



Основные задачи статистики

Лектор: доцент Мирзаев С.С.

11 октября 2020 г.



План:

Введение

1. Корреляционные модели
2. Парный регрессионный анализ
3. Множественный регрессионный анализ



Определение эконометрики

Эконометрика – наука, исследующая количественные закономерности и взаимозависимости в экономике при помощи методов математической статистики



Основные задачи статистического анализа

- Выявление наличия или отсутствия взаимосвязи между изучаемыми факторами (корреляционный анализ)
- Определение вида взаимосвязи между изучаемыми факторами (регрессионный анализ)
- Проверка гипотезы о виде взаимосвязи между факторами



Функциональная, статистическая и корреляционная зависимости

- Две случайные величины могут быть связаны либо **функциональной** зависимостью, либо зависимостью другого рода, называемой **статистической**, либо быть **независимыми**.
- Функциональная зависимость: Если каждому значению случайной величины **X** соответствует одно возможное значение случайной величины **Y** , то **Y** называют функцией случайного аргумента **X** :

$$Y = \varphi(X)$$



Статистическая зависимость

- Статистической называют зависимость, при которой изменение одной из величин влечет изменение **распределения** другой
- В частности, статистическая зависимость проявляется в том, что при изменении одной из величин изменяется **среднее** значение другой; в этом случае статистическую зависимость называют **корреляционной**



Корреляционные модели

- Пусть изучается зависимость между факторами (X, Y) . В результате n независимых опытов получены n пар чисел:

$$(x_1; y_1), (x_2; y_2), \dots, (x_n; y_n)$$



Коэффициент ковариации

- Является мерой взаимосвязи между факторами:

$$\text{cov}(x, y) = \frac{1}{n} \sum \left(x_i - \bar{x} \right) \left(y_i - \bar{y} \right)$$

Тесноту связи между двумя

взаимозависимыми рядами

характеризует коэффициент линейной

корреляции, который показывает,

существует ли и насколько велика

связь изучаемых явлений

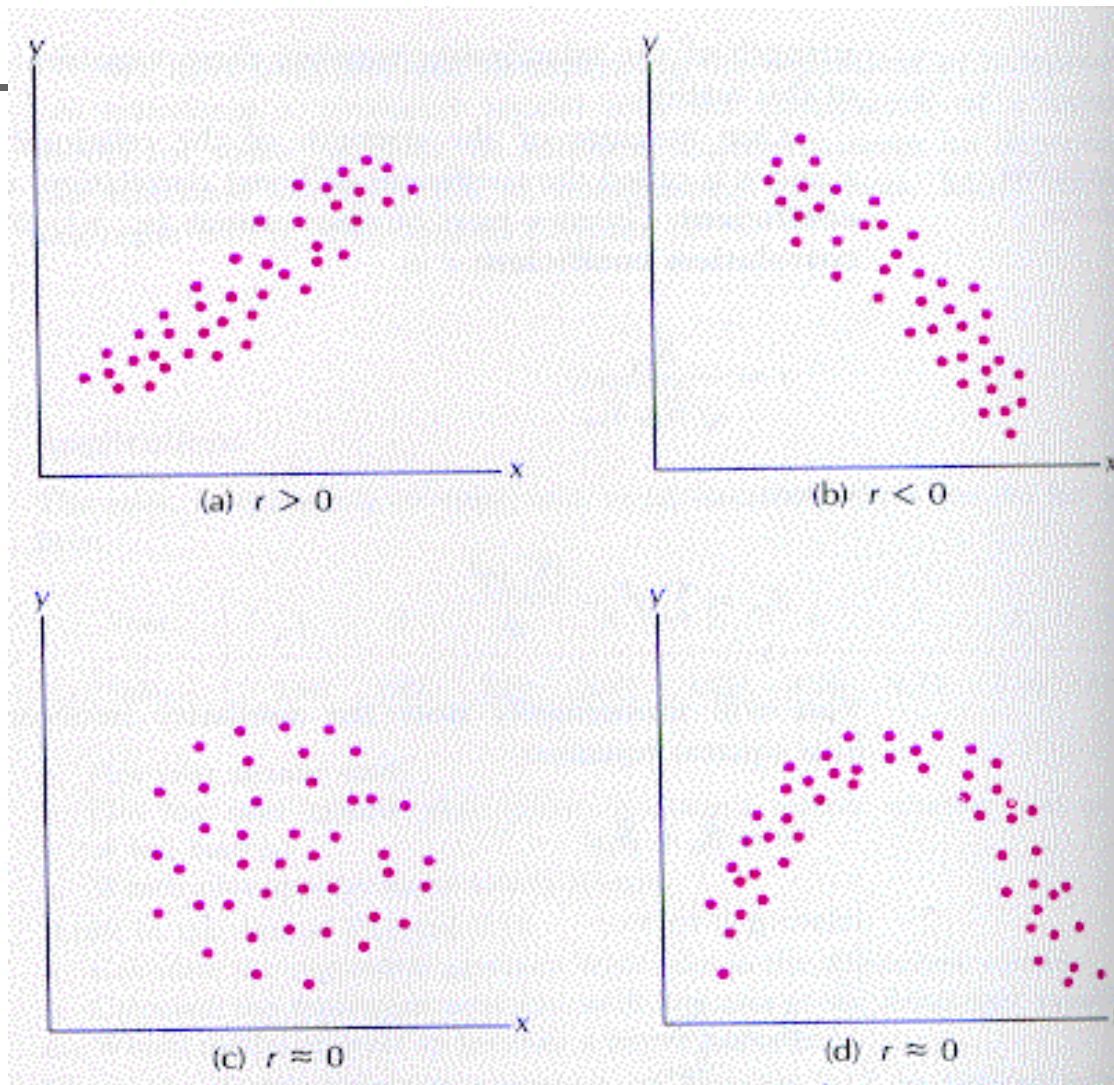
$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$



Коэффициент корреляции

- Коэффициент корреляции r может принимать значения от -1 до 1 . Если $r < 0$, то связь обратная, если $r > 0$, то связь прямая

Свойства r





Коэффициент корреляции

- Коэффициент корреляции указывает следующую степень связи:
- $0 \div \pm 0,15$ – отсутствие связи
- $\pm 0,16 \div \pm 0,20$ – плохая
- $\pm 0,21 \div \pm 0,30$ – слабая
- $\pm 0,31 \div \pm 0,40$ – умеренная
- $\pm 0,41 \div \pm 0,60$ – средняя
- $\pm 0,61 \div \pm 0,80$ – высокая
- $\pm 0,81 \div \pm 0,9$ – очень высокая
- $\pm 0,91 \div \pm 1,0$ – полная



Коэффициент детерминации

При анализе взаимосвязи между факторами также вычисляют коэффициент детерминации (r -квадрат)

$$D = r^2$$



Коэффициент детерминации

- Коэффициент детерминации показывает, какое влияние оказывают выбранные факторы на результирующий показатель
- Например, если $r = 0.9$ (r - коэффициент корреляции), то $D = 0.81$, то есть величина результирующего показателя на **81%** ($0.81 \cdot 100$) зависит от изменения исследуемых факторов

При анализе тесноты связи между переменной и несколькими факторами определяют коэффициент множественной корреляции по формуле


$$R = \sqrt{1 - \frac{\sum_i (y_i - \bar{Y}_i)^2}{\sum_i (y_i - \bar{y}_i)^2}}$$



Парный регрессионный анализ

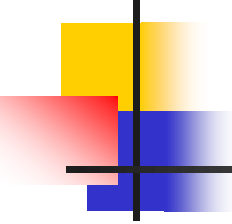
Коэффициент корреляции показывает, что две переменные связаны друг с другом, однако он не дает представления о том, каким образом они связаны.

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

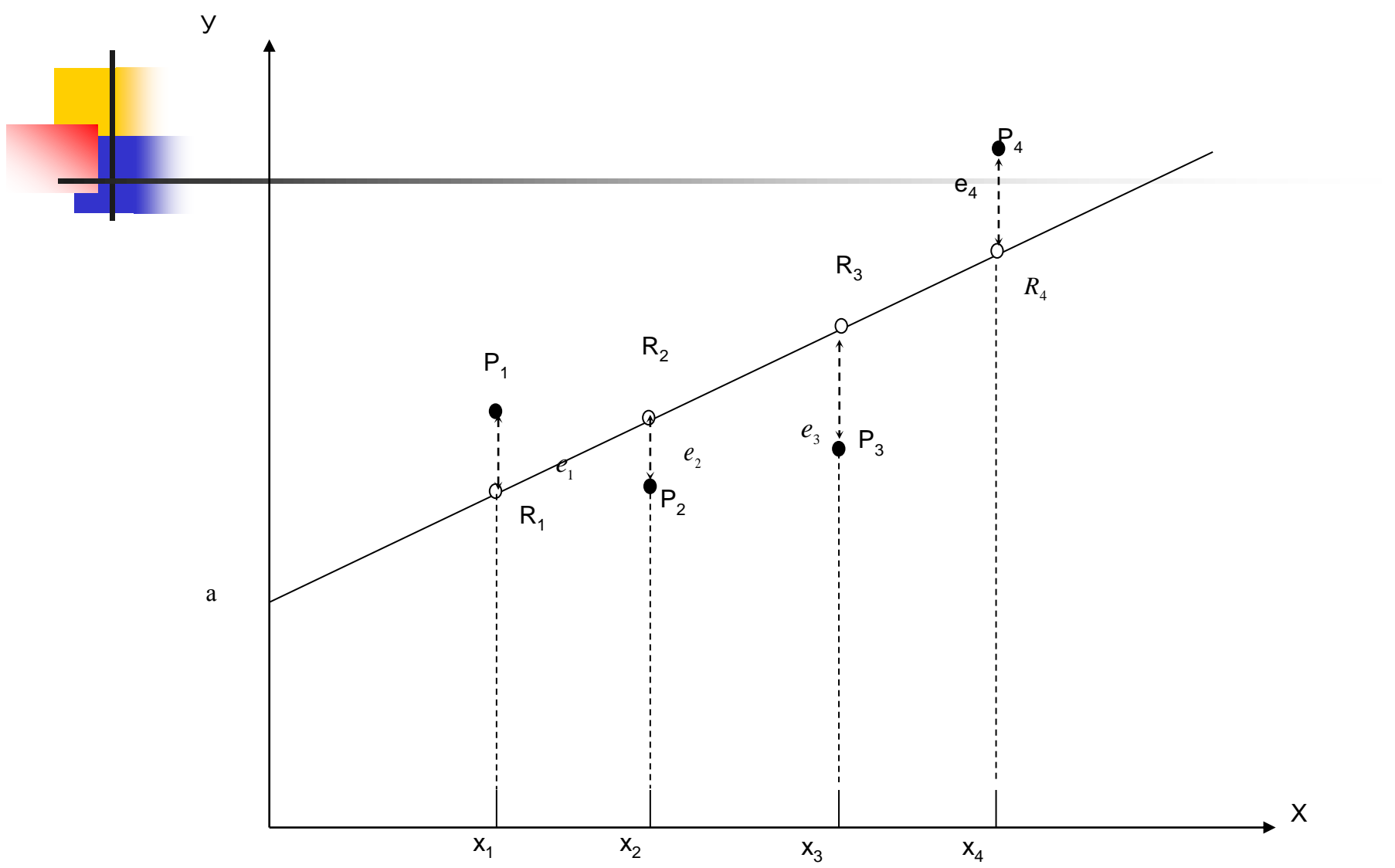
$$Y = \alpha + \beta x$$



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.



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

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



Система нормальных уравнений

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

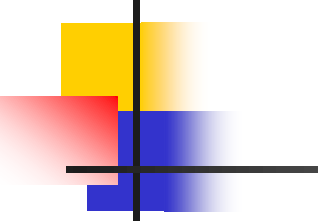


Решение системы

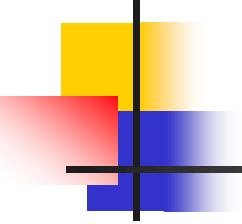
$$b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$$

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

Пример



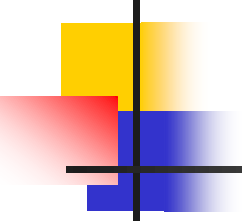
| № | Бонитет(X1) | Удобрения(X2) | Урожайность(Y) |
|----|-------------|---------------|----------------|
| 1 | 31 | 75 | 22 |
| 2 | 34 | 72 | 23 |
| 3 | 40 | 79 | 23 |
| 4 | 44 | 81 | 24 |
| 5 | 51 | 83 | 25 |
| 6 | 56 | 81 | 25 |
| 7 | 62 | 90 | 27 |
| 8 | 64 | 95 | 29 |
| 9 | 69 | 100 | 31 |
| 10 | 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:

$$\hat{y} = 1,39 + 0,28x$$

Полученный результат можно истолковать следующим образом. Если количество вносимых удобрений увеличить на 1 кг, то урожайность увеличится в среднем на 0,28 ц/га. Если $x = 0$, то прогнозируемый уровень урожайности равняется 1,39 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 (пример 2).

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

$$\hat{y} = 12,35 + 0,23x_1 + 0,02x_2$$

Полученное уравнение следует интерпретировать следующим образом. При каждом увеличении балла бонитета земли 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 $D = 0,7$, то есть изменение значений урожайности на 70% зависит от исследуемых факторов, и на 30% – от других факторов, не включенных в модель