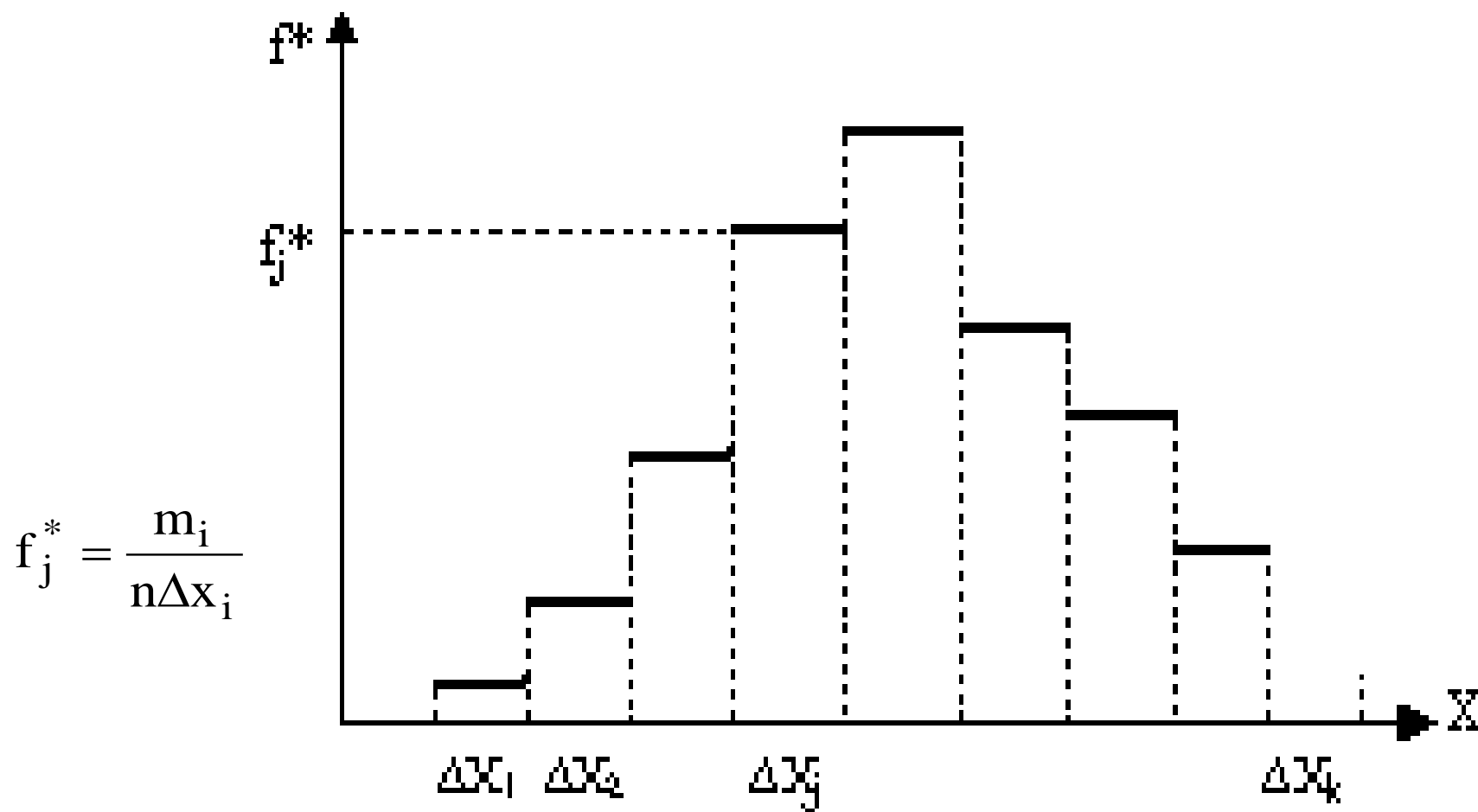


ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Основные понятия математической статистики

- **Совокупность** - это множество объектов (элементов совокупности), обладающих общим свойством.
- **Объем совокупности** - это число элементов совокупности.
- **Генеральная совокупность** - это наибольшая совокупность, объединяющая все элементы, обладающие каким-либо свойством (свойством, наличие которого позволяет причислить элементы к данной совокупности).
- **Выборочная совокупность (выборка)** - это часть генеральной совокупности, выбранная для изучения.
- **Вариантой** называют значение величины X для отдельного элемента выборки.

ГИСТОГРАММА



Точечные оценки:

Оптимальная выборочная оценка
математического ожидания (среднее
выборочное)

$$\hat{M}(X) = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \dots + x_n}{n}$$

- Оптимальная выборочная оценка дисперсии

$$\begin{aligned}\hat{D}(X) &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \\ &= \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n-1}\end{aligned}$$

$$\hat{D}(X) = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n(\bar{x})^2 \right)$$

Выборочная оценка среднего квадратичного отклонения

$$\hat{\sigma}(X) = S(X) = \sqrt{\hat{D}(X)}$$

Ошибка среднего

$$m_x = \frac{S}{\sqrt{n}}$$

Доверительные оценки (интервальные оценки)

$$n \geq 30$$

$$\bar{x} - t(\alpha) \sqrt{\frac{D(X)}{n}} \leq M(X) \leq \bar{x} + t(\alpha) \sqrt{\frac{D(X)}{n}}$$

$t(\alpha)$ находится по таблицам значений функции Лапласа, исходя из условия:

$$\alpha = 0,95 \quad (p = 0,05) \quad t(\alpha) = 1,96$$

$$\varphi(t) = \frac{\alpha}{2}$$

$$n < 30$$

$$\bar{x} - t(\alpha, k) \cdot \frac{S}{\sqrt{n}} \leq M(X) \leq \bar{x} + t(\alpha, k) \cdot \frac{S}{\sqrt{n}}$$

k - число степеней свободы ($k = n - 1$),

$t(\alpha, k)$ - коэффициент Стьюдента,

Проверка статистических гипотез

$$k = n_1 + n_2 - 2 \qquad H = (n_1 - 1)S_1^2 + (n_2 - 1)S_2^2$$

$$T = \sqrt{\frac{n_1 n_2 k}{(n_1 + n_2)H}} \cdot |\bar{x}_1 - \bar{x}_2|$$

Если $T > t$, то разность средних выборочных достоверна.

Если $T < t$, то разность не является достоверной.

$$T = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{m_1^2 + m_2^2}}$$

В результате проведения исследования получена выборка со следующими элементами: **28; 30; 32; 27; 30; 33; 28; 31; 31**. Найти оптимальные выборочные оценки математического ожидания, дисперсии и ошибки среднего и построить доверительный интервал для математического ожидания соответствующий доверительной вероятности $\alpha = 0,95$.

$$\hat{M}(X) = \bar{x} = 30$$

$$\hat{D}(X) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = 4$$

$$\hat{\sigma}(X) = S(X) = \sqrt{\hat{D}(X)} = 2$$

$$m_x = \frac{S}{\sqrt{n}} = 0,67$$

$$\bar{x} - t(\alpha, k) \cdot \frac{S}{\sqrt{n}} \leq M(X) \leq \bar{x} + t(\alpha, k) \cdot \frac{S}{\sqrt{n}}$$

$$28,45 \leq M(X) \leq 31,55$$

**Корреляционная
зависимость между
случайными величинами,
регрессия, коэффициенты
корреляции и регрессии**

Если условная плотность вероятности величины Y ($f(Y/x)$) зависит от величины X , а условная плотность вероятности величины X ($\varphi(X/y)$) зависит от величины Y , то говорят, что между величинами Y и X существует **корреляционная зависимость**.

Условное математическое ожидание случайной величины Y при заданном значении величины X

$$M(Y / x) = \psi(x),$$

где функция $\psi(x)$ называется **функцией регрессии Y на X** .

График функции регрессии называется **линией регрессии**.

Если в выражении для $\psi(x)$ присутствуют какие-нибудь постоянные коэффициенты, то они называются **коэффициентами регрессии**.

Условное математическое ожидание случайной величины X при заданном значении величины Y

$$M(X / y) = \xi(y), \quad \text{где } \xi(y) \text{ - это функция регрессии } X \text{ на } Y.$$

Корреляционное поле - это точки с координатами y_i и x_i ($i = 1, 2, \dots, n$), нанесенные на координатную плоскость XOY .

В случае линейной функции регрессии (при наличии зависимости вида $y = ax + b$) выборочный коэффициент корреляции определяется по формуле

$$R = \frac{\sum_{i=1}^n (x_i \cdot y_i) - n \cdot \bar{x} \cdot \bar{y}}{(n-1) \cdot S(X) \cdot S(Y)} = \frac{\sum_{i=1}^n (x_i \cdot y_i) - n \cdot \bar{x} \cdot \bar{y}}{\sqrt{\left(\sum_{i=1}^n x_i^2 - n \cdot \bar{x}^2 \right) \cdot \left(\sum_{i=1}^n y_i^2 - n \cdot \bar{y}^2 \right)}}$$

$$-1 \leq R \leq 1$$

Если $R > 0$, то функции регрессии Y на X и X на Y – это *возрастающие функции*, а если $R < 0$ – *убывающие*.

Коэффициент корреляции является мерой степени линейности зависимости между случайными величинами,

При определении функции регрессии принято считать оптимальными те оценки коэффициентов регрессии, которые получены на основе применения *метода наименьших квадратов*.

В случае линейной регрессии вида $\psi(x) = ax + b$ значения коэффициентов a и b определяют, минимизируя сумму

$$\sum_{i=1}^n (y_i - ax_i - b)^2$$

В итоге для оптимальных выборочных оценок коэффициентов регрессии a и b получают следующие выражения

$$a = \frac{\sum_{i=1}^n x_i y_i - n \cdot \bar{x} \cdot \bar{y}}{\sum_{i=1}^n x_i^2 - n \cdot \bar{x}^2}$$

$$b = \frac{\bar{y} \cdot \sum_{i=1}^n x_i^2 - \bar{x} \cdot \sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2 - n \cdot \bar{x}^2}$$

последние выражения могут быть преобразованы к виду

$$a = R \frac{S(Y)}{S(X)}$$

$$b = \bar{y} - a\bar{x}$$

В случае регрессии X на Y функция регрессии имеет вид

$$\xi(y) = a_1 y + b_1,$$

а коэффициенты регрессии a_1 и b_1 вычисляются по формулам

$$a_1 = R \frac{S(X)}{S(Y)}, \quad b_1 = \bar{x} - a_1 \bar{y}.$$

Методы оценки корреляционной зависимости в случае, если хотя бы один из признаков является качественным (не количественным).

Коэффициент ранговой корреляции (коэффициент корреляции *Спирмена*)

$$R = 1 - \frac{6 \sum_{i=1}^n (x_i - y_i)^2}{n(n^2 - 1)},$$

где x_i ($i = 1, 2, \dots, n$) – это значение ранга по одному признаку для i -го элемента выборки,

y_i – значение ранга по другому признаку для того же элемента выборки.