	Генеральная средняя	Генеральная доля
Интервал	$\lvert\bar{x}-\bar{x}_{\text{в}}\rvert \leqslant \Delta$ $\bar{x}_{\text{в}}-\Delta \leqslant \bar{x} \leqslant \bar{x}_{\text{в}}+\Delta$	$\lvert p-\omega_{\text{в}}\rvert \leqslant \Delta$ $w-\Delta \leqslant p \leqslant w+\Delta$
Отклонение	$\Delta = t\sigma_{\overline{x}}$ $\gamma = 2\Phi(t)$	$\Delta = t\sigma_{\omega}$ $\gamma = 2\Phi(t)$
Тип выборки		
Повторная	$\sigma_x = \sqrt{\frac{\sigma^2}{n}}$ $\sigma_x \approx \sqrt{\frac{s^2}{n}}$	$\sigma_w = \sqrt{\frac{p(1-p)}{n}}$ $\sigma_w \approx \sqrt{\frac{\omega(1-\omega)}{n}}$
Бесповторная	$\sigma_x = \sqrt{\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)}$ $\sigma_x \approx \sqrt{\frac{s^2}{n}\left(1-\frac{n}{N}\right)}$	$\sigma_w = \sqrt{\frac{p(1-p)}{n}\left(\frac{N-n}{N-1}\right)}$ $\sigma_w \approx \sqrt{\frac{\omega(1-\omega)}{n}\left(1-\frac{n}{N}\right)}$

Пример I

Случайная величина X имеет нормальное распределение с известным средним квадратическим отклонением $\sigma = 3$. Найти доверительный интервал для оценки неизвестного математического ожидания μ по выборочным средним $\bar{x}$, если объем выборки $n = 36$ и задана надежность оценки $\gamma = 0.95$.

- Так как объем генеральной совокупности не указан, считаем выборку повтороной.
- Найдем параметр t.

$$2\Phi(t) = 0.95$$
$$\Phi(t) = 0.475$$
$$t \approx 1.96$$

Пример II

- Найдем $\sigma_{\bar{x}}$:

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{3^2}{36}} = \frac{1}{2}.$$

- Найдем точность

$$\Delta = t\sigma_{\bar{x}} = \frac{1.96}{2} = 0.98$$

Ответ: Математическое ожидание отклоняется от выборочного математического ожидания не более чем на 0.98.

Пример

С целью размещения рекламы опрошено 420 телезрителей, из которых данную передачу смотрят 170 человек. С доверительной вероятностью $\gamma = 0.91$ найти долю телезрителей, охваченных рекламой в лучшем случае.

- Объем генеральной совокупности не известен. Считаем выборку посторной.
- Для оценки неизвестной доли телезрителей используем доверительный интервал:

$$w - \Delta \leqslant p \leqslant w + \Delta, \qquad \Delta = t\sigma_{\bar{x}}, \quad \gamma = 2\Phi(t)$$

- Найдем выборочную долю:

$$w = \frac{170}{420} \approx 0.405$$

- Найдем параметр t:

Оценка объема выборки

Объем выборки влияет на качество выборочного исследования и определяет необходимые временные, трудовые и стоимостные затраты.

Для определения объема выборки n необходимо задать два параметра:

- надежность γ оценки исследуемой характеристики
- точность Δ оценки исследуемой характеристики

Объем повторной выборки

Характеристика

Среднее

$$\boxed{n = \frac{t^2\sigma^2}{\Delta^2}}$$

Доля

$$\boxed{n = \frac{t^2 pq}{\Delta^2}}$$

Объем бесповторной выборки

$$n' = \frac{nN}{n+N}$$

Пример вычисления объема выборки I

Строительная компания хочет оценить среднюю стоимость ремонтных работ, выполняемых для клиентов. Каким должен быть объем бесповторной выборки среди 1200 клиентов строительной фирмы, если среднее квадратическое отклонение по результатам пробного обследования составило 85 тыс. руб., а предельная ошибка выборки не должна превышать 20 тыс. руб с вероятностью 0.95?

- Имеем $N = 1200$, $\gamma = 0.95$, $\Delta = 20$, $\sigma = 85$

Пример вычисления объема выборки II

- Найдем параметр t

$$2\Phi(t) = 0.95$$
$$\Phi(t) = 0.475$$
$$t \approx 1.96$$

- Найдем объем повторной выборки для оценки средней стоимости:

$$n = \frac{t^2\sigma^2}{\Delta^2} = \frac{1.96^2 \cdot 85^2}{20^2} \approx 69.3889$$

Не округляем, так как это промежуточный результат.

- Найдем объем бесповторной выборки

$$n' = \frac{nN}{n+N} = \frac{69.3889 \cdot 1200}{69.3889 + 1200} \approx 66$$

Округляем в большую сторону.

Ответ: Достаточно исследовать данные работ по 66 клиентам.

Квантиль

α–кванти́ль распределения — такое значение x_α, что для С. В. X, имеющей данное распределение, выполнено свойство

$$P(X < x_\alpha) = \alpha$$

p–й Перцентиль — квантиль уровня $\alpha = \frac{p}{100}$

Найдем доверительный интервал для среднего квадратического отклонения нормально распределенной величины

$$X \sim N(\bar{x}, \sigma)$$

с заданной надежностью γ

Если генеральное среднее $\bar{x}$ известно, то доверительный интервал для генеральной дисперсии σ^2 имеет вид:

$$\frac{ns^2}{\chi_2^2} < \sigma < \frac{ns^2}{\chi_1^2},$$

где

- n — объем выборки
- $\chi_1^2 = \chi^2_{\frac{1+\gamma}{2}, n}$, $\chi_2^2 = \chi^2_{\frac{1-\gamma}{2}, n}$ — квантили χ^2-распределения с n степенями свободы.

Оценка дисперсии II

Если математическое ожидание $\bar{x}$ неизвестно, то доверительный интервал для генеральной дисперсии σ^2 имеет вид:

$$\frac{(n-1)\hat{s}^2}{\chi_2^2} < \sigma < \frac{(n-1)\hat{s}^2}{\chi_1^2},$$

где

- n — объем выборки
- $\chi_1^2 = \chi^2_{\frac{1+\gamma}{2},n-1}$, $\chi_2^2 = \chi^2_{\frac{1-\gamma}{2},n-1}$ — квантили χ^2-распределения с $n-1$ степенями свободы.

Пример оценки дисперсии I

Для оценки параметра нормально распределенной случайной величины была сделана выборка объема в 30 единиц и вычислено $s = 1.5$. Найти доверительный интервал, покрывающий σ с надежностью $\gamma = 0.90$.

- Имеем $n = 30$, $\gamma = 0.9$.
- Математическое ожидание не известно.
- Исправим дисперсию $\hat{s}^2 = \frac{30}{29}(1.5)^2 \approx 2.328$
- Вычислим квантили

$$\chi_1^2 = \chi^2_{\frac{1+0.9}{2},30-1} = \chi^2_{0.95,29} = 17.7$$
$$\chi_2^2 = \chi^2_{\frac{1-0.9}{2},30-1} = \chi^2_{0.05,29} = 42.6$$

Пример оценки дисперсии II

- Найдем интервал

$$\frac{n-1}{\chi_2^2}\hat{s}^2 < \sigma^2 < \frac{\sqrt{n-1}\hat{s}^2}{\chi_1^2}$$
$$\frac{29 \cdot 2.328}{42.6} < \sigma < \frac{29 \cdot 2.328}{17.76}$$
$$1,585 < \sigma^2 17 < 1.579$$