

Университет ИТМО

Факультет Программной Инженерии и Компьютерной Техники

Теория вероятностей

Лабораторная работа № 6

Вариант: **15**

Выполнил:
Чжун Цзяцзюнь
Группа: Р3210
Преподаватель:
Селина Елена Георгиевна

Санкт-Петербург
2025 г.

Содержание

1 Постановка задачи	2
2 Исходные данные	2
3 Решение	2
3.1 Расчетная таблица	2
3.2 Вычисление выборочных средних	3
3.3 Вычисление выборочных дисперсий	4
3.4 Вычисление корреляционного момента	4
3.5 Коэффициент корреляции	5
3.6 Уравнение регрессии	5
4 Графическое представление	6
5 Выводы	6

1 Постановка задачи

Дана таблица распределения 100 предприятий по производственным средствам X (тыс. ден. ед.) и по суточной выработке Y (тыс. единиц продукции). Известно, что между X и Y существует линейная корреляционная зависимость.

Требуется:

- а) найти уравнение прямой регрессии y на x ;
- б) построить уравнение эмпирической линии регрессии и случайные точки выборки (X, Y) .

2 Исходные данные

Таблица распределения для варианта 15:

XY	1200	2700	4200	6700	8200	9700	11200	12700	m_x
20	4	2	5	—	—	—	—	—	11
520	—	—	7	5	2	—	—	—	14
1020	—	—	—	9	14	6	—	—	29
1520	—	—	—	7	8	6	—	—	21
2020	—	—	—	—	4	5	7	—	16
2520	—	—	—	—	—	3	2	4	9
m_y	4	2	12	21	28	20	9	4	100

Таблица 1: Исходная таблица распределения

Где:

- X — производственные средства (тыс. ден. ед.)
- Y — суточная выработка (тыс. единиц продукции)
- m_x — частоты по строкам
- m_y — частоты по столбцам
- $n = 100$ — общее число наблюдений

3 Решение

3.1 Расчетная таблица

Для подсчета числовых характеристик (выборочных средних \bar{x} и \bar{y} , выборочных средних квадратичных отклонений s_x и s_y и выборочного корреляционного момента s_{xy}) составляем расчетную таблицу.

i	x_i	m_{x_i}	$m_{x_i}x_i$	$x_i^2m_{x_i}$	$\sum m_{ij}y_j$	y_1	y_2	\dots	y_8	$\sum m_{ij}x_iy_j$
1	20	11	220	4400	$4(1200) + 2(2700) + 5(4200) = 30600$	4	2	5	—	$4(20)(1200) + \dots$
2	520	14	7280	3785600	$7(4200) + 5(6700) + 2(8200) = 79800$	—	—	7	—	...
3	1020	29	29580	30171600	$9(6700) + 14(8200) + 6(9700) = 233100$	—	—	—	—	...
4	1520	21	31920	48518400	$7(6700) + 8(8200) + 6(9700) = 170700$	—	—	—	—	...
5	2020	16	32320	65286400	$4(8200) + 5(9700) + 7(11200) = 164000$	—	—	—	—	...
6	2520	9	22680	57153600	$3(9700) + 2(11200) + 4(12700) = 393900$	—	—	—	4	...
Σ	—	100	124000	204919600	1072100	—	—	—	—	...

Вычисления для столбца $\sum m_{ij}y_j$ по каждой строке:

$$\begin{aligned}
i = 1 : & 4 \cdot 1200 + 2 \cdot 2700 + 5 \cdot 4200 = 4800 + 5400 + 21000 = 31200 \\
i = 2 : & 7 \cdot 4200 + 5 \cdot 6700 + 2 \cdot 8200 = 29400 + 33500 + 16400 = 79300 \\
i = 3 : & 9 \cdot 6700 + 14 \cdot 8200 + 6 \cdot 9700 = 60300 + 114800 + 58200 = 233300 \\
i = 4 : & 7 \cdot 6700 + 8 \cdot 8200 + 6 \cdot 9700 = 46900 + 65600 + 58200 = 170700 \\
i = 5 : & 4 \cdot 8200 + 5 \cdot 9700 + 7 \cdot 11200 = 32800 + 48500 + 78400 = 159700 \\
i = 6 : & 3 \cdot 9700 + 2 \cdot 11200 + 4 \cdot 12700 = 29100 + 22400 + 50800 = 102300
\end{aligned}$$

$$\text{Сумма: } 31200 + 79300 + 233300 + 170700 + 159700 + 102300 = 776500$$

Проверка по столбцам:

$$\begin{aligned}
\sum m_y y_j &= 4 \cdot 1200 + 2 \cdot 2700 + 12 \cdot 4200 + 21 \cdot 6700 + 28 \cdot 8200 \\
&\quad + 20 \cdot 9700 + 9 \cdot 11200 + 4 \cdot 12700 \\
&= 4800 + 5400 + 50400 + 140700 + 229600 + 194000 + 100800 + 50800 \\
&= 776500
\end{aligned}$$

3.2 Вычисление выборочных средних

Выборочное среднее \bar{x} :

$$\begin{aligned}
\bar{x} &= \frac{\sum m_{x_i}x_i}{n} = \frac{20 \cdot 11 + 520 \cdot 14 + 1020 \cdot 29 + 1520 \cdot 21 + 2020 \cdot 16 + 2520 \cdot 9}{100} \\
\bar{x} &= \frac{220 + 7280 + 29580 + 31920 + 32320 + 22680}{100} = \frac{124000}{100} = 1240 \text{ тыс. ден. ед.}
\end{aligned}$$

Выборочное среднее \bar{y} :

$$\begin{aligned}
\bar{y} &= \frac{\sum m_{y_j}y_j}{n} \\
&= \frac{1200 \cdot 4 + 2700 \cdot 2 + 4200 \cdot 12 + 6700 \cdot 21 + 8200 \cdot 28 + 9700 \cdot 20 + 11200 \cdot 9 + 12700 \cdot 4}{100} \\
&= \frac{4800 + 5400 + 50400 + 140700 + 229600 + 194000 + 100800 + 50800}{100} \\
&= \frac{776500}{100} = 7765 \text{ тыс. ед.}
\end{aligned}$$

3.3 Вычисление выборочных дисперсий

Для вычисления выборочной дисперсии s_x^2 нужно найти $\sum m_{x_i} x_i^2$:

$$\begin{aligned}\sum m_{x_i} x_i^2 &= 11 \cdot 20^2 + 14 \cdot 520^2 + 29 \cdot 1020^2 + 21 \cdot 1520^2 + 16 \cdot 2020^2 + 9 \cdot 2520^2 \\ &= 11 \cdot 400 + 14 \cdot 270400 + 29 \cdot 1040400 + 21 \cdot 2310400 \\ &\quad + 16 \cdot 4080400 + 9 \cdot 6350400 \\ &= 4400 + 3785600 + 30171600 + 48518400 + 65286400 + 57153600 \\ &= 204920000\end{aligned}$$

Выборочная дисперсия s_x^2 :

$$\begin{aligned}s_x^2 &= \frac{1}{n-1} \left(\sum m_{x_i} x_i^2 - \frac{1}{n} \left(\sum m_{x_i} x_i \right)^2 \right) \\ s_x^2 &= \frac{1}{99} \left(204920000 - \frac{124000^2}{100} \right) = \frac{1}{99} (204920000 - 153760000) \\ s_x^2 &= \frac{51160000}{99} = 516767.68 \\ s_x &= \sqrt{516767.68} \approx 718.87 \text{ тыс. ден. ед.}\end{aligned}$$

Для вычисления s_y^2 нужно найти $\sum m_{y_j} y_j^2$:

$$\begin{aligned}\sum m_{y_j} y_j^2 &= 4 \cdot 1200^2 + 2 \cdot 2700^2 + 12 \cdot 4200^2 + 21 \cdot 6700^2 \\ &\quad + 28 \cdot 8200^2 + 20 \cdot 9700^2 + 9 \cdot 11200^2 + 4 \cdot 12700^2 \\ &= 4 \cdot 1440000 + 2 \cdot 7290000 + 12 \cdot 17640000 + 21 \cdot 44890000 \\ &\quad + 28 \cdot 67240000 + 20 \cdot 94090000 + 9 \cdot 125440000 + 4 \cdot 161290000 \\ &= 5760000 + 14580000 + 211680000 + 942690000 \\ &\quad + 1882720000 + 1881800000 + 1128960000 + 645160000 \\ &= 6713350000\end{aligned}$$

Выборочная дисперсия s_y^2 :

$$\begin{aligned}s_y^2 &= \frac{1}{99} \left(6713350000 - \frac{776500^2}{100} \right) = \frac{1}{99} (6713350000 - 602952250) \\ s_y^2 &= \frac{683827500}{99} = 6907348.49 \\ s_y &= \sqrt{6907348.49} \approx 2628.18 \text{ тыс. ед.}\end{aligned}$$

3.4 Вычисление корреляционного момента

Для вычисления s_{xy} нужно найти $\sum \sum m_{ij} x_i y_j$. Вычисляем для каждой ячейки с ненулевой частотой:

$$\begin{aligned}
\sum \sum m_{ij} x_i y_j = & 4(20)(1200) + 2(20)(2700) + 5(20)(4200) \\
& + 7(520)(4200) + 5(520)(6700) + 2(520)(8200) \\
& + 9(1020)(6700) + 14(1020)(8200) + 6(1020)(9700) \\
& + 7(1520)(6700) + 8(1520)(8200) + 6(1520)(9700) \\
& + 4(2020)(8200) + 5(2020)(9700) + 7(2020)(11200) \\
& + 3(2520)(9700) + 2(2520)(11200) + 4(2520)(12700)
\end{aligned}$$

Вычисляем по строкам:

$$\begin{aligned}
\text{Строка 1: } & 96000 + 108000 + 420000 = 624000 \\
\text{Строка 2: } & 15288000 + 17420000 + 8528000 = 41236000 \\
\text{Строка 3: } & 61506000 + 117096000 + 59418000 = 238020000 \\
\text{Строка 4: } & 71176000 + 99712000 + 88584000 = 259472000 \\
\text{Строка 5: } & 66256000 + 97970000 + 158368000 = 322594000 \\
\text{Строка 6: } & 73332000 + 56448000 + 128016000 = 257796000 \\
\text{Сумма: } & 1119742000
\end{aligned}$$

Корреляционный момент:

$$\begin{aligned}
s_{xy} &= \frac{1}{n-1} \left(\sum \sum m_{ij} x_i y_j - \frac{1}{n} \left(\sum m_{x_i} x_i \right) \left(\sum m_{y_j} y_j \right) \right) \\
s_{xy} &= \frac{1}{99} \left(1119742000 - \frac{124000 \cdot 776500}{100} \right) \\
s_{xy} &= \frac{1}{99} (1119742000 - 962860000) = \frac{156882000}{99} = 1584040.40
\end{aligned}$$

3.5 Коэффициент корреляции

$$r_{xy} = \frac{s_{xy}}{s_x \cdot s_y} = \frac{1584040.40}{718.87 \cdot 2628.18} = \frac{1584040.40}{1889332.89} \approx 0.8384$$

3.6 Уравнение регрессии

Уравнение эмпирической линии регрессии y на x имеет вид:

$$y - \bar{y} = r_{xy} \frac{s_y}{s_x} (x - \bar{x})$$

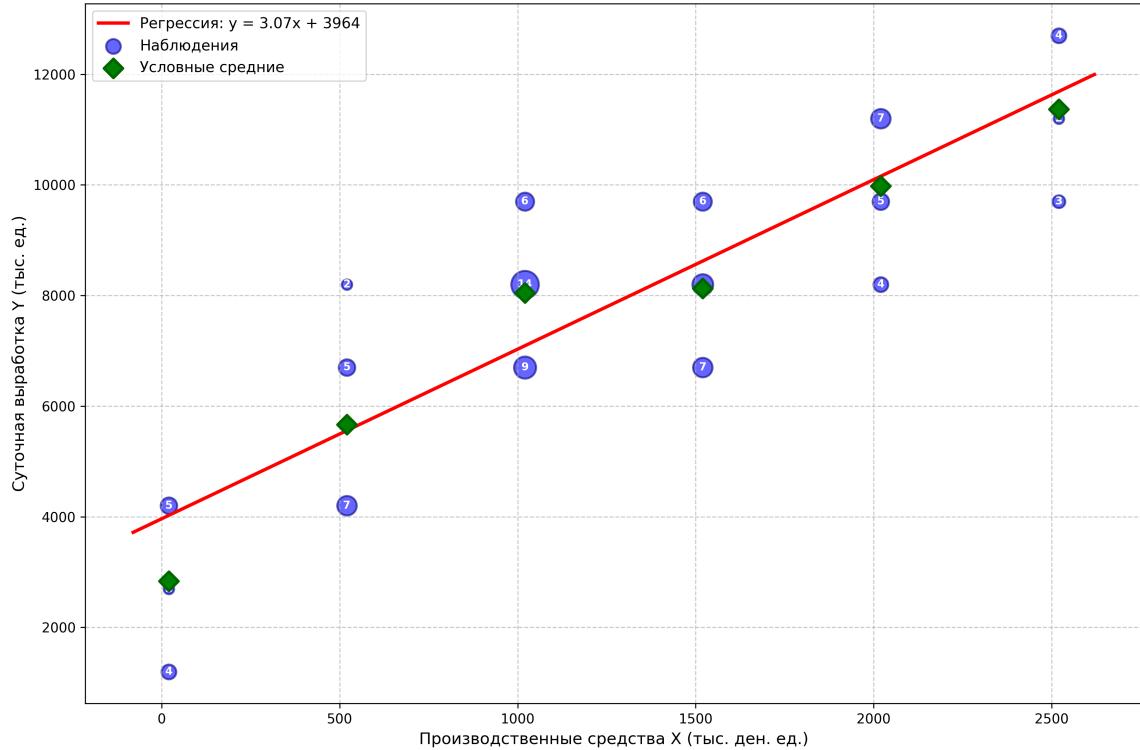
Подставляем значения:

$$\begin{aligned}
y - 7765 &= 0.8384 \cdot \frac{2628.18}{718.87} (x - 1240) \\
y - 7765 &= 0.8384 \cdot 3.6558(x - 1240) \\
y - 7765 &= 3.065(x - 1240) \\
y &= 3.065x - 3800.6 + 7765 \\
y &= 3.065x + 3964.4
\end{aligned}$$

Или в округленном виде:

$$y \approx 3.07x + 3964$$

4 Графическое представление



5 Выводы

- Получено уравнение регрессии: $y = 3.07x + 3964$
- Коэффициент корреляции $r_{xy} \approx 0.8384$ указывает на сильную положительную линейную связь между производственными средствами и суточной выработкой
- С увеличением производственных средств на 1 тыс. ден. ед. суточная выработка в среднем увеличивается примерно на 3.07 тыс. единиц продукции
- Высокое значение коэффициента корреляции (близкое к 1) свидетельствует о том, что линейная модель хорошо описывает зависимость между переменными