

А.Л.КРОХИН

МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ
ПРОВЕДЕНИЯ
лабораторной работы

Екатеринбург 2011

Цель работы: Изучение статистической зависимости между двумя случайными величинами.

Основные задачи:

1. Сбор статистического материала. Получить выборку двух СВ, связанных с одним ЭСС. Например, температура воздуха и время, необходимое для поездки от дома до университета. Объем выборки не менее 20 пар.
2. Выполнение первичной обработки выборки: группировка по необходимости, нахождение числовых характеристик, построение диаграммы рассеивания.
3. Вывод уравнения прямой линии регрессии.
4. Вычисление параметров линии регрессии и оценка коэффициента корреляции. Построить линию регрессии на диаграмме рассеивания.
5. Проанализировать полученные результаты.
6. Подготовить отчет по работе.

Краткое описание методики

Статистическое описание и выборочные характеристики двумерного случайного вектора. Пусть $(x_i, y_i), i = 1, 2, \dots, n$, — выборка объема n из наблюдений случайного двумерного вектора (ξ, η) . Предварительное представление о двумерной генеральной совокупности можно получить, изображая элементы выборки точками на плоскости с выбранной декартовой прямоугольной системой координат. Это представление выборки называется *диаграммой рассеивания*.

Распределением двумерной выборки называется распределение двумерного дискретного случайного вектора, принимающего значения $(x_i, y_i), i = 1, 2, \dots, n$, с вероятностями, равными $1/n$ (если значения не повторяются). Выборочные числовые характеристики, вычисляются как соответствующие числовые характеристики двумерного случайного вектора дискретного типа

Вычислить выборочные средние; дисперсии и коэффициент корреляции для выборки, приведенной в таблице 1.4. Построить диаграмму рассеивания.

Вычисление указанных выборочных характеристик удобно выполнять в следующей последовательности. Сначала вычисляют суммы

$$\sum x_i, ; \sum y_i, ; \sum x_i^2, ; \sum y_i^2, ; \sum x_i y_i, ; \sum (x_i + y_i)^2.$$

Выборочные средние отсюда находятся по формулам

$$\bar{x}_в = \frac{1}{n} \sum_{i=1}^n x_i, ; \bar{y}_в = \frac{1}{n} \sum_{i=1}^n y_i, ; \tag{1}$$

затем вычисляем

$$Q_x = \sum (x_i - \bar{x}_B)^2 = \sum x_i^2 - \frac{1}{n} \left(\sum x_i \right)^2; \quad (2)$$

$$Q_y = \sum (y_i - \bar{y}_B)^2 = \sum y_i^2 - \frac{1}{n} \left(\sum y_i \right)^2 \quad (3)$$

$$Q_{xy} = \sum (x_i - \bar{x}_B)(y_i - \bar{y}_B) = \sum x_i y_i - \frac{1}{n} \sum x_i \sum y_i. \quad (4)$$

Оценка коэффициента корреляции

$$r = \frac{Q_{xy}}{\sqrt{Q_x Q_y}}.$$

Выборочная прямая линия регрессии имеет уравнение

$$y = \tilde{\beta}_0 + \tilde{\beta}_1 x,$$

где

$$\tilde{\beta}_1 = \frac{Q_{xy}}{Q_x}, \quad \tilde{\beta}_0 = \bar{y}_B - \tilde{\beta}_1 \bar{x}_B.$$

Пример

Выборка значений случайного вектора.

x	y	x	y	x	y	x	y	x	y
8,35	3,50	10,50	6,00	11,35	9,50	12,15	6,00	12,85	9,50
8,74	1,49	10,75	2,50	11,50	6,00	12,25	8,05	13,15	9,02
9,25	6,40	10,76	5,74	11,50	9,00	12,35	5,01	13,25	6,49
9,50	4,50	11,00	8,50	11,62	8,50	12,50	7,03	13,26	10,50
9,75	5,00	11,00	5,26	11,75	10,00	12,76	7,53	13,40	7,51
10,24	7,00	11,25	8,00	12,00	9,00	12,85	6,01	13,50	10,00
13,65	9,50	14,50	10,00	13,75	8,51	14,75	12,00	14,00	11,00
15,25	12,50	14,23	8,40	16,00	11,50	14,26	10,00	16,00	13,00
14,51	9,50	16,25	12,00						

Найдем оценки параметров линии регрессии, коэффициента корреляции, построим график.

Объем выборки $n = 42$. Предварительно вычислим

$$\sum x_i = 522,23, \quad \sum y_i = 336,41 \quad \sum x_i^2 = 6652,25, \quad \sum y_i^2 = 2987,80, \quad \sum x_i y_i = 4358,626.$$

Тогда по формуле (1)

$$\bar{x}_B = 12,34, \quad \bar{y}_B = 8,011.$$

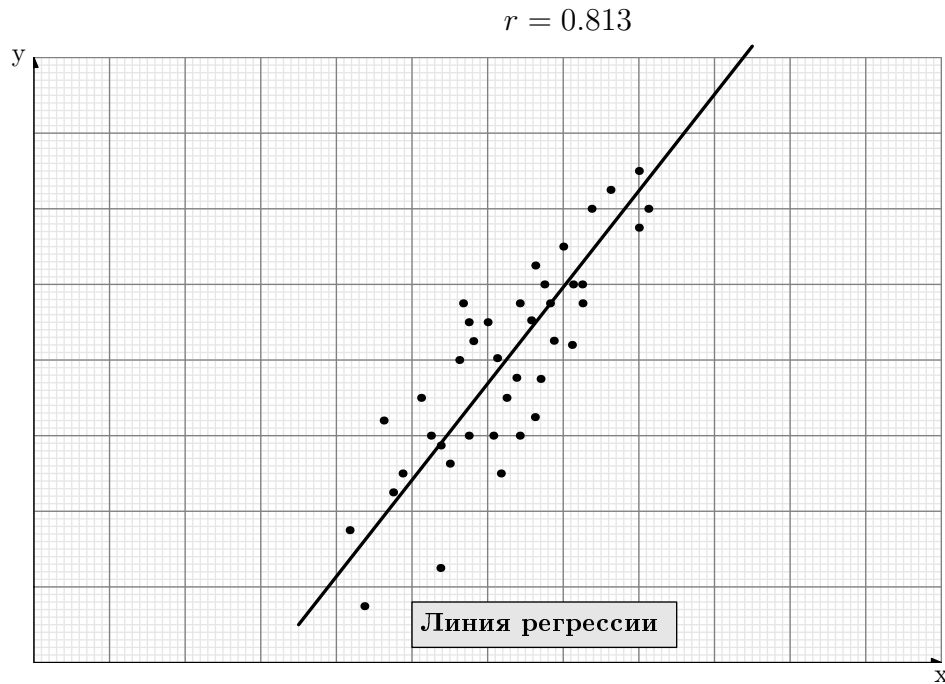
По формулам (2) находим

$$Q_x = 6552.25 - \frac{522.23^2}{42} = 158,3182; \quad (5)$$

$$Q_y = 2987.80 - \frac{336,41^2}{42} = 292,5958; \quad (6)$$

$$Q_{xy} = 4358,626 - \frac{522,23 \cdot 336,41}{42} = 175,19. \quad (7)$$

Окончательно из соотношений (15) получаем



Можно сделать вывод о довольно тесной корреляционной зависимости между величинами. Предполагая, что в нашем примере именно ξ влияет на η , мы можем используя уравнение регрессии количественно оценить это влияние.

Порядок оформления отчета

- Отчет оформляется на листах бумаги А4 и должен иметь стандартный титульный лист.
- Первая часть: Постановка задачи исследования, где приводится выборка.
- Вторая часть: Теоретический вывод параметров уравнения прямой линии регрессии МНК методом.
- Третья часть: Протокол выполнения практической части работы. При использовании (с разрешения преподавателя) специализированного ПО прилагается распечатка. Обязательно сделать чертеж, на котором на диаграмме рассеивания нарисовать линию регрессии.
- Сделать выводы по работе.