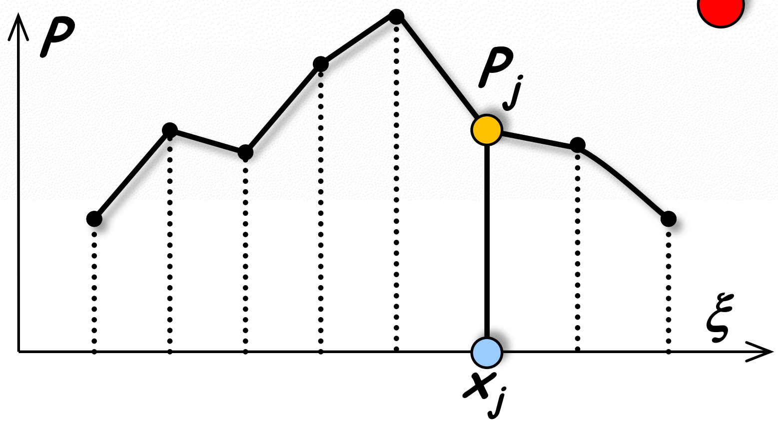
Числовые характеристики случайных величин

{ определение случайных величин - математическое ожидание - примеры – дисперсия – центральные и начальные моменты – коэффициент корреляции - ковариация }

- *Математическим ожиданием дискретной случайной величины* $\xi(\omega)$ называется величина

$$M\xi = \sum_{\omega \in \Omega} \xi(\omega) p(\omega)$$

Если д.с.в. ξ имеет математическое ожидание, то оно находится как сумма всех произведений значений случайных величин на их вероятности



$$M\xi = \sum_{j=1}^{n} x_j p_j$$

Случайных величин без математического ожидания не бывает, так как, если у нас есть случайная величина мы всегда в праве от нее что-нибудь ожидать.

Из студенческой контрольной работы

@ *Найти математическое ожидание для биномиального распределения дискретной случайной величины.*

Решение

$$M(\xi)=\sum_{j=1}^{n}x_j p_j=\sum_{m=1}^{n}mC_n^m p^m g^{n-m}=\sum_{m=1}^{n}m\frac{n!}{m!(n-m)!}p^m g^{n-m}=$$

$$=np\sum_{m=1}^{n}\frac{(n-1)!}{(m-1)!((n-1)-(m-1))!}p^{m-1}g^{n-m}=$$

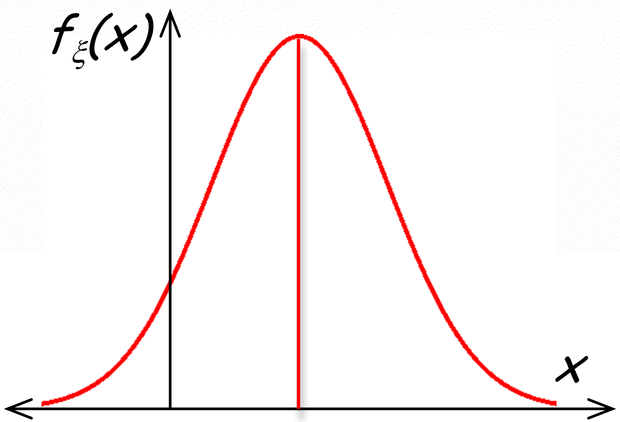
$$=np\sum_{m=1}^{n}C_{n-1}^{m-1}p^{m-1}g^{(n-1)-(m-1)}=np\sum_{m^*=0}^{n^*}C_{n^*}^{m^*}p^{m^*}g^{n^*-m^*}=np$$

$$M(\xi)=np$$

Математическое ожидание случайной величины

- *Математическим ожиданием* непрерывной случайной величины ξ называется величина

$$M\xi = \int\limits_{-\infty}^{+\infty} x f_\xi(x)\,dx$$



Свойства:

- $Mc = c$
- $Mc\xi = cM\xi$
- $M\sum\limits_{j=1}^{n}\xi_j = \sum\limits_{j=1}^{n} M\xi_j$
- $M(\xi\cdot\eta) = M(\xi)M(\eta)$

  (ξ и η – независимые СВ)

@ *Найти математическое ожидание для нормального распределения непрерывной случайной величины.*

Решение

$$M = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{+\infty} t e^{-\frac{(t-m)^2}{2\sigma^2}}dt \Rightarrow \frac{(t-m)}{\sigma} = z \Rightarrow \frac{d(t-m)}{\sigma} = \frac{1}{\sigma}dt = dz \Rightarrow$$

$$M = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}(\sigma z + m)e^{-\frac{z^2}{2}}dz =$$

$$= \frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} z e^{-\frac{z^2}{2}}dz + m\left(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}}dz\right) = m$$

$= 1\,!$

$= 0\,!$

$$M(\xi) = m$$

Дисперсия случайной величины

- *Дисперсия случайной величины* ξ определяется как математическое ожидание квадрата отклонения случайной величины от её математического ожидания

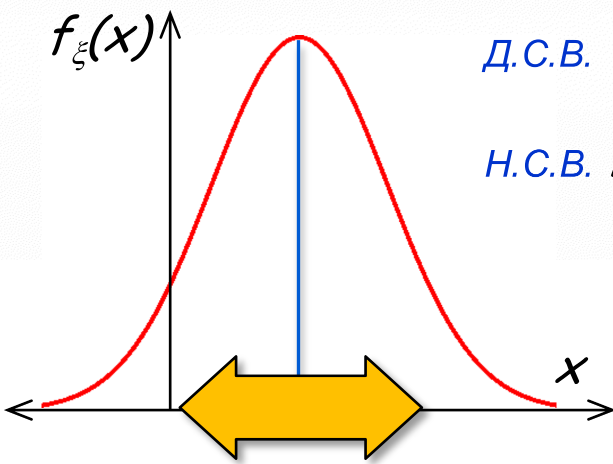
$$M(\xi - M\xi) = M(\xi) - M(\xi) = 0$$

- $$D\xi = M(\xi - M\xi)^2 = \sigma^2$$

σ - среднее квадратичное отклонение

$$D\xi = M(\xi - M\xi)^2 = M(\xi^2) - M(2\xi M) + M(M^2\xi)$$

$$\Rightarrow \quad D\xi = M(\xi^2) - (M\xi)^2$$



Д.С.В. $D\xi = \sum_j p_j(x_j - M\xi)^2 = \sum_j x_j^2 p_j - (\sum_j x_j p_j)^2$

Н.С.В. $D\xi = \int_{-\infty}^{+\infty} f_\xi(x)(x - M\xi)^2 dx = \int_{-\infty}^{+\infty} x^2 f_\xi(x) dx - (\int_{-\infty}^{+\infty} x f_\xi(x) dx)^2$

Свойства:

- $Dc = 0$
- $Dc\xi = c^2 D\xi$
- $D\sum_j \xi_j = \sum_j D\xi_j$

ξ_j независимы в совокупности

- $D(\xi - \eta) = D(\xi) + D(\eta)$

@ Найти дисперсию для биномиального распределения

Решение

Предварительно найдем дисперсию в распределении Бернулли

X	0	1
P	q	p

$$MX = 0 \cdot q + 1 \cdot p = p \quad MX^2 = 0 \cdot q + 1 \cdot p = p$$

$$DX = MX^2 - (MX)^2 = p - p^2 = p(1-p) = pg$$

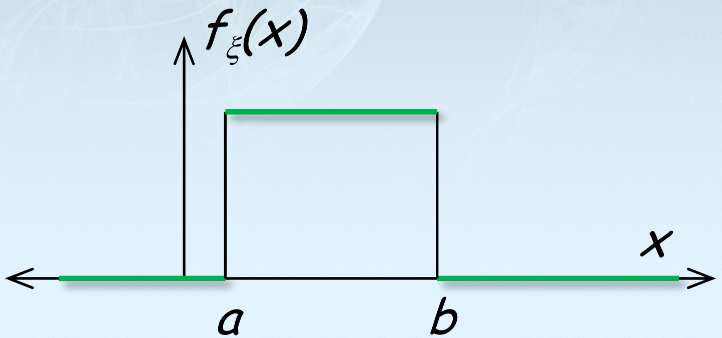
$$\xi = X_1 + X_2 + \ldots. + X_n \qquad M\xi = \sum_{j=1}^{n} MX_j = np$$

$$D\xi = \sum_{j=1}^{n} DX_j = npg \Rightarrow \sigma = \sqrt{npq}$$

Найти дисперсию для равномерного распределения

$$f_\xi(x) = \begin{cases} \dfrac{1}{b-a}, & x \in [a,b] \\ 0, & x \notin [a,b] \end{cases}$$

$$M\xi = \int_a^b x \frac{1}{b-a} dx = \left.\frac{x^2}{2(b-a)}\right|_a^b = \frac{a+b}{2}$$



$$M\xi^2 = \int_a^b x^2 \frac{1}{b-a} dx = \left.\frac{x^3}{3(b-a)}\right|_a^b = \frac{a^2+ab+b^2}{3}$$

$$D\xi = M\xi^2 - (M\xi)^2 = \frac{a^2+ab+b^2}{3} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}$$

Математическим моментом случайной величины называется математическое ожидание величины $(\xi - a)^k$: $\mu_k(a) = M(\xi - a)^k$

Если $a = M(\xi)$, момент называется *центральным*, если $a = 0$, то *начальным* - ν_k

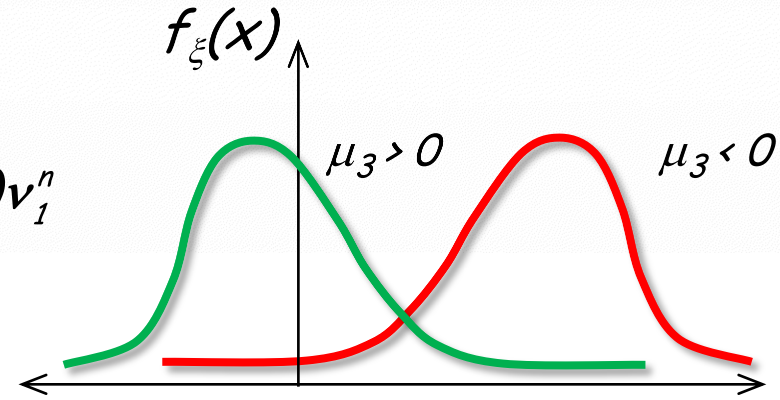
Абсолютный момент k-го порядка: $m_k(a) = M\,|\,\xi - a\,|^k$

$$\nu_1 = M(\xi),\ \mu_0 = 1,\ \mu_1 = 0,\ \mu_2(a) = M(\xi - a)^2 = M(\xi^2) - (M\xi)^2 = \nu_2 - \nu_1^2$$

$$\mu_3 = \nu_3 - 3\nu_2\nu_1 + 2\nu_1^3$$

$$\mu_4 = \nu_4 - 4\nu_3\nu_1 + 6\nu_2\nu_1^2 - 3\nu_1^4$$

$$\mu_n = \sum_{k=2}^{n} (-1)^{n-k} C_n^k \nu_k \nu_1^{n-k} + (-1)^{n-1}(n-1)\nu_1^n$$



$$A = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\mu_3}{\sigma^3} \qquad E = \frac{\mu_4}{\sigma^4} - 3$$

Коэффициент ассиметрии

Эксцесс – степень крутости распределения

E=0 для случая Н.З. распределения

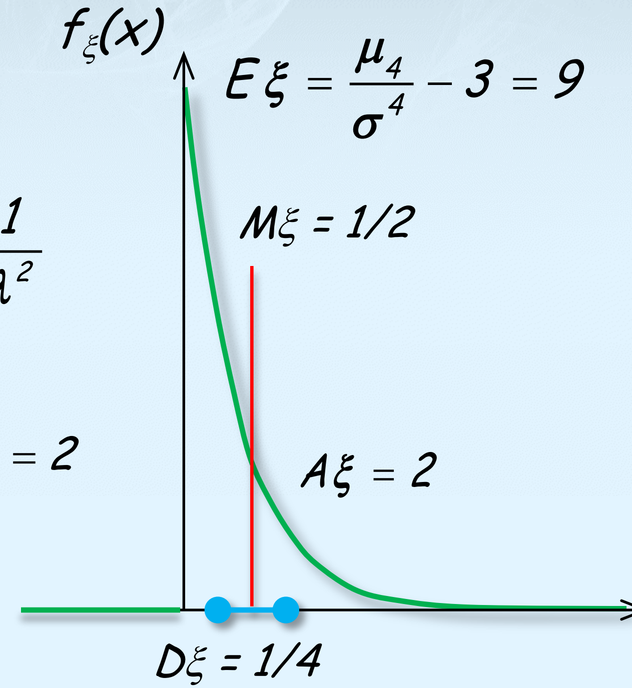
@ *Определить числовые характеристики для экспоненциального распределения*

$$f_\xi(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \le 0 \end{cases} \qquad \lambda = 2 \qquad M\xi = \frac{1}{\lambda}$$

$$E\xi = \frac{\mu_4}{\sigma^4} - 3 = 9$$

$$D\xi = \nu_2 - \nu_1^2 = \int_0^\infty x^2 \lambda e^{-\lambda x} dx = \frac{2}{\lambda^2} - (1/\lambda)^2 = \frac{1}{\lambda^2}$$

$$A\xi = \frac{\mu_3}{\sigma^3} = \frac{\nu_3 - 3\nu_1\nu_2 + 2\nu_1^3}{\left(\frac{1}{\lambda}\right)^3} = \frac{\frac{6}{\lambda^3} - 3\frac{1}{\lambda}\frac{2}{\lambda^2} + 2\frac{1}{\lambda^3}}{\left(\frac{1}{\lambda}\right)^3} = 2$$



Числовые характеристики зависимости двух случайных величин

Для независимых случайных величин с конечными вторыми моментами дисперсия их суммы равна сумме их дисперсий. Чему равна дисперсия суммы в общем случае?

$$D(\xi+\eta)=M(\xi+\eta)^2-(M(\xi+\eta))^2=M(\xi^2+\eta^2+2\xi\eta)-(M\xi)^2-(M\eta)^2-2M\xi M\eta$$

$$D(\xi+\eta)=D\xi+D\eta+2(M(\xi\eta)-M\xi M\eta)$$

Величина $M(\xi\eta)$ - $M\xi M\eta$ равняется нулю, если случайные величины ξ и η независимы. Однако из равенства ее нулю вовсе не следует независимость. Эту величину часто используют как «индикатор наличия зависимости» пары случайных величин.

- *Ковариацией* $cov(\xi,\eta)$ случайных величин ξ и η называется *число*

$$cov(\xi,\eta)=M((\xi-M\xi)(\eta-M\eta))$$

$$cov(\xi,\eta)=M((\xi-M\xi)(\eta-M\eta))=M(\xi\eta)-M\xi M\eta$$

$$cov(\xi,\xi)=D\xi \quad cov(\xi,\eta)=cov(\eta,\xi)$$

$$D(\xi_1+\ldots+\xi_n)=\sum_{i=1}^{n}D\xi_i+\sum_{i\neq j}cov(\xi_i,\xi_j)=\sum_{i=1}^{n}D\xi_i+2\sum_{i<j}cov(\xi_i,\xi_j)=\sum_{i,j}cov(\xi_i,\xi_j)$$

- Если ковариация $cov(\xi,\eta)$ отлична от нуля, то величины ξ и η *зависимы* !

@ *Пусть **ξ** и **η** — независимые случайные величины, и дисперсия **ξ** отлична от нуля. Доказать, что **ξ** и **ξ** + **η** зависимы.*

$$M(\xi(\xi+\eta)) = M\xi^2 + M\xi M\eta$$

$$M\xi M(\xi+\eta)) = (M\xi)^2 + M\xi M\eta$$

$$cov(\xi,\xi+\eta) = M\xi^2 + M\xi M\eta - ((M\xi)^2 + M\xi M\eta) =$$

$$= D\xi > 0$$

Следовательно, ξ и $\xi + \eta$ зависимы

Ковариация – размерная величина. Её нужно нормировать так, чтобы она не менялась при умножении или сдвиге случайных величин и свидетельствовала о силе их зависимости.
Говоря о «силе» зависимости между с.в., мы имеем в виду следующее. Самая сильная зависимость – функциональная, а из функциональных – линейная зависимость, когда $\xi = a\eta + b$.

- ***Коэффициентом корреляции*** $\rho(\xi, \eta)$ случайных величин $\boldsymbol{\xi}$, $\boldsymbol{\eta}$, дисперсии которых существуют и отличны от нуля, называется число

$$\rho(\xi,\eta) = \frac{cov(\xi,\eta)}{\sqrt{D\xi}\sqrt{D\eta}}$$

1. Если с. в. ξ и η независимы, то $\rho(\xi, \eta) = cov(\xi, \eta) = 0$
2. $|\rho(\xi, \eta)| \leq 1$
3. $|\rho(\xi, \eta)| = 1$, если и только если случайные величины ξ и η с вероятностью 1 линейно связаны, т.е. существуют числа $a \neq 0$ и b такие, что $P(\eta = a, \xi = b) = 1$

@ *Найти коэффициент корреляции для* ξ *и* $\xi+\eta$.
ξ, η - независимые случайные величины, дисперсии ξ и η равны между собой и отличны от нуля.

Решение

$$cov(\xi,\xi+\eta)=M\xi^2+M\xi M\eta-((M\xi)^2+M\xi M\eta)=D\xi$$

$$\rho(\xi,\xi+\eta)=\frac{cov(\xi,\xi+\eta)}{\sqrt{D\xi}\sqrt{D(\xi+\eta)}}=$$

$$=\frac{D\xi}{\sqrt{D\xi}\sqrt{D\xi+D\eta}}=\frac{D\xi}{\sqrt{D\xi}\sqrt{2D\xi}}=\frac{1}{\sqrt{2}}$$

ξ и $\xi+\eta$ зависимы с коэффициентом корреляции $\rho = 0.707$

@ *Найти коэффициент корреляции между числом выпадений единицы и числом выпадений шестерки при n подбрасываниях симметричного кубика.*

Решение

Обозначим для $i = 1, 2, 3, 4, 5, 6$ через ξ_i случайную величину, равную числу выпадений грани с i очками при n подбрасываниях кубика. Подсчитаем $cov(\xi_1, \xi_6)$

Каждая из случайных величин ξ_i имеет биномиальное распределение с параметрами n и $p = 1/6$, поэтому

$$M\xi_i = np = \frac{n}{6} \qquad D\xi_i = npg = np(1-p) = n\frac{1}{6}\frac{5}{6} = \frac{5n}{36}$$

@ Сумма $\xi_1 + \ldots + \xi_n$ равна n. В силу симметрии кубика, все математические ожидания $M_{i,j}$ равны, но отличаются от $M_{i,i}$

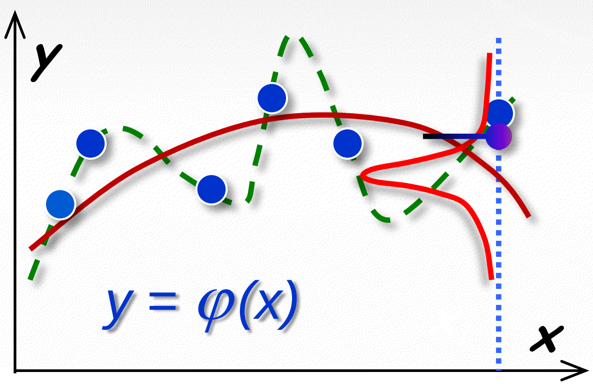
$$M\xi_1\xi_2 = M\xi_1\xi_3 \ldots = M\xi_1\xi_6 \qquad M\xi_1\xi_1 = M\xi_1^2 = D\xi_1 + (M\xi_1)^2 = \frac{5n}{36} + \frac{n^2}{36}$$

$$M\xi_1(\xi_1 + \ldots + \xi_6) = M\xi_1 n = \frac{n^2}{6}$$ или по другому

$$M\xi_1(\xi_1 + \ldots + \xi_6) = M\xi_1^2 + 5M\xi_1\xi_6 = \frac{5n}{36} + \frac{n^2}{36} + 5M\xi_1\xi_6$$

$$\rho(\xi_1, \xi_6) = \frac{M(\xi_1\xi_6) - M\xi_1 M\xi_6}{\sqrt{D\xi_1}\sqrt{D\xi_6}} \qquad \rho(\xi_1, \xi_6) = \frac{\dfrac{n^2 - n}{36} - \dfrac{n^2}{36}}{\dfrac{5n}{36}} = -\frac{1}{5}$$

- Даны два множества величин x_i и y_i, где y_i – измеряется с ошибкой. Рассмотрим задачу *установления функциональной зависимости между двумя величинами* $y=\varphi(x)$ *по результатам измерений, в которых значения функции определяются с ошибкой* – случайной величиной, которая подчиняется нормальному закону.



$$y_i \in \left(y_1, y_2, y_3, \ldots, y_n\right)$$

$$m = \varphi(x_i) \qquad \sigma_1 = \sigma_2 = \ldots = \sigma_i = \ldots = \sigma$$

$$dP_i = f_i(y_i)dy_i = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{[y_i - \varphi(x_i)]^2}{2\sigma^2}} dy_i$$

min

$$dP = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{[y_i - \varphi(x_i)]^2}{2\sigma^2}} dy_i = Ke^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[y_i - \varphi(x_i)]^2}$$

- *Данная совокупность измеренных значений величин* y_i *будет наивероятнейшей, если сумма квадратов отклонений наблюдаемых значений* y_i *от предполагаемых* $\varphi(x_i)$ *будет минимальной.*

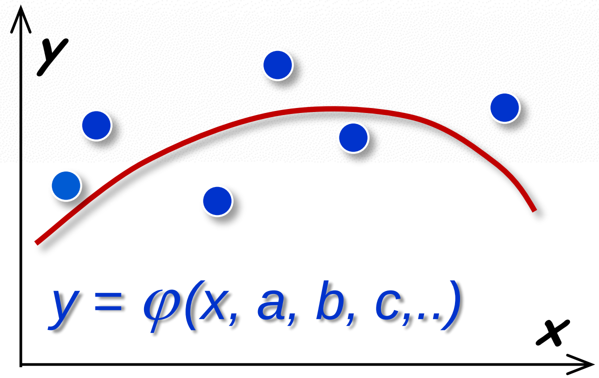
Получаем использование метода максимального правдоподобия для получения статистических оценок.

$$y = \varphi(x, a, b, c, \ldots)$$

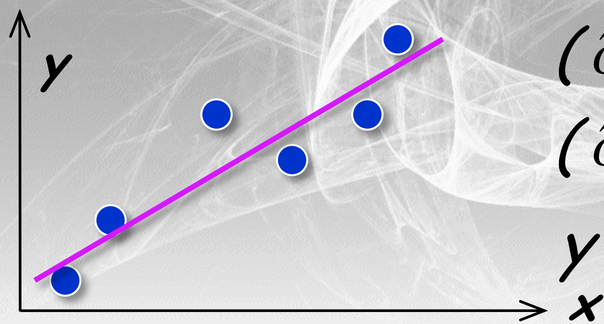
$$a = ?, b = ?, c = ?, \ldots$$

$$\sum_{i=1}^{n} [y_i - \varphi(x_i, a, b, c, \ldots)]^2 \Rightarrow min$$

$$\begin{cases} \sum\limits_{i=1}^{n} [y_i - \varphi(x_i, a, b, c, \ldots)](\partial\varphi/\partial a)_i = 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots = 0 \\ \sum\limits_{i=1}^{n} [y_i - \varphi(x_i, a, b, c, \ldots)](\partial\varphi/\partial c)_i = 0 \end{cases}$$



Подбор параметров для линейной функции



$$\left(\partial\varphi/\partial a\right)_i = x_i$$

$$\left(\partial\varphi/\partial b\right)_i = 1$$

$$y = \varphi(x,a,b) = ax + b$$

$$\begin{cases} \sum_{i=1}^{n}[y_i - ax_i - b]x_i = 0 \\ \sum_{i=1}^{n}[y_i - ax_i - b] = 0 \end{cases}$$

$$\begin{cases} \dfrac{\sum_{i=1}^{n} x_i y_i}{n} - a\dfrac{\sum_{i=1}^{n} x_i^2}{n} - b\dfrac{\sum_{i=1}^{n} x_i}{n} = 0 \\ \dfrac{\sum_{i=1}^{n} y_i}{n} - a\dfrac{\sum_{i=1}^{n} x_i}{n} - b = 0 \end{cases}$$

$$\begin{cases} \sum_{i=1}^{n}[y_i x_i - a x_i x_i - b x_i] = 0 \\ \sum_{i=1}^{n}[y_i - a x_i] - bn = 0 \end{cases}$$

$$\tilde{m}_x = \frac{\sum_{i=1}^{n} x_i}{n},\quad \tilde{m}_y = \frac{\sum_{i=1}^{n} y_i}{n},\quad \tilde{\nu}_{1,1} = \frac{\sum_{i=1}^{n} x_i y_i}{n},\quad \tilde{\nu}_2 = \frac{\sum_{i=1}^{n} x_i^2}{n}$$

Подбор параметров для линейной функции

$$\begin{pmatrix} \tilde{v}_2[x] & \tilde{m}_x \\ \tilde{m}_x & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \tilde{v}_{1,1}[x,y] \\ \tilde{m}_y \end{pmatrix}$$

Система уравнений примет вид

Решение $a = \dfrac{\tilde{v}_{1,1}[x,y] - \tilde{m}_x \tilde{m}_y}{\tilde{v}_2[x] - \tilde{m}_x^2}, \quad b = \tilde{m}_y - a\tilde{m}_x$

Запись уравнений упрощается, если ввести “центральные моменты”

$$\tilde{K}_{xy} = \tilde{v}_{1,1}[x,y] - \tilde{m}_x \tilde{m}_y$$
$$\tilde{D}_x = \tilde{v}_2[x] - (\tilde{m}_x)^2$$

Решение

$$a = \frac{\tilde{K}_{xy}}{\tilde{D}_x}, \quad b = \tilde{m}_y - a\tilde{m}_x$$

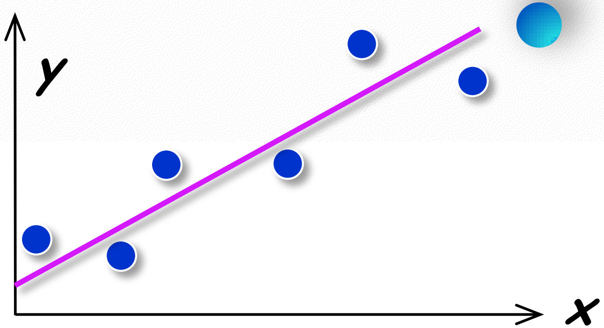
В приведенных формулах появляются выборочные средние, дисперсия и корреляционный момент

$$\tilde{m}_x = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \tilde{m}_y = \frac{1}{n}\sum_{i=1}^{n} y_i, \quad \tilde{D}_x = \frac{1}{n}\sum_{i=1}^{n}(x_i - \tilde{m}_x)^2$$

$$\tilde{K}_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \tilde{m}_x)(y_i - \tilde{m}_y) \Leftrightarrow cov\,(x,y) = M((x - Mx)(y - My))$$

$$\tilde{K}_{xy} = \frac{\sum_{i=1}^{n} x_i y_i}{n} - \tilde{m}_x \tilde{m}_y$$

$$\tilde{D}_x = \frac{\sum_{i=1}^{n} x_i^2}{n} - \tilde{m}_x^2$$



$$y - \tilde{m}_y = \frac{\tilde{K}_{xy}}{\tilde{D}_x}(x - \tilde{m}_x)$$

Линейная регрессия