ЛЕКЦИЯ 7

Теория вероятностей и математическая статистика

Числовые характеристики двумерных и многомерных случайных величин

Характеристики двумерной случайной величины

- **Характеристики двумерной случайной величины (ξ, η) – это характеристики одномерных величин ξ и η, и характеристики связи между ними. Дальше мы будем рассматривать именно статистическую связь, которая называется *корреляцией*. Вначале рассмотрим линейную связь и ее характеристики – *ковариацию*, *коэффициент корреляции*, *уравнение линейной регрессии*, *остаточную дисперсию.***

Ковариация

- **Определение.** *Ковариацией* случайной величины (ξ, η) называется центральный смешанный момент второго порядка

$K_{\xi,\eta} = \mathrm{cov}(\xi, \eta) = M[(\xi - M\xi)\cdot(\eta - M\eta)].$

Ковариация есть мера линейной зависимости между ξ, η.

Ковариация

Величины ξ,η называются

- *некоррелированными* при $\mathrm{cov}(\xi, \eta) = 0$,
- *положительно коррелированными* при $\mathrm{cov}(\xi, \eta) > 0$,
- *отрицательно коррелированными* при $\mathrm{cov}(\xi, \eta) < 0$.

Для вычисления ковариации часто используют формулу

- $\mathrm{cov}(\xi, \eta) = M(\xi\cdot\eta) - M\,\xi\cdot M\,\eta.$

Коэффициент корреляции

- **<u>Определение.</u>** Коэффициентом корреляции между случайными
- величинами ξ, η называется число

$$\boxed{\rho_{\xi,\eta} = \frac{\operatorname{cov}(\xi,\eta)}{\sigma_{\xi}\cdot\sigma_{\eta}}}$$

Свойства коэффициента корреляции

- 1. $|\rho_{\xi\eta}| \le 1$.
- 2. Если ξ,η независимы, то $\rho_{\xi\eta} = 0$.
- Если $|\rho_{\xi\eta}| = 1$, то ξ, η линейно зависимы,

то есть существуют такие a и b, что

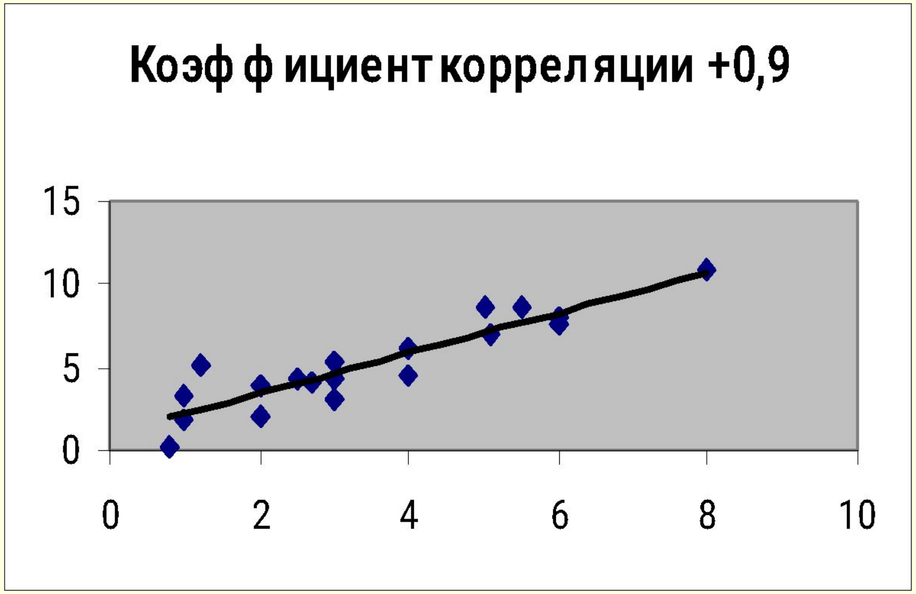
$\xi = a\eta + b$.

Смысл коэффициента корреляции

- Коэффициент корреляции есть мера линейной зависимости между ξ, η.
- Его модуль указывает на силу линейной связи

  (чем ближе к 1, тем сильнее),
- а знак указывает на направление связи.

Пример: ρ = +0,9

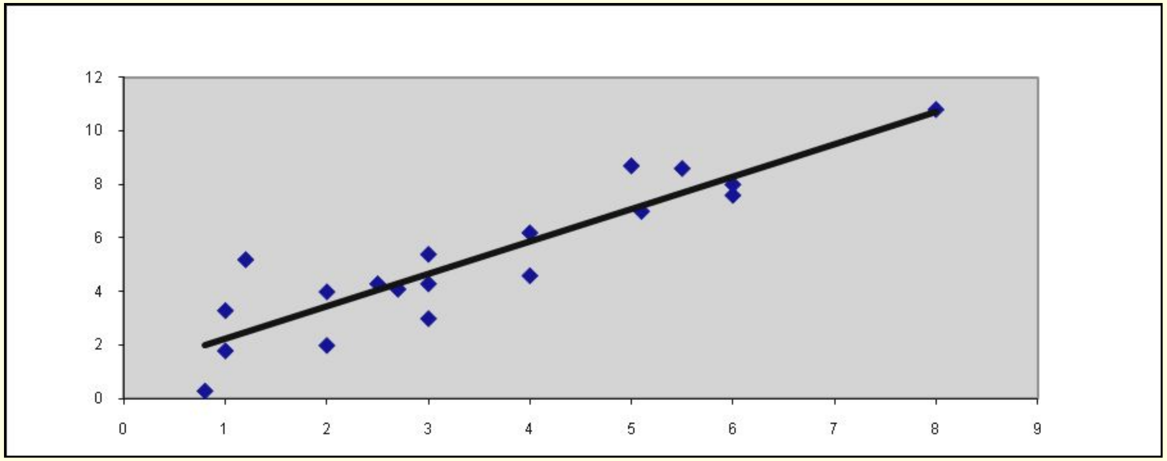


Пример : ρ = +0,2

Пример: ρ = – 0,6

Линейная зависимость

- Проблема: найти функцию, описывающую линейную зависимость (уравнение прямой).



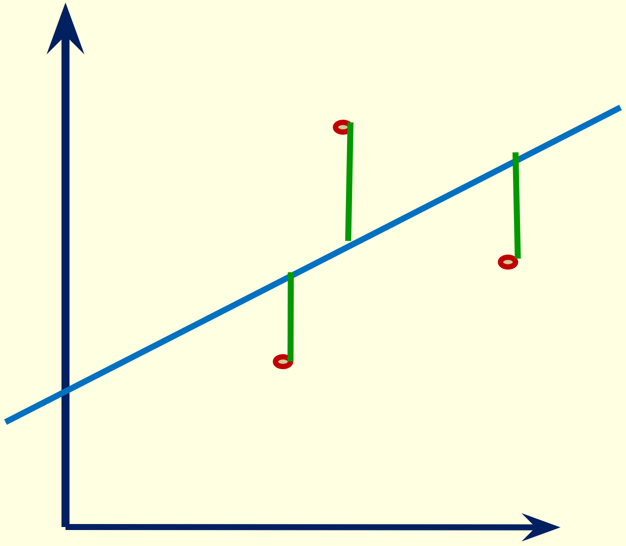
Уравнение линейной регрессии

- **Определение.** Уравнением линейной регрессии η на ξ называется уравнение

$\hat{\eta} = a\xi + b$, параметры которого минимизируют остаточную дисперсию

$$S^2_{ост} = M(\eta - \hat{\eta})^2 = M(\eta - (a\xi + b))^2.$$

Смысл. Уравнение линейной регрессии η на ξ выражает линейную зависимость η от ξ.



- Надо найти минимум остаточной дисперсии
- $S^2_{ост} = M\,(\eta - \hat{\eta})^2$

Нахождение коэффициентов уравнения линейной регрессии

$S^2_{ост} = M[\eta - (a\xi + b)]^2 =$

$M[(\eta - M\eta) - a(\xi - M\xi) + (M\eta - aM\xi - b)]^2 =$

$M(\eta - M\eta)^2 + a^2M(\xi - M\xi)^2 + M[(M\eta - aM\xi - b)]^2 -$

$2aM[(\eta - M\eta)(\xi - M\xi)] + 2M[(\eta - M\eta)(M\eta - aM\xi - b)] -$

$2aM[(\xi - M\xi)(M\eta - aM\xi - b)].$

- $(M\eta - aM\xi - b)$ – постоянная величина, ее можно вынести за знак матожидания.
- $M(\eta - M\eta) = M\eta - M[M\eta] = M\eta - M\eta = 0$,
- $M(\xi - M\xi) = 0$
- Подставляя, получаем:

$$\mathbf{S^2_{ост}} = M(\eta - M\eta)^2 + a^2M(\xi - M\xi)^2 +$$

$$+ (M\eta - aM\xi - b)^2 - 2aM[(\eta - M\eta)(\xi - M\xi)].$$

- Поскольку $M(\eta - M\eta)^2 = D\eta = \sigma^2{}_{\eta}$,

$$M(\xi - M\,\xi)^2 = D\xi = \sigma^2{}_{\xi},$$

$M[(\xi - M\xi)\cdot(\eta - M\eta)] = \mathrm{cov}(\xi,\eta) = \rho\sigma_{\xi}\sigma_{\eta}$, то

$\mathbf{S^2{}_{ост}} = \sigma^2{}_{\eta} + a^2\sigma^2{}_{\xi} + (M\eta - aM\xi - b)^2 - 2a\,\rho\sigma_{\xi}\sigma_{\eta}.$

$\mathbf{S^2_{ост}}$ – функция переменных a и b, надо найти min $\mathbf{S^2_{ост}}$, то есть найти значения

a и b, при которых достигается минимум.

Найдем производные от $S^2_{ост}$ по a и b.

$$S^2_{ост} = \sigma^2_\eta + a^2\sigma^2_\xi + (M\eta - aM\xi - b)^2 - 2a\,\rho\sigma_\xi\sigma_\eta.$$

$$(S^2_{ост})'_b = -2(M\eta - aM\xi - b) = 0$$

$$(S^2_{ост})'_a = 2a\sigma^2_\xi - 2M\xi\,(M\eta - aM\xi - b) -$$

$$-2\rho\sigma_\xi\sigma_\eta = 0$$

Из первого уравнения находим:

$$b = M\eta - aM\xi.$$

Подставляя во второе, получаем:

$$a = \rho\cdot\sigma_\eta/\sigma_\xi.$$

Подставим

$$\mathbf{a = \rho \cdot \sigma_\eta/\sigma_\xi, \quad b = M\eta - aM\xi}$$

В уравнение $\hat{\eta} = a\xi + b$.

Получим:

$\hat{\eta} = \rho \cdot \sigma_\eta/\sigma_\xi \cdot \xi + M\eta - \rho \cdot \sigma_\eta/\sigma_\xi \cdot M\xi$, или

$$\boxed{\hat{\eta} - M\eta = \rho \frac{\sigma_\eta}{\sigma_\xi}(\xi - M\xi).}$$

Замечание

- Коэффициент уравнения линейной регрессии $\boldsymbol{\rho \cdot \sigma_\eta / \sigma_\xi}$ можно записать в виде:
- $\boldsymbol{\rho \cdot \sigma_\eta / \sigma_\xi = \mathrm{cov}(\xi, \eta) / \sigma^2_\xi}$.
- Тогда уравнение линейной регрессии примет вид:

$$\boxed{\hat{\eta} - M\eta = \frac{\mathrm{cov}(\xi, \eta)}{\sigma^2_\xi}(\xi - M\xi).}$$

Остаточная дисперсия

Найдем значение $\mathbf{S^2_{ост} = M(\eta - \hat{\eta})^2 = M(\eta - (a\xi+b))^2}$. Для этого подставим полученные значения a и b.

$S^2_{ост} = M(\eta - (a\xi + b))^2 = M(\eta - (a\xi + b))^2 =$

$M[\eta - M\eta - \rho\cdot\sigma_\eta/\sigma_\xi(\xi - M\xi)]^2 = M(\eta - M\eta)^2 +$

$(\rho\cdot\sigma_\eta/\sigma_\xi)^2 M(\xi - M\xi)]^2 - 2\rho\cdot\sigma_\eta/\sigma_\xi M[(\xi - M\xi)\cdot$

$(\eta - M\eta)] = \sigma^2_\eta + (\rho\cdot\sigma_\eta/\sigma_\xi)^2\sigma^2_\xi -$

$2\rho\cdot\sigma_\eta/\sigma_\xi \cdot \rho\sigma_\xi\sigma_\eta =$

Остаточная дисперсия

- $\sigma^2_\eta + (\rho \cdot \sigma_\eta)^2 - 2\,\rho^2 \cdot \sigma_\eta{}^2 = \sigma^2_\eta - \rho^2 \cdot \sigma_\eta{}^2 =$

$$= \boldsymbol{\sigma^2_\eta\,(1 - \rho^2)}.$$

Смысл: остаточная дисперсия выражает ошибку приближения при замене η на $\hat{\eta} = a\xi + b$.

Пример

- **Дискретная двумерная случайная величина (X,Y) задана таблицей распределения:**

Y \ X	0	1	2
-1	0,1	0	0,2
0	0,1	0,2	0,1
3	0,3	0	0

Пример

- Найдем одномерные законы распределения:

X	**0**	**1**	**2**
P	0,5	0,2	0,3
Y	**-1**	**0**	**3**
P	0,3	0,4	0,3

Пример

- Вычислим числовые характеристики.
- $MX = 0 \cdot 0{,}5 + 1 \cdot 0{,}2 + 2 \cdot 0{,}3 = 0{,}8$.
- $DX = 0^2 \cdot 0{,}5 + 1^2 \cdot 0{,}2 + 2^2 \cdot 0{,}3 - 0{,}8^2 = 0{,}76$.
- $MY = (-1) \cdot 0{,}3 + 0 \cdot 0{,}4 + 3 \cdot 0{,}3 = 0{,}6$.
- $DY = (-1)^2 \cdot 0{,}3 + 0^2 \cdot 0{,}4 + 3^2 \cdot 0{,}3 - 0{,}6^2 = 2{,}64$.
- $M(XY) = (-1) \cdot 2 \cdot 0{,}2 = -0{,}4$.

Пример

- Найдем ковариацию:
- $\operatorname{cov}(\xi, \eta) = M(\xi \cdot \eta) – M\xi \cdot M\eta$.
- В наших обозначениях
- $\operatorname{cov}(X, Y) = M(X \cdot Y) – MX \cdot MY$.
- $\operatorname{cov}(X,Y) = – 0,4 – 0,8 \cdot 0,6 = – 0,88$.
- Величины X,Y отрицательно коррелированы.

Коэффициент корреляции

$$\boxed{\rho_{\xi,\eta} = \frac{\operatorname{cov}(\xi,\eta)}{\sigma_\xi \cdot \sigma_\eta}}$$

$$\rho_{XY} = \frac{-0,88}{\sqrt{0,76 \cdot 2,64}} \approx -0,64$$

Уравнение линейной регрессии

$$\hat{Y} - MY = \frac{\operatorname{cov}(X,Y)}{\sigma^2{}_X}(X - MX).$$

Запишем уравнение линейной регрессии Y на X.

Подставим MX = 0,8, DX = 0,76, MY = 0,6. cov(X,Y) = – 0,88.

Yˆ – 0,6 = – 0,88/0,76·(X – 0,8).

Остаточная дисперсия

- $\hat{Y} - 0,6 = -1,16(X - 0,8).$

$\underline{\hat{Y} = -1,16X + 1,53.}$

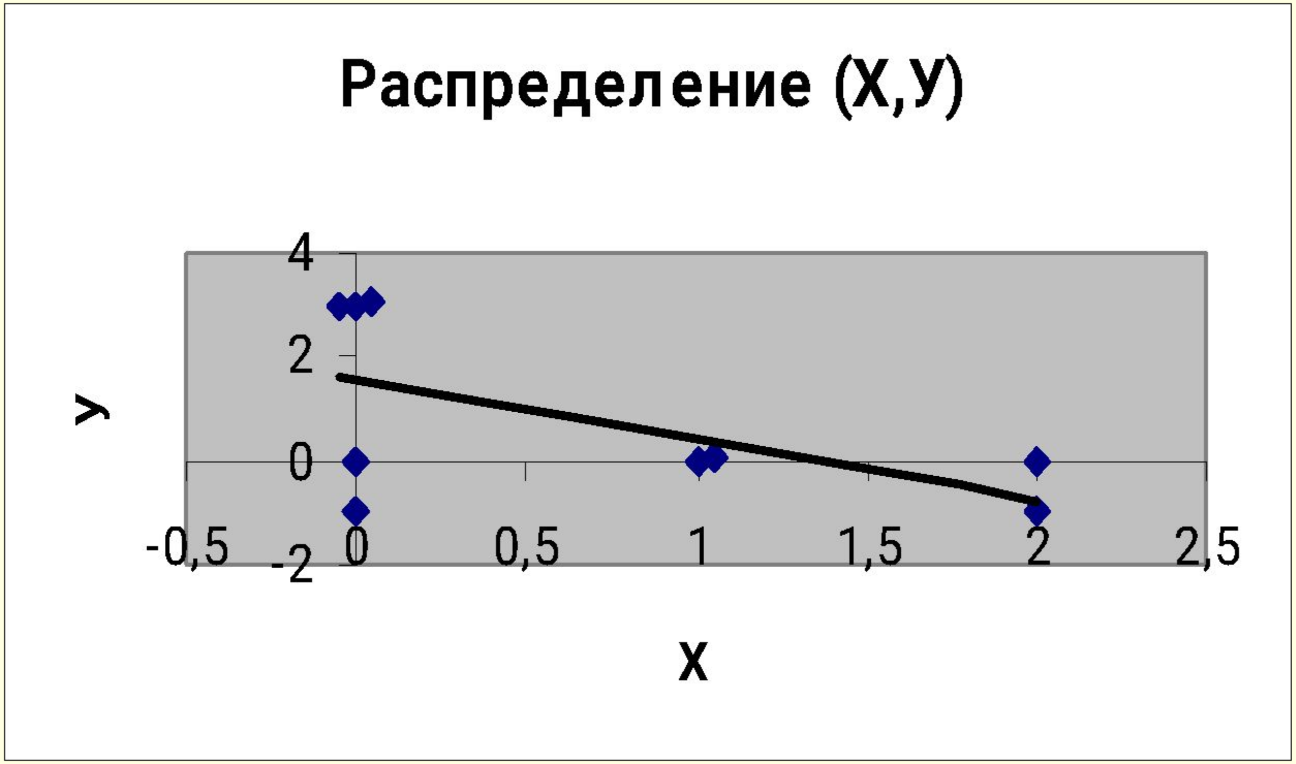
Найдем остаточную дисперсию:

$$\mathbf{S^2_{ост.} = \sigma^2_Y (1 - \rho^2).}$$

$$S^2_{ост.} = 2,64 \cdot (1 - 0,64^2) \approx \underline{1,56.}$$

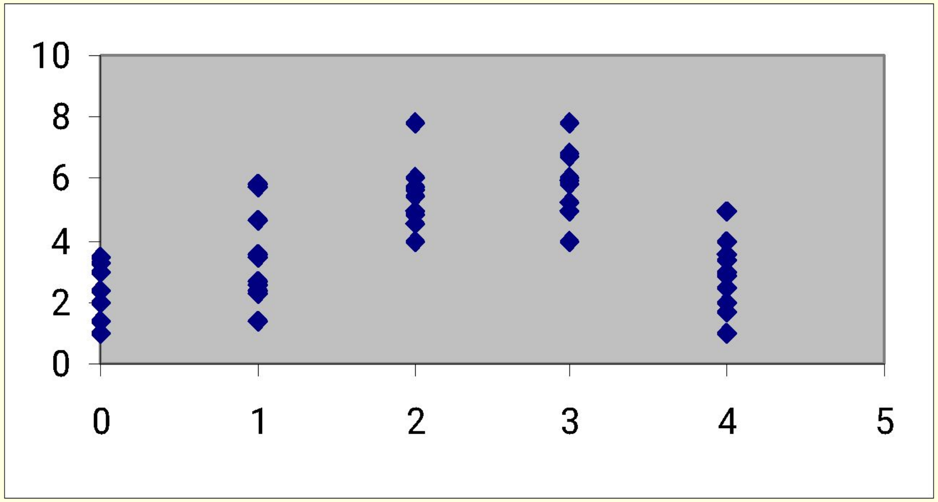
График линейной регрессии

$$\hat{Y} = -1{,}16X + 1{,}53.$$



Нелинейная зависимость

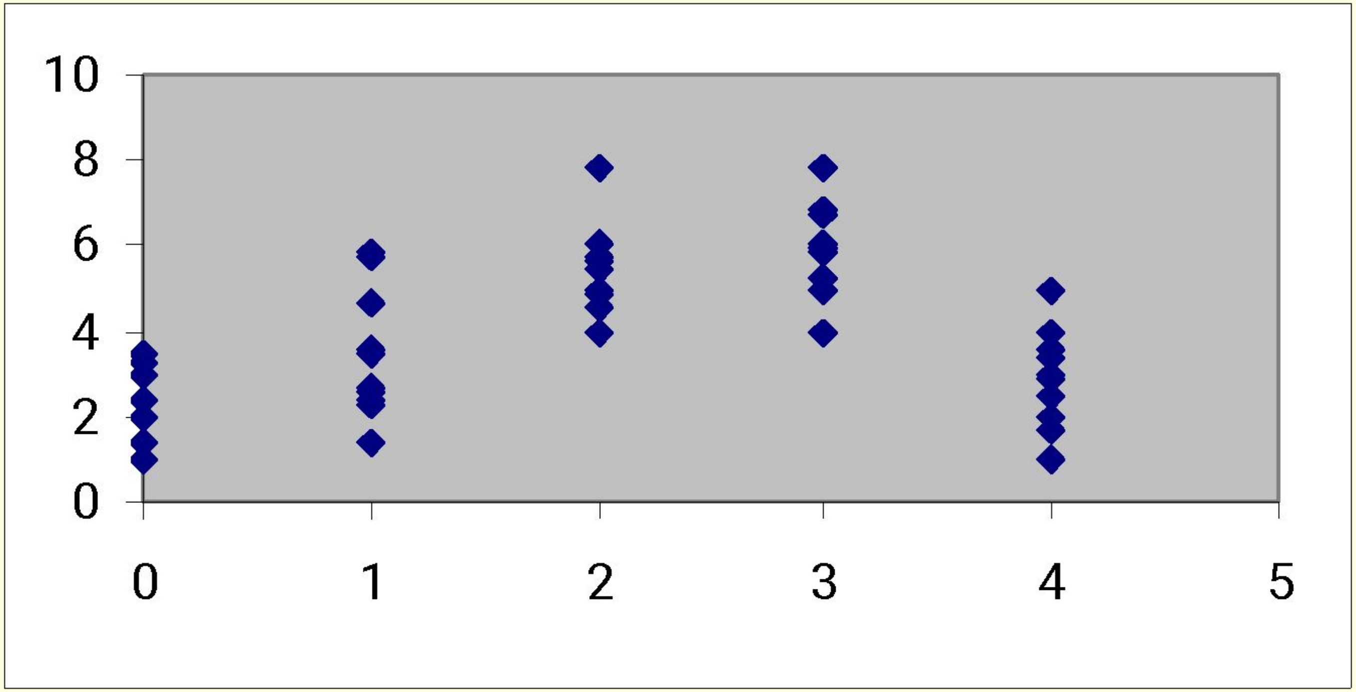
- Проблема: найти функцию, описывающую нелинейную зависимость.



Условные распределения

- Пусть (ξ, η) – двумерная случайная величина. Рассмотрим распределение η при условии, что ξ = x. Оно называется условным.

Условные распределения η при разных значениях ξ.



Способы нахождения условных распределений в дискретном случае

- Рассмотрим пример. Пусть дискретная двумерная случайная величина (X, Y) задана таблицей:

Y \ X	0	1	2
-1	0,1	0	0,2
0	0,1	0,2	0,1
3	0,3	0	0

Пример

Найдем условный закон распределения $Y/X = 0$:

$Y/X=0$	-1	0	3
P	$1/5$	$1/5$	$3/5$

Действительно,

$P(Y = -1/X = 0) = P(Y = -1, X = 0)/P(X = 0)$ т.к. по формуле условной вероятности,

$P(A/B) = P(AB)/P(B)$.

$P(Y = -1, X = 0) = 0{,}1$

$P(X = 0) = 0{,}5$.

Отсюда $P(Y = -1/X = 0) = 0{,}1 : 0{,}5 = 1/5$.

Аналогично $P(Y = 0/X = 0) = 0{,}1 : 0{,}5 = 1/5$,

$P(Y = 3/X = 0) = 0{,}3 : 0{,}5 = 3/5$.

Найдем другие условные законы.

Условный закон распределения Y/X = 1:

Y/X=1	–1	0	3
P	0	1	0

- Такой закон распределения записывается в виде ряда распределения

Y/X=1	**0**
P	**1**

Условный закон распределения Y/X=2:

Y/X=2	**–1**	**0**
P	**2/3**	**1/3**

Условное математическое ожидание

- **Определение.** Условным математическим ожиданием случайной

величины η при условии, что ξ = x,

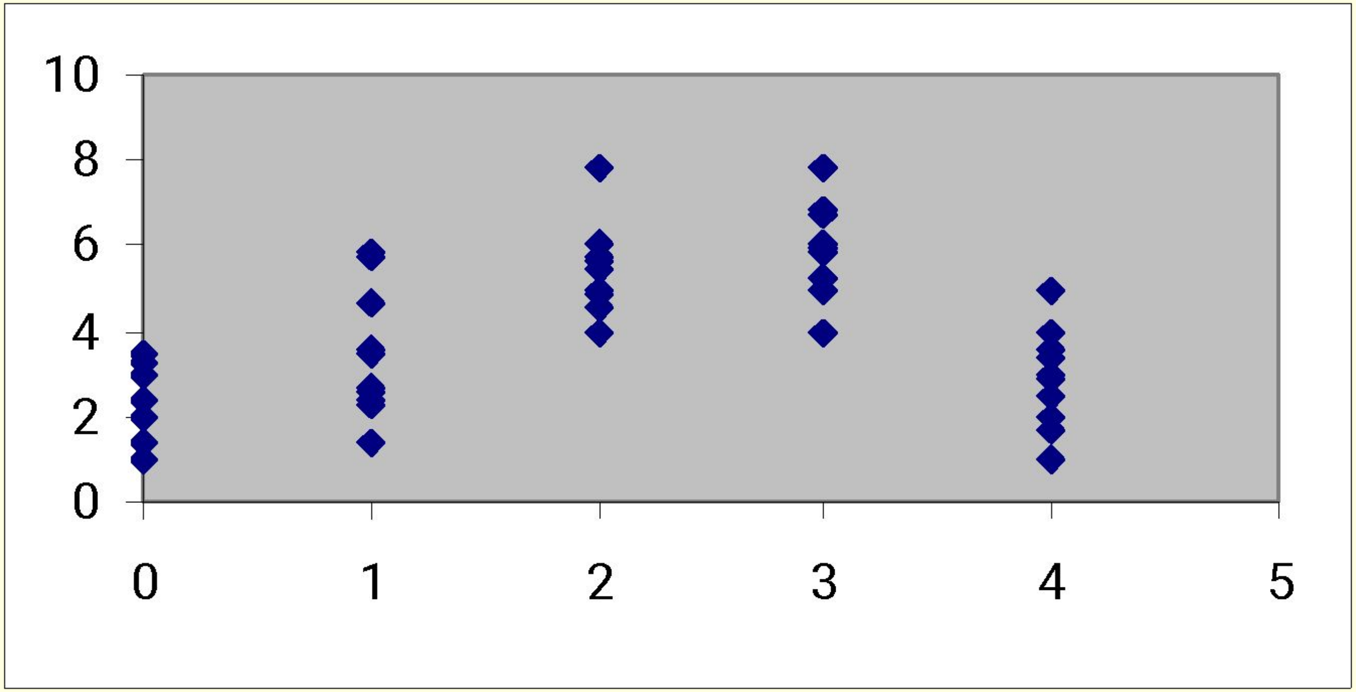
называется математическое ожидание, найденное с помощью условного закона распределения.

Обозначение: **$M(\eta/\xi = x)$.**

Замечание

- Условное математическое ожидание обладает свойствами математического ожидания .

Условное математическое ожидание



Вспомним предыдущий пример.

Найдем условное матожидание Y/X=0:

M(Y/X=0)= (–1)·1/5 + 0·1/5 + 3·3/5 = 8/5

Y/X=0	–1	0	3
P	1/5	1/5	3/5

Аналогично, условные матожидания

$M(Y/X=1) = 0 \cdot 1 = 0$,

$M(Y/X=2) = (-1) \cdot 2/3 + 0 \cdot 1/3 = -2/3$.

Регрессия

- **Определение.** Регрессией η на ξ называется случайная величина $r(\xi)$, равная при каждом x условному математическому ожиданию случайной величины η при условии, что $\xi = x$.
Определение. Линией регрессии называется линия $y = r(x)$, где

$$r(x) = M(\eta/\xi = x).$$

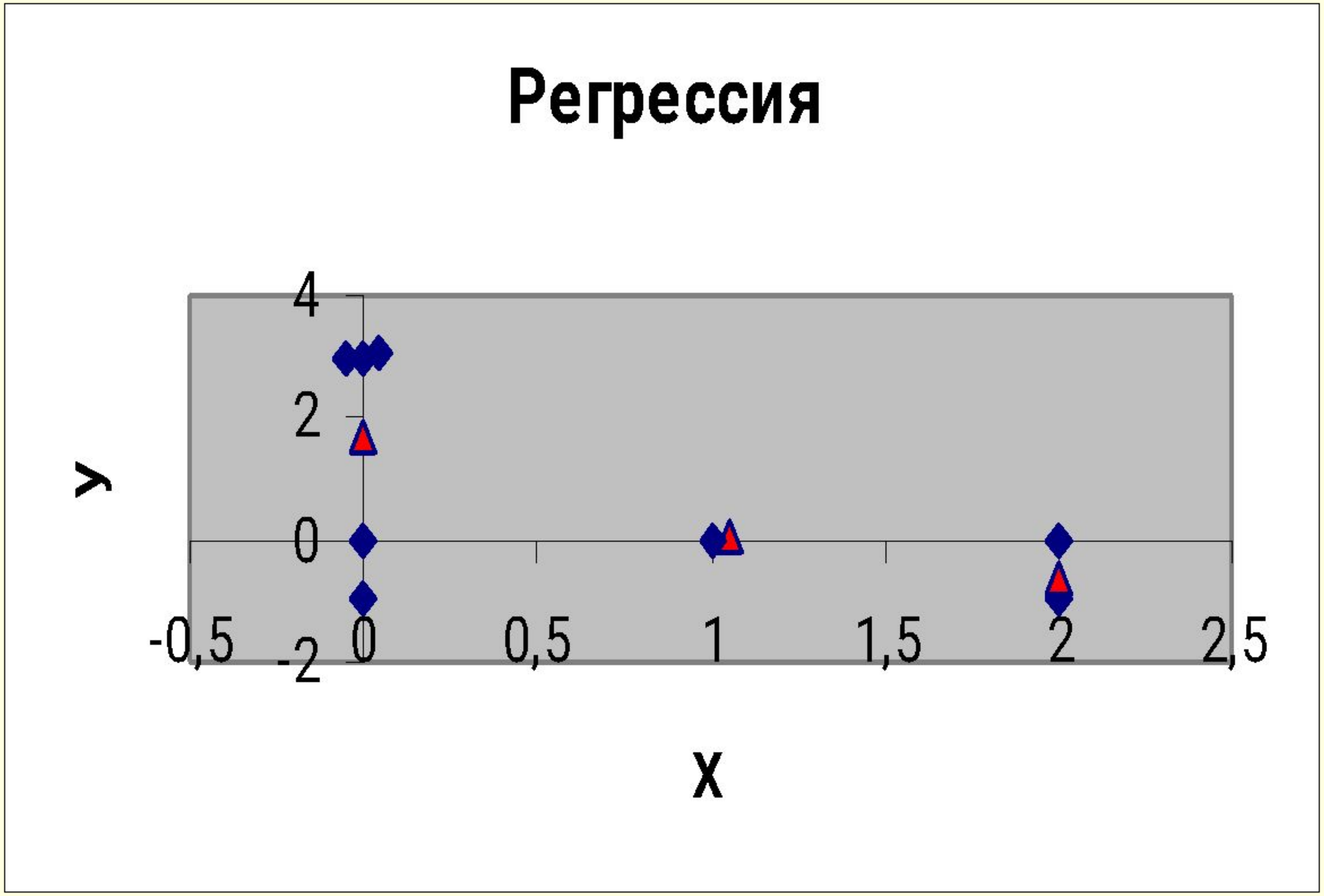
Пример

- В условиях предыдущего примера регрессия Y на X равна:

$$\mathrm{r}(x) = \begin{cases} 8/5, & x = 0 \\ 0, & x = 1 \\ -2/3, & x = 2 \end{cases}$$

Другой способ записи регрессии

r(x)	-2/3	0	8/5
P	0,5	0,2	0,3



Регрессия
4
2
0
-2
y
-0,5
0
0,5
1
1,5
2
2,5
X

Корреляционное отношение

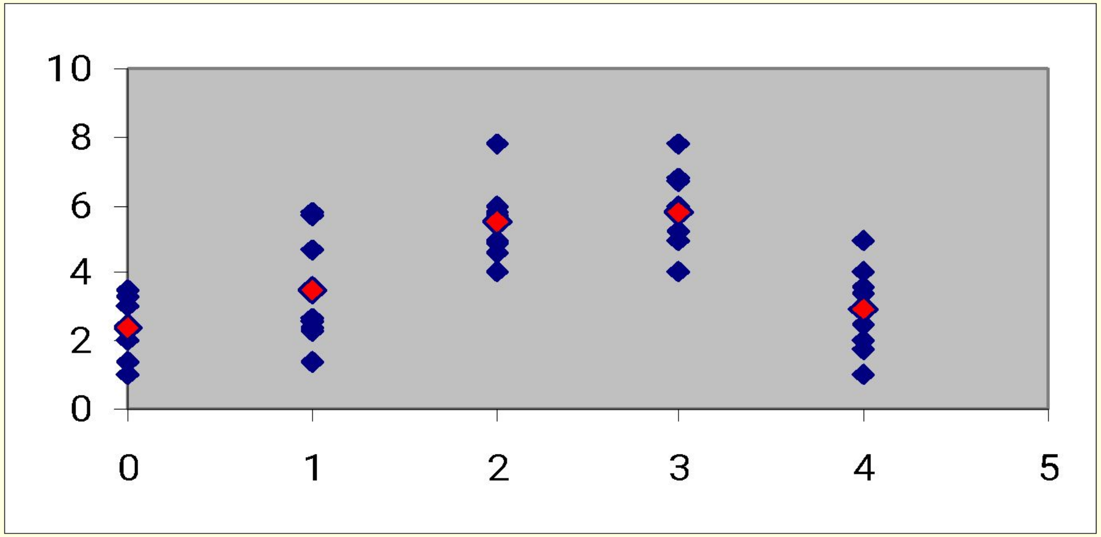
- **Определение.** Корреляционным отношением η на ξ называется числовая характеристика, равная

$$\boxed{\theta^2_{\eta\xi} = \frac{M(r(\xi) - M\eta)^2}{\sigma^2_{\eta}}}$$

Свойства корреляционного отношения

- **1. $0 \le \theta^2_{\eta,\xi} \le 1$.**
- **2. $\theta^2_{\eta,\xi} \ge \rho^2$.**
- **3. $\theta^2_{\eta,\xi} = \rho^2 \leftrightarrow r(\xi) = a\xi + b$ (т.е., линейная зав–ть).**
- **4. $\theta^2_{\eta,\xi} = 0 \leftrightarrow r(\xi) = M\eta$ ($r(\xi)$=const, нет связи).**
- **5. $\theta^2_{\eta,\xi} = 1 \leftrightarrow \eta = r(\xi)$ (т.е., функц–я зав–ть).**

Смысл: корреляционное отношение измеряет силу зависимости η от ξ



$$\theta^2_{\eta\xi} = \frac{M(r(\xi) - M\eta)^2}{\sigma^2_\eta}$$

Пример.

$$\theta^2_{YX} = \frac{M(r(X) - MY)^2}{\sigma^2_Y}$$

Y \ X	0	1	2
-1	0,1	0	0,2
0	0,1	0,2	0,1
3	0,3	0	0

- Чтобы найти θ^2_{YX}, надо сначала найти MY и DY.
- Мы их недавно находили с помощью одномерного закона.

X	0	1	2
P	0,5	0,2	0,3

- **MY = (–1)·0,3 + 0·0,4 + 3·0,3 = 0,6.**
- **DY = (–1)^2·0,3 + 0^2·0,4 + 3^2·0,3 – 0,6^2 = 2,64.**

$$\mathrm{r}(x) = \begin{cases} 8/5, & x = 0 \\ 0, & x = 1 \\ -2/3, & x = 2 \end{cases}$$

$$\theta^2{}_{YX} = \frac{1}{2,64}\left[(8/5-0,6)^2\cdot 0,5+(0-0,6)^2\cdot 0,2+(-2/3-0,6)^2\cdot 0,3\right].$$

Смысл полученного числа: корреляционное отношение измеряет силу зависимости Y от X.

- Чем ближе к 1, тем связь сильнее, чем ближе к 0, тем слабее.
- <u>Напоминание</u>: корреляционное отношение принимает значения от 0 до 1.
- Если надо найти θ^2_{XY}, а не θ^2_{YX} , то в формуле надо поменять местами X и Y.

$$\boxed{\theta^2{}_{XY} = \frac{M(r(Y) - MX)^2}{\sigma^2{}_X}.}$$

Условные распределения

- **Определение.** Условной функцией распределения случайной величины

η при условии, что ξ = x, называется

$$F_{\eta/\xi = x} = P(\eta < y/\xi = x).$$

Условная плотность

- <u>**Определение.**</u> Если условная функция распределения случайной величины η при условии, что $\xi = x$, непрерывна, то производная от нее называется условной плотностью распределения случайной величины η при условии, что $\xi = x$.

- Обозначается условная плотность

$$f_{\eta/\xi = x}(y)$$

(плотность распределения η в точке y при условии, что ξ = x).

Нахождение условной функции распределения

- **Условная функция распределения случайной величины η при условии, что ξ = x**

$$\mathrm{F}_{\eta/\xi=\mathrm{x}}(\mathrm{y}) = \frac{\int\limits_{-\infty}^{y} f_{\xi,\eta}(x,v)dv}{\int\limits_{-\infty}^{+\infty} f_{\xi,\eta}(x,v)dv}.$$

Нахождение условной плотности распределения

- Условная плотность распределения сл. в. η при условии, что ξ = x

$$\mathrm{f}_{\eta/\xi=\mathrm{x}}(\mathrm{y}) = \frac{f_{\xi,\eta}(x,y)}{\int\limits_{-\infty}^{+\infty} f_{\xi,\eta}(x,v)\,dv}.$$

- Поскольку

$$\int_{-\infty}^{+\infty} f_{\xi,\eta}(x,v)dv = f_{\xi}(x),$$

$$\boxed{\mathrm{f}_{\eta/\xi=\mathrm{x}}(\mathrm{y}) = \frac{f_{\xi,\eta}(x,y)}{f_{\xi}(x)}.}$$

Числовые характеристики многомерных случайных величин

- **Определение.** Ковариационной матрицей случайных величин $\xi_1, \xi_2, \ldots, \xi_n$ называется матрица K размерности n x n с элементами a_{ij}, равными ковариациям $cov(\xi_i, \xi_j) = k_{ij}$.

$$K = (k_{ij})_{n \times n} = (cov(\xi_i, \xi_j))_{n \times n}$$

Ковариационная матрица К

$$K = \begin{pmatrix} \sigma_1{}^2 \; K_{12} \; \ldots K_{1n} \\ K_{21} \; \sigma_2{}^2 \; \ldots K_{2n} \\ \ldots \quad \ldots \; \ldots \; \ldots \\ K_{n1} \; K_{n2} \ldots \sigma_n{}^2 \end{pmatrix}$$

Корреляционная матрица R

- **Наряду с ковариационной матрицей рассматривают и матрицу R, составленную из коэффициентов корреляции $\rho_{ij} = \rho(\xi_i, \xi_j)$.**

$$R = \begin{pmatrix} 1 & \rho_{12} & \dots & \rho_{1n} \\ \rho_{21} & 1 & \dots & \rho_{2n} \\ \dots & \dots & \dots & \dots \\ \rho_{n1} & \rho_{n2} & \dots & 1 \end{pmatrix}$$

Уравнение множественной линейной регрессии

- Рассмотрим случайные величины

$$\xi_0\ \xi_1,\ \xi_2\ ,\ \ldots,\ \xi_n$$

с математическими ожиданиями $M\xi_i = a_i$,

с дисперсиями $D\xi_i = \sigma^2{}_i$,

$$i = 0,1,\ldots,\ n,$$

и с корреляционной матрицей R размерности $(n+1) \times (n+1)$.

- **<u>Определение.</u>** Уравнением линейной регрессии ξ_0 на ξ_1, ξ_2 , …, ξ_n называется уравнение

$$\boxed{\hat{\xi}_0 = a_0 + \sum_{i=1}^{n} b_i(\xi_i - a_i).}$$

- Здесь b_i (i =1,…, n) – параметры, минимизирующие остаточную дисперсию

$$\boxed{M(\xi_0 - \hat{\xi}_0)^2.}$$

- Минимизируя остаточную дисперсию, получаем, что

$$\boxed{b_i = -\frac{R_{0i}}{R_{00}} \cdot \frac{\sigma_0}{\sigma_i},}$$

Остаточная дисперсия

Здесь и далее через R_{ij} обозначено алгебраическое дополнение элемента a_{ij} матрицы R, а через |R| – определитель матрицы R.

Остаточная дисперсия равна

$$\boxed{S^2_{ост.} = \sigma_0{}^2 \cdot \frac{|R|}{|R_{00}|},}$$

Частный коэффициент корреляции

- Частный коэффициент корреляции используется как мера линейной зависимости между двумя какими –либо случайными величинами за вычетом влияния остальных случайных величин.

$$\rho_{0i} = \frac{-R_{0i}}{\sqrt{R_{00} \cdot R_{ii}}}$$

Множественный (сводный) коэффициент корреляции

- Выражает зависимость между ξ_0 и всей совокупностью $\xi_1, \xi_2, \dots, \xi_n$.

$$\rho_{0(1\dots n)} = \sqrt{1 - \frac{|R|}{R_{00}}}.$$