



Регрессионный анализ

Лектор: доцент Мирзаев С.С.

21 октября 2019 г.



План:


1. Парный регрессионный анализ
2. Метод наименьших квадратов
3. Множественный регрессионный анализ



Парный регрессионный анализ

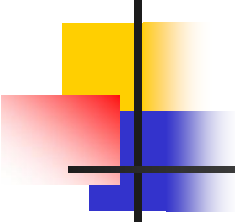
Корреляционный анализ показывает, что две переменные связаны друг с другом, однако он не дает представления о том, каким образом они связаны.

Для определения вида зависимости между переменными проводят регрессионный анализ



Рассмотрим более подробно
случай, когда одна переменная
зависит от другой. Рассмотрим
простейшую, т.е. линейную
модель

$$y = \alpha + \beta x + u$$



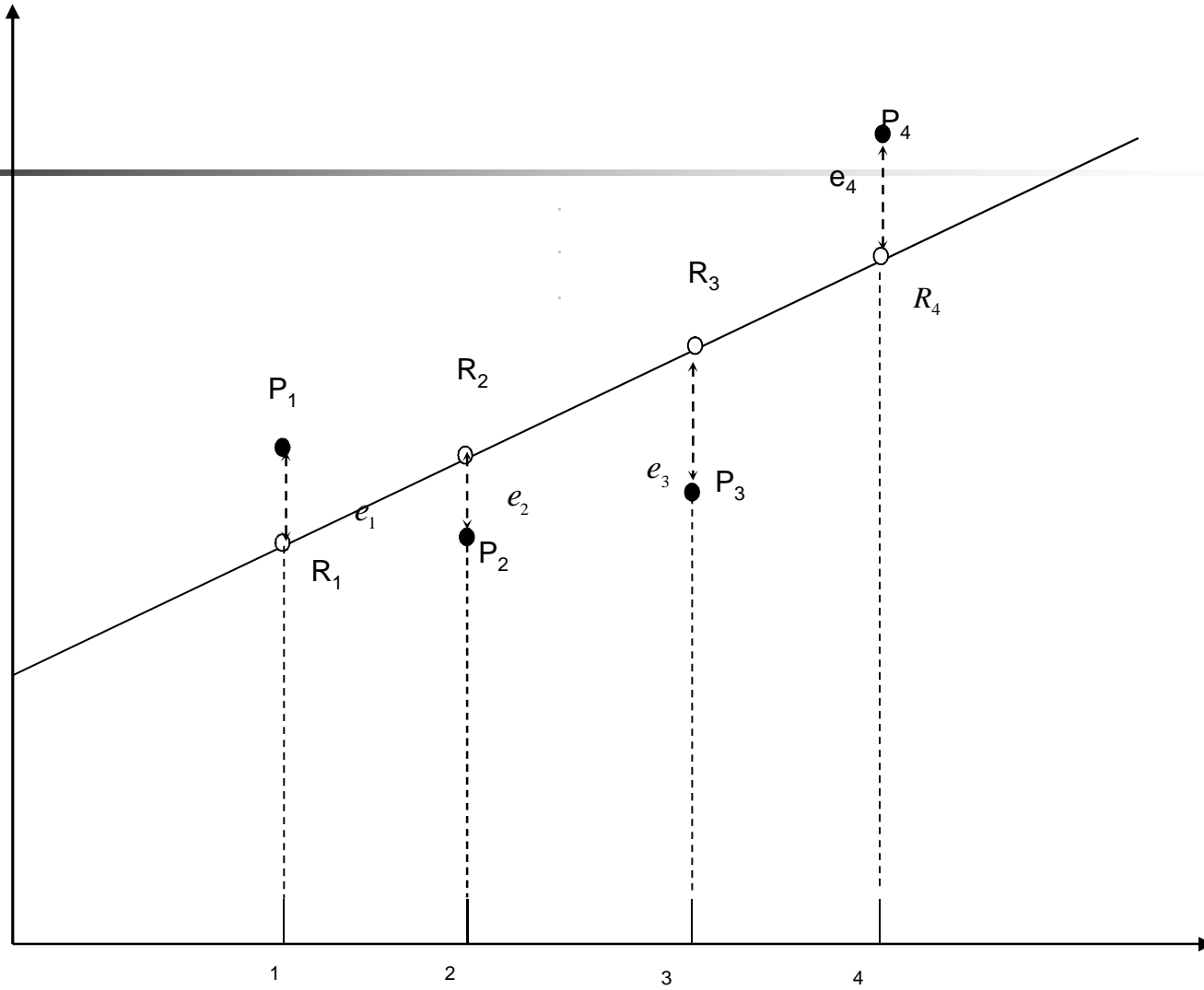
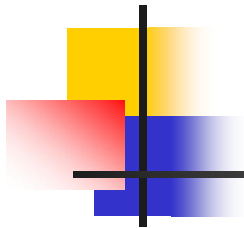
Оценка параметров по методу наименьших квадратов

Допустим, что мы имеем четыре наблюдения для X и Y .

Мы хотим построить линию регрессии таким образом, чтобы отклонения были минимальными

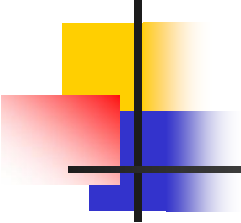
Уравнение этой прямой:

$$Y = a + bx$$



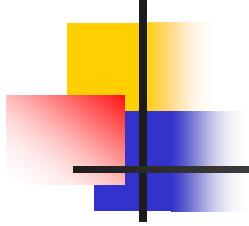
Одним из способов решения поставленной проблемы состоит в минимизации суммы квадратов отклонений

$$S(a,b) = \sum_i (y_i - Y_i)^2 = \sum_{i=1}^n (y_i - a - bx_i)^2 \rightarrow \min$$



Система нормальных уравнений

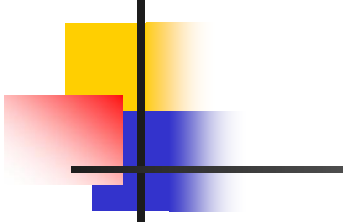
$$\begin{cases} \frac{\partial S}{\partial a} = -2 \sum_{i=1}^n (y_i - a - bx_i) = 0 \\ \frac{\partial S}{\partial b} = -2 \sum_{i=1}^n (y_i - a - bx_i)x_i = 0 \end{cases}$$



Решение системы

$$a = \frac{\sum x^2 \sum y - \sum x \sum xy}{n \sum x^2 - (\sum x)^2}$$
$$b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$$

Пример

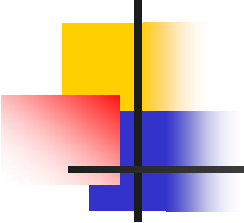


	(X1)	(X2)	(Y)
1	31	75	22
2	34	72	23
3	40	79	23
4	44	81	24
5	51	83	25
6	56	81	25
7	62	90	27
8	64	95	29
9	69	100	31
10	75	95	33
11	81	110	34
12	83	115	36
13	88	110	35
14	95	120	36
15	98	130	37

Получено следующее уравнение
регрессии:

$$Y = 0,92 + 0,29x$$

Полученный результат можно истолковать следующим образом. Если количество вносимых удобрений увеличить на 1 кг, то урожайность увеличится в среднем на 0,29 ц/га. Если $x = 0$, то прогнозируемый уровень урожайности равняется 0,92 ц/га



Множественный регрессионный анализ

Множественный регрессионный анализ является развитием парного регрессионного анализа применительно к случаям, когда зависимая переменная гипотетически связана с более чем одной независимой переменной

Рассмотрим пример, который исследован для парной зависимости.

Теперь исследуем зависимость урожайности от двух факторов: балла бонитета и количества вносимых удобрений. Получим следующее уравнение регрессии:

$$Y = 12,35 + 0,23x_1 + 0,02x_2$$

Полученное уравнение следует интерпретировать следующим образом. При каждом увеличении балла бонитета почвы на 1 балл урожайность увеличится на 0,23 ц/га. На каждую единицу увеличения вносимых удобрений урожайность увеличится на 0,02 ц/га. Коэффициент множественной корреляции $r = 0,81$, то есть связь между исследуемыми факторами и урожайностью хлопка очень высокая; коэффициент детерминации $D = 0,66$, то есть изменение значений урожайности на 66% зависит от исследуемых факторов, и на 34% – от других факторов, не включенных в модель