

ВЗАИМОСВЯЗЬ ЯВЛЕНИЙ

Корреляция и регрессия

*The hardness of the butter
is proportional
to the softness of the bread*

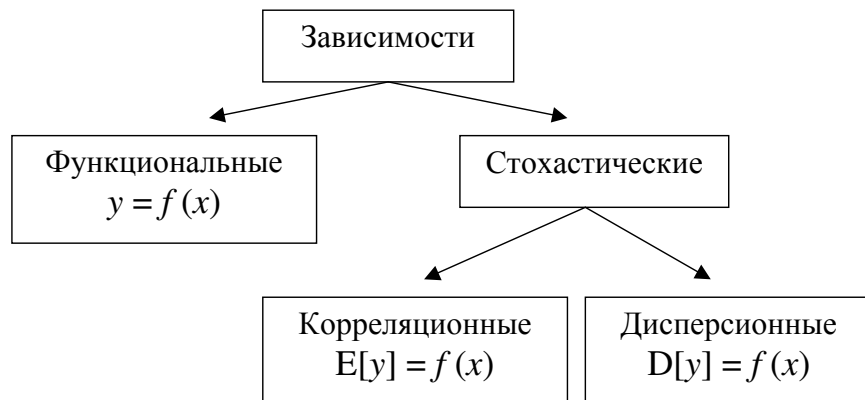
Статистические закономерности

- взаимосвязь явлений
- причинно-следственные отношения
- Изменение одного признака сопровождается изменением другого
 - причина и следствие
 - общая причина
 - одновременное изменение по разным причинам

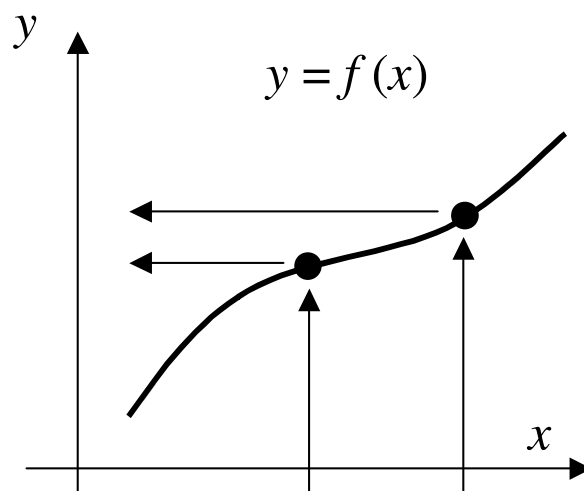
Признаки

- факторные и результативные
- независимая и зависимая переменная
- возможные причина и следствие
- аргумент и функция
- X и Y

Виды взаимосвязей



Функциональная зависимость



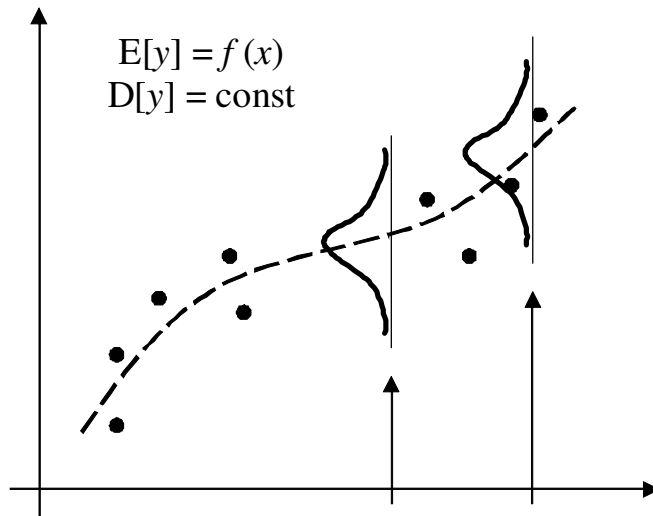
Функциональная зависимость

- Результативный признак Y – *функция* факторного признака X
- при изменении X изменяется *значение* Y
- точная, однозначная, полностью предсказуемая зависимость
- точки на графике расположены на линии

Стохастическая зависимость

- случайный, вероятностный, непредсказуемый
- точное значение Y предсказать невозможно
- при изменении X изменяются характеристики *распределения* Y
- зависимость на фоне случайных отклонений
- проявляется в среднем при большом числе данных как тенденция
- разброс точек вокруг линии
 - [Gr. $\sigma\tau\omicron\chi\alpha\sigma\tau\iota\kappa\omicron\varsigma$ – способный угадать, догадаться (найти ответ при нехватке данных)
 - *stokhasis* – догадка
 - $\sigma\tau\omicron\chi\omicron\varsigma$ – цель, мишень]

Корреляционная зависимость

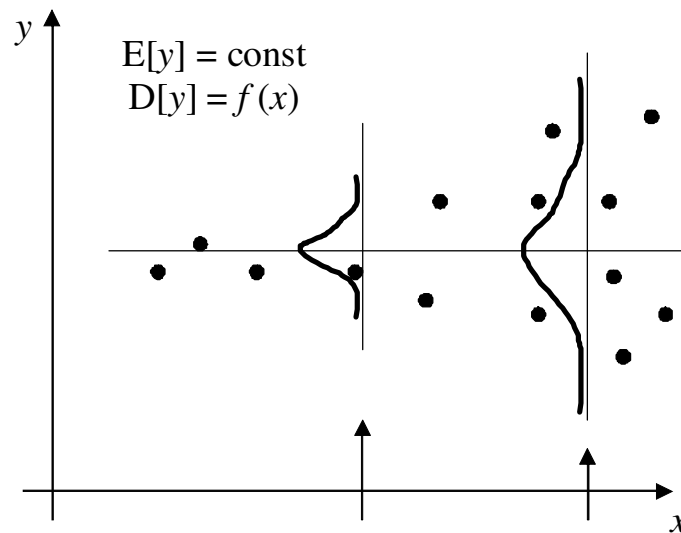


Корреляционная зависимость

- среднее значение результативного признака Y – функция факторного признака X
- при изменении X изменяется *среднее* значение Y
- условное среднее $\bar{y}_x = E[y | x]$

– [L. *correlatio* = *com* (вместе, взаимо-) + *relatio* (отношение, связь) = взаимосвязь, зависимость]
» реляционные БД

Дисперсионная зависимость



Направление связи

- | | |
|--|---|
| • прямая | • обратная |
| • положительная | • отрицательная |
| • с ростом фактора
результат растет | • с ростом фактора
результат уменьшается |
| • прямо
пропорционально | • обратно
пропорционально |

Форма связи

- | | |
|--|---|
| <ul style="list-style-type: none">• линейная<ul style="list-style-type: none">• X в первой степени• уравнение прямой (плоскости)• прямая линия на графике | <ul style="list-style-type: none">• нелинейная<ul style="list-style-type: none">• все остальные степени X• уравнение кривой• кривая на графике |
|--|---|

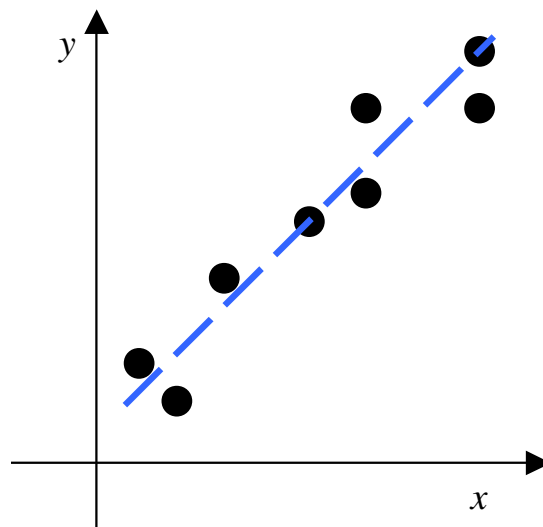
Методы изучения взаимосвязи

- качественный (визуальный) анализ
 - поле корреляции
- количественный анализ
 - изучение показателей связи
 - корреляционный анализ
 - построение уравнения связи
 - регрессионный анализ
 - дисперсионный анализ

Корреляционное поле

- поле корреляции
- диаграмма рассеяния
- диаграмма разброса
- графическое изображение двух признаков
 - увидеть всю картину
 - оценить вид зависимости

Корреляционное поле



Расположение точек

- «полоса, трубка, канал» вдоль линии
 - есть стохастическая связь
 - можно строить уравнение связи
- равномерно по всему полю
 - нет явной взаимосвязи
- различная концентрация точек
 - можно строить распределение
- центры группирования
 - несколько классов, видов единиц
 - статистические методы классификации

Корреляция

Корреляционный анализ

- количественное исследование тесноты связи, наличия корреляции, степени коррелированности переменных
 - Чем сильнее, теснее взаимосвязь, тем ближе точки расположены к линии

Коэффициент линейной корреляции

- Количественный показатель тесноты линейной связи

$$r_{yx} = \frac{n\sum yx - \sum x \sum y}{\sqrt{(n\sum x^2 - (\sum x)^2)(n\sum y^2 - (\sum y)^2)}}$$

- Принимает значения в диапазоне

$$-1 \leq r \leq +1$$

Коэффициент линейной корреляции

$$r_{yx} = \frac{\overline{(x - \bar{x}) \cdot (y - \bar{y})}}{\sigma_x \cdot \sigma_y}$$

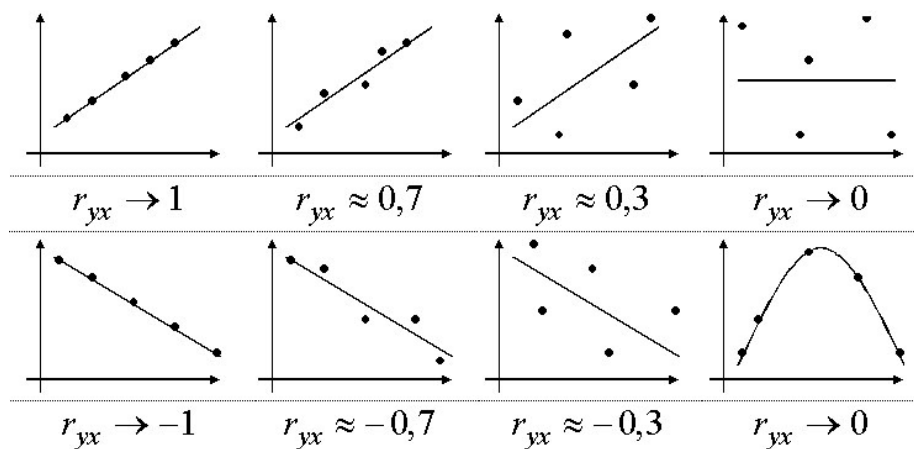
Знак коэффициента корреляции

- Знак говорит о направлении зависимости
 - положительная (прямая)
 - отрицательная (обратная)

Величина коэффициента корреляции

- Величина коэффициента корреляции говорит о тесноте связи:
 - $|r| = 1$ – функциональная связь, все точки лежат на прямой линии
 - $0 < |r| < 1$ – линейная зависимость на фоне случайных отклонений
 - $r = 0$
 - (1) взаимосвязь отсутствует
 - (2) взаимосвязь незаметна на фоне случайных отклонений
 - (3) связь есть, но она нелинейная

Примеры



Коэффициент корреляции рангов

ранг – порядковый номер после сортировки

- Коэффициент корреляции рангов Спирмена ρ
- Ранговый коэффициент корреляции
 - теснота нелинейной монотонной зависимости
 - возрастание или убывание результирующего признака при возрастании факторного

Коэффициент корреляции рангов Спирмена

$$\rho_{xy} = 1 - \frac{6 \cdot \sum d^2}{n \cdot (n^2 - 1)}$$

число «шесть»



$$-1 \leq \rho \leq +1$$

d – разность рангов

$$d_i = R(x_i) - R(y_i)$$

Пример

x	y	R_x	R_y	d	d^2
5	3	2	2	0	0
10	5	4	3	+1	1
7	9	3	4	-1	1
2	1	1	1	0	0

$$\rho_{xy} = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

$$\rho = 1 - \frac{6 \cdot 2}{4 \cdot 15} = \frac{4}{5} = 0,8$$

X	R_x
2	1
5	2
7	3
10	4

Регрессия

Регрессионный анализ

- Построение уравнения связи
 - уравнение регрессии
 - линия регрессии
- Регрессия (в статистике) – зависимость в среднем, «средняя зависимость»
 - [L. *regressus* = *re-* (назад) + *gradi* (идти), *gradus* – шаг, ступень]
 - » (1) возвращение, отступление, движение назад
 - » (2) противоположность развитию, прогрессу

Ф.Гальтон: Регрессия

<http://galton.org>

- взаимосвязь между ростом родителей и взрослых детей
- «тенденция регрессии»: возвращение к среднему уровню
 - у высоких родителей дети ниже, чем родители
 - у родителей низкого роста дети выше, чем родители

В НАСТОЯЩЕЕ ВРЕМЯ

- уравнение связи – «функция регрессии»
- в статистике *регрессия* **НЕ ОЗНАЧАЕТ** «уменьшение, возвращение, движение назад»!!!

Возвращение к среднему значению: рост человека определяется наследственностью

ANTHROPOLOGICAL MISCELLANEA.

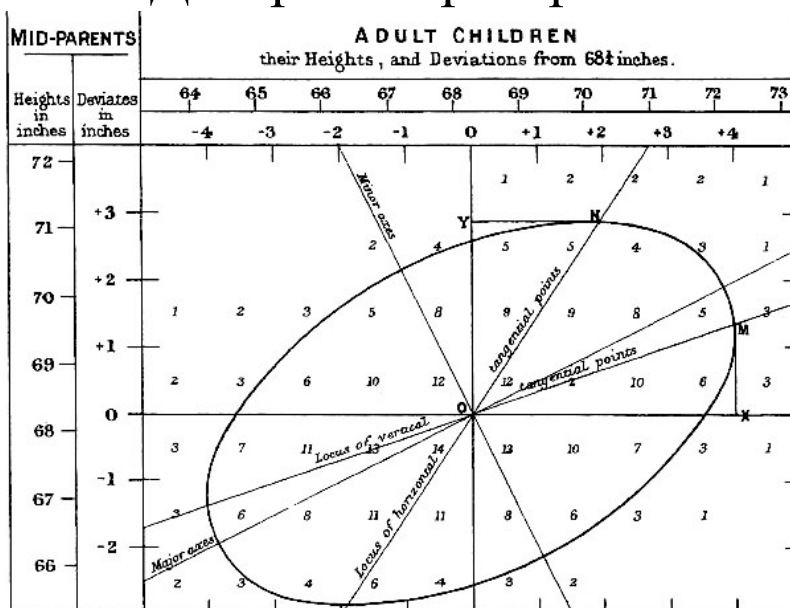
REGRESSION *towards* MEDIOCRITY in HEREDITARY STATURE.

By FRANCIS GALTON, F.R.S., &c.

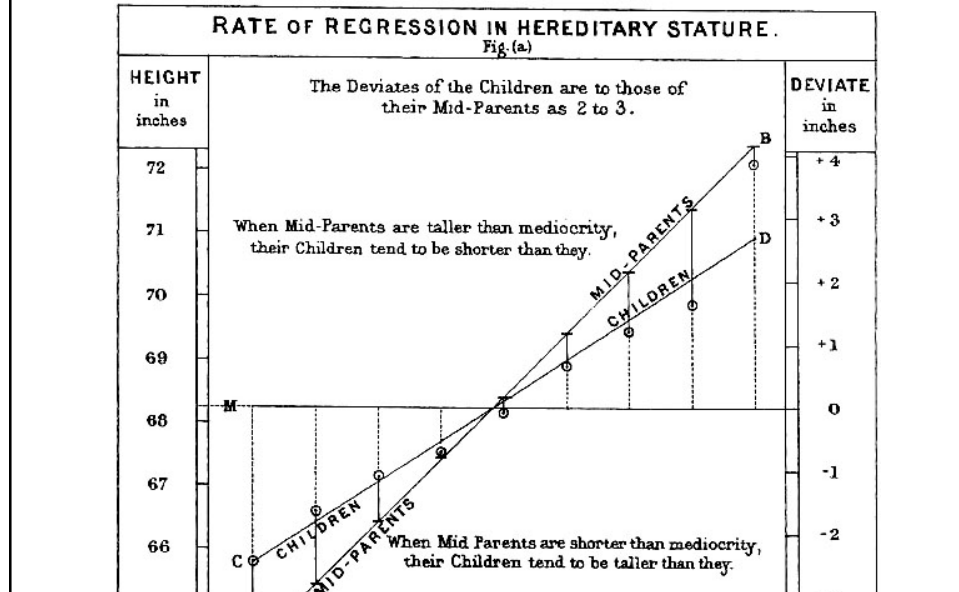
[WITH PLATES IX AND X.]

THIS memoir contains the data upon which the remarks on the Law of Regression were founded, that I made in my Presidential Address to Section H, at Aberdeen. That address, which will appear in due course in the Journal of the British Association, has already been published in "Nature," September 24th. I reproduce here

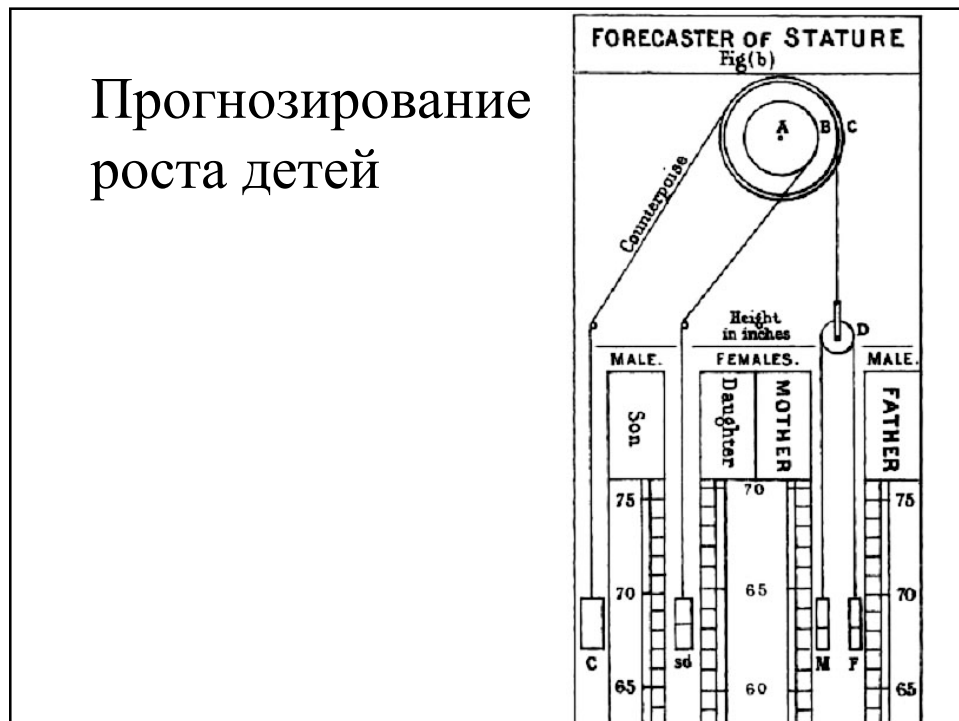
Диаграмма разброса



Скорость возвращения: наклон линии регрессии



Прогнозирование роста детей



Парная линейная регрессия

– взаимосвязь двух признаков (пары)

$$y = a + b \cdot x$$

a – свободный член

b – коэффициент регрессии

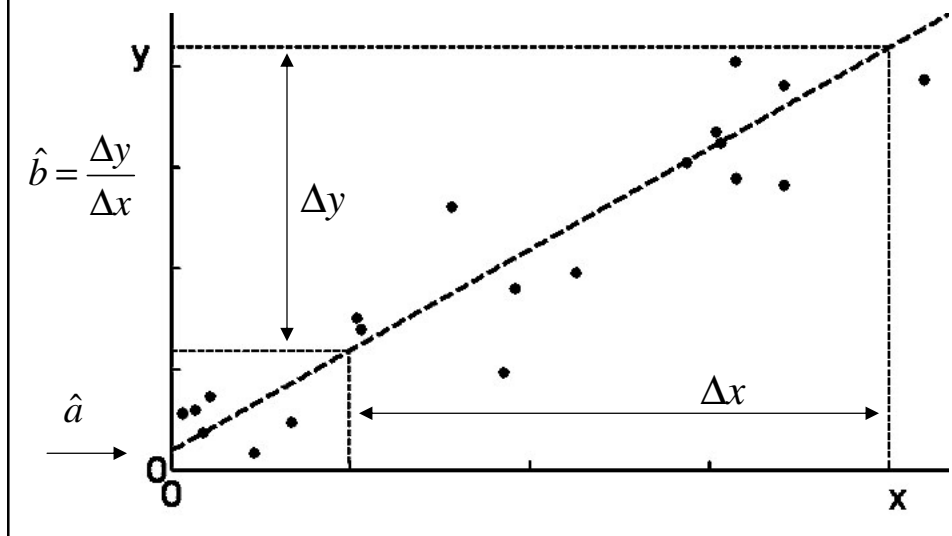
- Цель регрессионного анализа – найти вид уравнения и коэффициенты

«регрессия Y на X »

Объем выборки

- Чтобы найти два коэффициента, нужно два уравнения
- Для изучения зависимости «в среднем», требуется гораздо больше данных
 - объем выборки «на порядок» больше числа параметров уравнения связи
 - «на порядок» = в десять раз
 - объем выборки для построения уравнения парной линейной регрессии: $n = ?$

Графический метод



Оценки коэффициентов

Выбираем две точки на линии регрессии

$$\{x_1, y_1\} \quad \text{и} \quad \{x_2, y_2\}$$

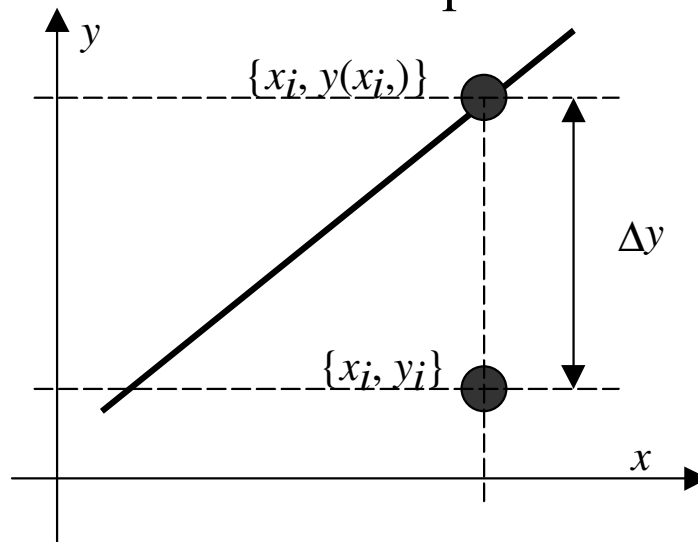
$$\begin{cases} y_1 = a + b \cdot x_1 \\ y_2 = a + b \cdot x_2 \end{cases} \Rightarrow \hat{a}, \hat{b}$$

Метод наименьших квадратов (МНК)

- сумма *квадратов* отклонений должна быть *наименьшей*
 - найти такие параметры модели, которые минимизируют сумму квадратов отклонений эмпирических значений результативного признака от теоретических в соответствии с уравнением регрессии

$$\sum \Delta y^2 \rightarrow \min$$

Расстояние по вертикали



Отклонение от линии регрессии

- Отклонение Δy – разница между фактическим и теоретическим значениями результативного признака
 - расстояние от точки до линии
 - по вертикали $y = f(x)$
 - по горизонтали $x = f(y)$
 - по диагонали (диагональная регрессия)
 - перпендикулярно к линии (ортогональная регрессия)
 - решение зависит от выбранного критерия

Регрессия Y на X : Постановка задачи

Дано: выборка $\{x_i, y_i\}, i = 1..n$

Найти: коэффициенты уравнения регрессии

$$y(x) = a + b \cdot x$$

такие, что

$$\sum_{i=1}^n (\Delta y_i)^2 \rightarrow \min$$

Оценки коэффициентов

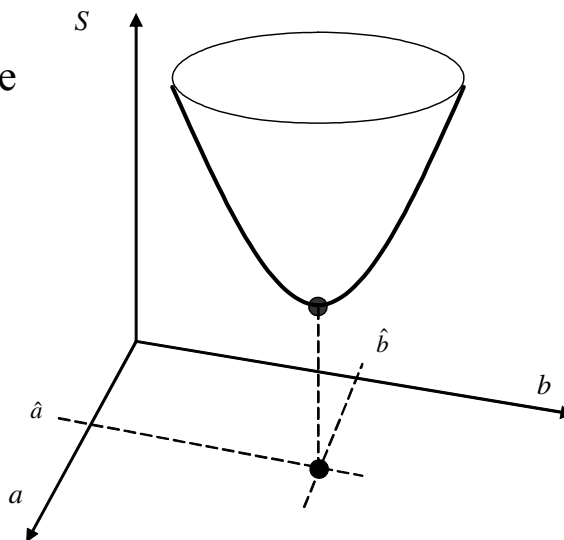
- Объем выборки n во много раз больше, чем число неизвестных коэффициентов
 - переопределенная система уравнений
 - число уравнений больше числа неизвестных
 - точного решения нет
 - можно найти только *приближённое* решение
 - оценки коэффициентов a и b

$$\hat{a} \quad \hat{b}$$

Квадратичный критерий

- Единственное решение

$$\begin{cases} \frac{\partial S}{\partial a} = 0 \\ \frac{\partial S}{\partial b} = 0 \end{cases}$$



- «Правильная» постановка = половина решения задачи

Производная сложной функции

- производная внешней функции по внутренней * на производную внутренней функции по аргументу

$$\frac{df(g(x))}{dx} = \frac{df}{dg} \cdot \frac{dg}{dx} = \frac{df}{dx}$$

Частные производные

$$\begin{aligned}\frac{\partial S}{\partial a} &= \frac{\partial}{\partial a} \left(\sum (y_i - a - bx_i)^2 \right) = \\ &= \sum \left(\frac{\partial}{\partial a} \left((y_i - a - bx_i)^2 \right) \right) = \\ &= \sum \left(\frac{\partial}{\partial (y_i - a - bx_i)} \left((y_i - a - bx_i)^2 \right) \frac{\partial}{\partial a} (y_i - a - bx_i) \right) = \\ &= \sum (2 \cdot (y_i - a - bx_i) \cdot (-1))\end{aligned}$$

Частные производные

$$\begin{aligned}\frac{\partial S}{\partial a} &= 0 \\ (-2) \sum (y_i - a - bx_i) &= 0 \\ \sum (y_i - a - bx_i) &= 0 \\ \sum y - \sum a - \sum bx &= 0 \\ \sum y &= an + b \sum x\end{aligned}$$

Частные производные

$$\begin{aligned}\frac{\partial S}{\partial b} &= \frac{\partial}{\partial b} \left(\sum (y_i - a - bx_i)^2 \right) = \\ &= \sum (2 \cdot (y_i - a - bx_i) \cdot (-x_i))\end{aligned}$$

Частные производные

$$\begin{aligned}\frac{\partial S}{\partial b} &= 0 \\ -2 \sum (yx - ax - bx^2) &= 0 \\ \sum yx - \sum ax - \sum bx^2 &= 0 \\ \sum yx &= a \sum x + b \sum x^2\end{aligned}$$

Система нормальных уравнений

- После преобразований получаем два уравнения с двумя неизвестными

$$\begin{cases} an + b\Sigma x = \Sigma y \\ a\Sigma x + b\Sigma x^2 = \Sigma yx \end{cases}$$

- Система имеет единственное *точное* решение – *оценки* коэффициентов

$$\hat{a} \quad \hat{b} \Rightarrow y = \hat{a} + \hat{b}x$$

Правило запоминания

- система нормальных уравнений без вывода

$$y = a + b \cdot x$$

- 1) берем уравнение регрессии и ставим знак суммы

$$\Sigma a + \Sigma bx = \Sigma y \qquad an + b\Sigma x = \Sigma y$$

- 2) берем уравнение регрессии, умножаем на x и ставим знак суммы

$$\Sigma ax + \Sigma bx^2 = \Sigma yx \qquad a\Sigma x + b\Sigma x^2 = \Sigma yx$$

Задача

- система нормальных уравнений без вывода

$$y = a + b \cdot x$$

- 1) берем уравнение регрессии и ставим знак суммы
- 2) берем уравнение регрессии, умножаем на x и ставим знак суммы

Решение через определители

$$\begin{cases} an + b\Sigma x = \Sigma y \\ a\Sigma x + b\Sigma x^2 = \Sigma yx \end{cases} \quad \Delta = \begin{vmatrix} n & \Sigma x \\ \Sigma x & \Sigma x^2 \end{vmatrix}$$

$$\Delta_a = \begin{vmatrix} \Sigma y & \Sigma x \\ \Sigma yx & \Sigma x^2 \end{vmatrix} \quad \Delta_b = \begin{vmatrix} n & \Sigma y \\ \Sigma x & \Sigma yx \end{vmatrix}$$

$$a = \frac{\Delta_a}{\Delta}$$

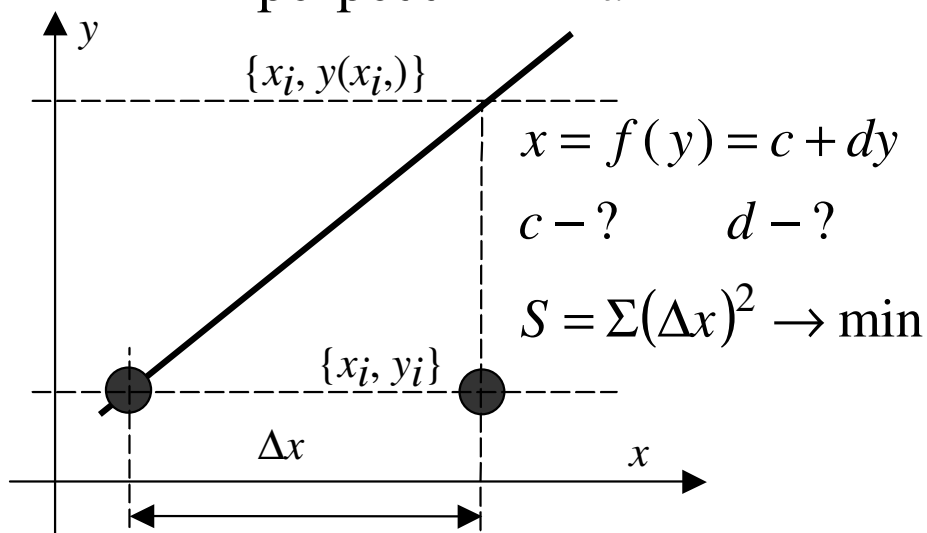
$$b = \frac{\Delta_b}{\Delta}$$

Решение

$$a = \frac{\Delta_a}{\Delta} = \frac{\Sigma y \Sigma x^2 - \Sigma yx \Sigma x}{(\Sigma x)^2 - n \Sigma x^2}$$

$$b = \frac{\Delta_b}{\Delta} = \frac{n \Sigma yx - \Sigma y \Sigma x}{(\Sigma x)^2 - n \Sigma x^2}$$

Расстояние по горизонтали:
регрессия X на Y



Задача

- система нормальных уравнений без вывода

$$x = c + d \cdot y$$

- 1) берем уравнение регрессии и ставим знак суммы
- 2) берем уравнение регрессии, умножаем на y и ставим знак суммы

Система нормальных уравнений

$$x = c + dy$$

$$\begin{cases} cn + d\Sigma y = \Sigma x \\ c\Sigma y + d\Sigma y^2 = \Sigma yx \end{cases}$$

$$\hat{c} \quad \hat{d} \Rightarrow x = \hat{c} + \hat{d}y$$

Решение

$$c = \frac{\Delta_c}{\Delta} = \frac{\Sigma x \Sigma y^2 - \Sigma y x \Sigma y}{(\Sigma y)^2 - n \Sigma y^2}$$

$$d = \frac{\Delta_d}{\Delta} = \frac{n \Sigma y x - \Sigma y \Sigma x}{(\Sigma y)^2 - n \Sigma y^2}$$

Коэффициент линейной корреляции

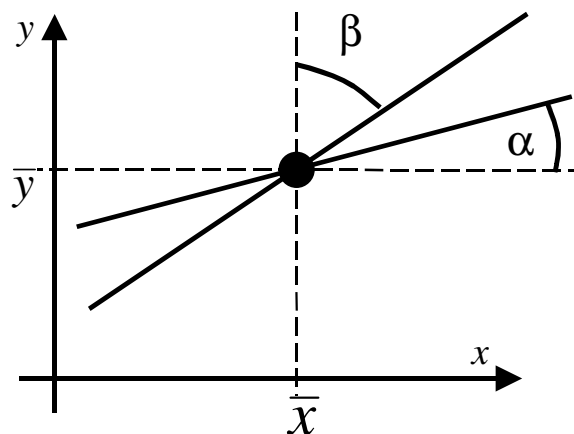
связан с коэффициентами регрессии

$$r_{yx} = \frac{n \Sigma y x - \Sigma x \Sigma y}{\sqrt{(n \Sigma x^2 - (\Sigma x)^2)(n \Sigma y^2 - (\Sigma y)^2)}}$$

$$r^2 = b \cdot d$$

Коэффициент корреляции

$$r^2 = b \cdot d = \operatorname{tg} \alpha \cdot \operatorname{tg} \beta$$



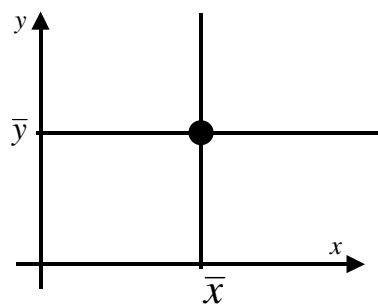
Отсутствие линейной связи

- При отсутствии взаимосвязи линия регрессии проходит горизонтально:
 - при изменении значения фактора результат *в среднем* остается постоянным.

Крайние случаи

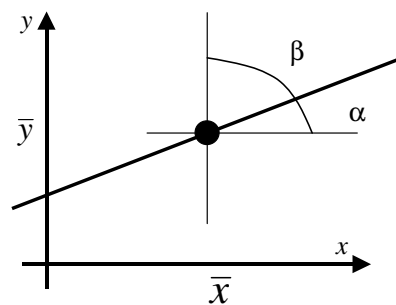
$$\alpha = 0^\circ \quad \beta = 0^\circ$$

$$r^2 = 0 \cdot 0 = 0$$



$$\beta = 90^\circ - \alpha$$

$$\operatorname{tg} \alpha = \frac{1}{\operatorname{tg} \beta} \Rightarrow r^2 = 1$$



Геометрическая интерпретация

Коэффициент корреляции r показывает

- насколько различается наклон двух линий регрессии: $y(x)$ и $x(y)$
- насколько разные результаты даёт минимизация отклонений по x и по y

Линии регрессии по МНК

Линии $y = f(x)$ и $x = f(y)$

пересекаются в точке $\{\bar{x}, \bar{y}\}$

$$\begin{cases} an + b\Sigma x = \Sigma y \\ a\Sigma x + b\Sigma x^2 = \Sigma yx \end{cases} \quad \begin{cases} cn + d\Sigma y = \Sigma x \\ c\Sigma y + d\Sigma y^2 = \Sigma yx \end{cases}$$

Разделим первые уравнения на n :

$$a + b\bar{x} = \bar{y} \qquad c + d\bar{y} = \bar{x}$$

Дисперсия

$$D_x = \frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n} \right)^2 = \overline{x^2} - \bar{x}^2$$

средний квадрат минус квадрат среднего

Система нормальных уравнений: Матричный способ

Система уравнений = одно матричное уравнение

$$\begin{cases} a + bx_1 = y_1 \\ a + bx_2 = y_2 \\ \dots \\ a + bx_n = y_n \end{cases} \quad \begin{pmatrix} a + bx_1 \\ a + bx_2 \\ \dots \\ a + bx_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}$$

$$\mathbf{Z} = \mathbf{Y}$$

Система нормальных уравнений: Матричный способ

Произведение матриц:

сумма попарных произведений

«строка умножается на столбец»

$$\begin{pmatrix} a + bx_1 \\ a + bx_2 \\ \dots \\ a + bx_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} \quad \begin{pmatrix} 1 & | & x_1 \\ 1 & | & x_2 \\ \dots & | & \dots \\ 1 & | & x_n \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}$$

$$\mathbf{R} \cdot \mathbf{A} = \mathbf{Y}$$

«Псевдо-обращение» матрицы

$$\mathbf{R} \cdot \mathbf{A} = \mathbf{Y}$$

Умножим слева на транспонированную матрицу регрессоров $\mathbf{R}^T$

$$\mathbf{R}^T \cdot \mathbf{R} \cdot \mathbf{A} = \mathbf{R}^T \cdot \mathbf{Y}$$

Получаем квадратную матрицу: $\mathbf{R}^T \cdot \mathbf{R}$

Решение системы: $\mathbf{A} = (\mathbf{R}^T \cdot \mathbf{R})^{-1} \cdot (\mathbf{R}^T \cdot \mathbf{Y})$

Задача

Вывести систему нормальных уравнений

$$\mathbf{R} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \dots & \dots \\ 1 & x_n \end{pmatrix}$$

$$\mathbf{A} = \begin{pmatrix} a \\ b \end{pmatrix}$$

$$\mathbf{R} \cdot \mathbf{A} = ?$$

$$\mathbf{R}^T = ?$$

$$\mathbf{R}^T \cdot \mathbf{R} = ?$$

$$(\mathbf{R}^T \cdot \mathbf{R}) \cdot \mathbf{A} = ?$$

$$\mathbf{R}^T \cdot \mathbf{Y} = ?$$

$$\mathbf{R}^T \cdot \mathbf{R} \cdot \mathbf{A} = \mathbf{R}^T \cdot \mathbf{Y}$$

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}$$

Транспонирование матрицы

«Перестановка» элементов матрицы

Зеркальное отражение относительно главной диагонали

Е. *transpose* = **L**. *trans*- (через-, пере-) + *pose* (ставить) = переставить, поменять местами

$$\mathbf{R} = \left(\begin{array}{c|c} 1 & x_1 \\ 1 & x_2 \\ \dots & \dots \\ 1 & x_n \end{array} \right) \quad \mathbf{R}^T = \left(\begin{array}{cccc} 1 & 1 & \dots & 1 \\ \hline x_1 & x_2 & \dots & x_n \end{array} \right)$$

Произведение матриц

Сумма попарных произведений: «строка на столбец»

$$\mathbf{R}^T \cdot \mathbf{R} = ?$$

$$\begin{aligned} \mathbf{R}^T \cdot \mathbf{R} &= \left(\begin{array}{cccc} 1 & 1 & \dots & 1 \\ \hline x_1 & x_2 & \dots & x_n \end{array} \right) \cdot \left(\begin{array}{c|c} 1 & x_1 \\ 1 & x_2 \\ \dots & \dots \\ 1 & x_n \end{array} \right) = \\ &= \left(\begin{array}{c|c} 1+1+\dots+1 & x_1+x_2+\dots+x_n \\ \hline x_1+x_2+\dots+x_n & x_1^2+x_2^2+\dots+x_n^2 \end{array} \right) = \begin{pmatrix} n & \Sigma x \\ \Sigma x & \Sigma x^2 \end{pmatrix} \end{aligned}$$

Произведение матриц

Сумма попарных произведений: «строка на столбец»

$$(\mathbf{R}^T \cdot \mathbf{R}) \cdot \mathbf{A} = ?$$

$$\mathbf{R}^T \cdot \mathbf{R} \cdot \mathbf{A} = \begin{pmatrix} n & \Sigma x \\ \Sigma x & \Sigma x^2 \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a \cdot n + b \cdot \Sigma x \\ a \cdot \Sigma x + b \cdot \Sigma x^2 \end{pmatrix}$$

Произведение матриц

Сумма попарных произведений: «строка на столбец»

$$\mathbf{R}^T \cdot \mathbf{Y} = ?$$

$$\mathbf{R}^T \cdot \mathbf{Y} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} \Sigma y \\ \Sigma(yx) \end{pmatrix}$$

Система нормальных уравнений

$$\mathbf{R}^T \cdot \mathbf{R} \cdot \mathbf{A} = \mathbf{R}^T \cdot \mathbf{Y}$$

$$\mathbf{R}^T \cdot \mathbf{R} \cdot \mathbf{A} = \begin{pmatrix} \frac{a \cdot n + b \cdot \Sigma x}{a \cdot \Sigma x + b \cdot \Sigma x^2} \end{pmatrix} \quad \mathbf{R}^T \cdot \mathbf{Y} = \begin{pmatrix} \frac{\Sigma y}{\Sigma(yx)} \end{pmatrix}$$

$$\begin{pmatrix} \frac{a \cdot n + b \cdot \Sigma x}{a \cdot \Sigma x + b \cdot \Sigma x^2} \end{pmatrix} = \begin{pmatrix} \frac{\Sigma y}{\Sigma(yx)} \end{pmatrix} \quad \begin{cases} an + b\Sigma x = \Sigma y \\ a\Sigma x + b\Sigma x^2 = \Sigma yx \end{cases}$$

Задача

Вывести систему нормальных уравнений

$$\mathbf{R} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \dots & \dots \\ 1 & x_n \end{pmatrix}$$

$$\mathbf{A} = \begin{pmatrix} a \\ b \end{pmatrix}$$

$$\mathbf{R} \cdot \mathbf{A} = ?$$

$$\mathbf{R}^T = ?$$

$$\mathbf{R}^T \cdot \mathbf{R} = ?$$

$$(\mathbf{R}^T \cdot \mathbf{R}) \cdot \mathbf{A} = ?$$

$$\mathbf{R}^T \cdot \mathbf{Y} = ?$$

$$\mathbf{R}^T \cdot \mathbf{R} \cdot \mathbf{A} = \mathbf{R}^T \cdot \mathbf{Y}$$

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}$$

Множественная регрессия

- многофакторная регрессия

$$y = a + b_1 x_1 + b_2 x_2 + \dots + b_k x_k$$

$$y = a \cdot 1 + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_k \cdot x_k$$

$$R: \quad 1 \quad x_1 \quad x_2 \quad x_k$$

- «регрессоры»: то, на что умножаются коэффициенты уравнения линейной регрессии

Правило запоминания

система нормальных уравнений без вывода

- берем уравнение регрессии
- умножаем на регрессора
- ставим знак суммы
- повторяем со следующим регрессором

Пример: Множественная
линейная регрессия

$$y = a + bx + cz$$

$$R: \begin{array}{ccc} 1 & x & z \end{array}$$

$$\begin{cases} \Sigma \dots = \dots \\ \Sigma \dots = \dots \\ \Sigma \dots = \dots \end{cases}$$

Пример: Множественная
линейная регрессия

$$y = a + bx + cz$$

$$R: \begin{array}{ccc} 1 & x & z \end{array}$$

$$\begin{cases} \Sigma y = an + b\Sigma x + c\Sigma z \\ \Sigma yx = a\Sigma x + b\Sigma x^2 + c\Sigma zx \\ \Sigma yz = a\Sigma z + b\Sigma xz + c\Sigma z^2 \end{cases}$$

Задача

$$y = ax + b + cz$$

$$\begin{cases} \Sigma \dots = \dots \\ \Sigma \dots = \dots \\ \Sigma \dots = \dots \end{cases}$$

Пример: Парная
нелинейная регрессия

$$y = a + bx + cx^2$$

$$R: \quad 1 \quad x \quad x^2$$

$$\begin{cases} \Sigma \dots = \dots \\ \Sigma \dots = \dots \\ \Sigma \dots = \dots \end{cases}$$

Пример: Парная
нелинейная регрессия

$$y = a + bx + cx^2$$

$$R: \quad 1 \quad x \quad x^2$$

$$\begin{cases} \Sigma y = an + b\Sigma x + c\Sigma x^2 \\ \Sigma yx = a\Sigma x + b\Sigma x^2 + c\Sigma x^3 \\ \Sigma yx^2 = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 \end{cases}$$

Задача

$$y = a + bx^2 + c \lg x$$

$$\begin{cases} \Sigma \dots = \dots \\ \Sigma \dots = \dots \\ \Sigma \dots = \dots \end{cases}$$

Домашнее задание



Кластерный анализ

