

23. Элементы математической статистики-1

23.1. Предварительные сведения

Определения генеральной совокупности, единицы генеральной совокупности, случайной выборки, выборочных значений, вариационного ряда, эмпирической функции распределения, статистики даны на лекциях (см. конспект).

Пусть x_1, \dots, x_n — выборочные значения.

Выборочное среднее — это число $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$.

Выборочная дисперсия — это число $\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}_n^2$.

Исправленная выборочная дисперсия — это число $s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2 = \frac{1}{n-1} \sum_{i=1}^n x_i^2 - \frac{n}{n-1} \bar{x}_n^2$.

Оценка $\hat{\theta}_n$ параметра θ называется

- несмещенной, если $\mathbb{E}\hat{\theta}_n = \theta$;
- асимптотически несмещенной, если $\mathbb{E}\hat{\theta}_n \rightarrow \theta$ при $n \rightarrow \infty$;
- состоятельной, если $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta$ при $n \rightarrow \infty$.

23.2. Практическое занятие

1. Попросите 15–25 студентов вашей группы назвать свой рост, после чего

- 1) запишите полученные выборочные значения;
- 2) постройте вариационный ряд;
- 3) постройте график эмпирической функции распределения;
- 4) сгруппировав выборочные значения по интервалам вида 180 ± 2.5 см, постройте гистограмму и полигон частот;
- 5) посмотрите, как изменятся гистограмма и полигон частот при увеличении ширины интервалов до 10 см;
- 6) предложите несколько статистик для оценки среднего роста студента и найдите численные значения этих оценок по выборочным данным.

2. Убедитесь в том, что выборочная функция распределения является несмещенной оценкой для функции распределения $F(x)$ генеральной совокупности X в точке $x \in \mathbb{R}$.

3. Пусть $(x_1, y_1), \dots, (x_n, y_n)$ — независимая выборка, соответствующая случайному вектору (X, Y) . Покажите, что величина

$$m = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}),$$

где $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$, является несмещенной и состоятельной оценкой ковариации $\text{Cov}(X, Y)$.

4. Пусть (X_1, X_2, X_3) — случайная выборка объема $n = 3$ из генеральной совокупности X , распределенной по экспоненциальному закону с плотностью

$$p(x; \lambda) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Имеются следующие три оценки математического ожидания, которое равно $1/\lambda$:

$$\hat{\theta} = \frac{X_1 + X_2 + X_3}{3}, \quad \tilde{\theta} = \frac{X_{(1)} + X_{(3)}}{2}, \quad \theta^* = X_{(2)}.$$

Проверьте эти оценки на несмещенность и исправить смещенные оценки соответствующим образом. Сравните дисперсии несмещенных оценок, полученных из $\hat{\theta}$ и θ^* .

Ответ. Оценка $\hat{\theta}$ является несмещенной, оценки $\tilde{\theta}$ и θ^* — смещенные; исправленные оценки равны $\frac{12}{13}\tilde{\theta}$ и $\frac{6}{5}\theta^*$ соответственно; $\text{Var } \hat{\theta} = \frac{1}{3\lambda^2}$, $\text{Var}(\frac{6}{5}\theta^*) = \frac{13}{25\lambda^2}$.

23.3. Домашнее задание

5. В результате измерения диаметров 200 валиков из партии, изготовленной одним станком-автоматом, получены отклонения измеренных диаметров от номинала (в микрометрах). Сгруппированные данные представлены в виде интервального статистического ряда. Постройте гистограмму и полигон частот. Найдите выборочное среднее, выборочную дисперсию и исправленную выборочную дисперсию.

J_i	$[-20, -15)$	$[-15, -10)$	$[-10, -5)$	$[-5, 0)$	$[0, 5)$	$[5, 10)$	$[10, 15)$	$[15, 20)$	$[20, 25)$	$[25, 30]$	\sum
n_i	7	11	15	24	49	41	26	17	7	3	200

Указание. В качестве выборочных значений использовать середины интервалов.

Ответ. $\bar{x}_{200} = 4.3$ мкм, $\hat{\sigma}_{200}^2 \approx 94.26$ мкм², $s_{200}^2 = 94.73$ мкм².

6. Дана случайная выборка объема n из генеральной совокупности X с плотностью распределения

$$p(x; \theta) = \begin{cases} e^{\theta-x}, & x \geq \theta, \\ 0, & x < \theta, \end{cases}$$

и оценка $\hat{\theta}_n = X_{(1)}$ неизвестного параметра θ .

- 1) Покажите, что эта оценка является смещенной.
- 2) Покажите, что эта оценка является асимптотически несмещенной;
- 3) Исправьте оценку $\hat{\theta}_n$ таким образом, чтобы она стала несмещенной;
- 4) Вычислите дисперсию $\text{Var } \hat{\theta}_n$;
- 5) Покажите, что оценка $\hat{\theta}_n$ является состоятельной.

Указание. Для доказательства состоятельности используйте тот факт, что $\text{Var } \hat{\theta}_n \rightarrow 0$ при $n \rightarrow \infty$.