СТРУКТУРА КУРСА

Теория пакет Statistica Приложения

Корреляционный анализ

Регрессионный анализ
- Простая
- Множественная
- Нелинейная

Дисперсионный анализ

Многомерный анализ
- Факторный
- Кластерный
- Дискриминантный

Планирование эксперимента
- Оптимизация планов
- Последовательное планирование
- Стохастическая аппроксимация

Статистика автокомбината

Центробанк

АСМАП

Маршрутная сеть

Террамеханика

Характеристика автомобилей

Поставки запчастей

ИНТЕРПРЕТАЦИЯ ДАННЫХ

Корреляционный анализ

Простая регрессия

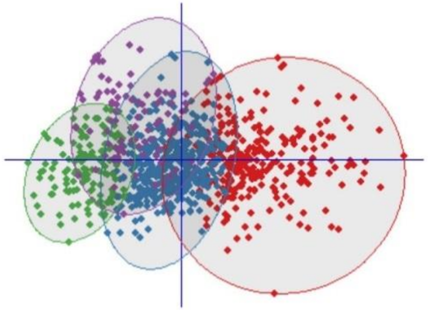
$$\sum_{i=1}^{n} d_i^2 \to \boldsymbol{min}$$

Множественная регрессия

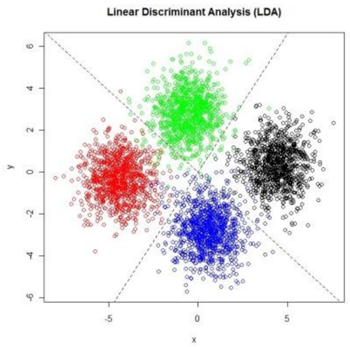
Нелинейная регрессия

Дисперсионный анализ

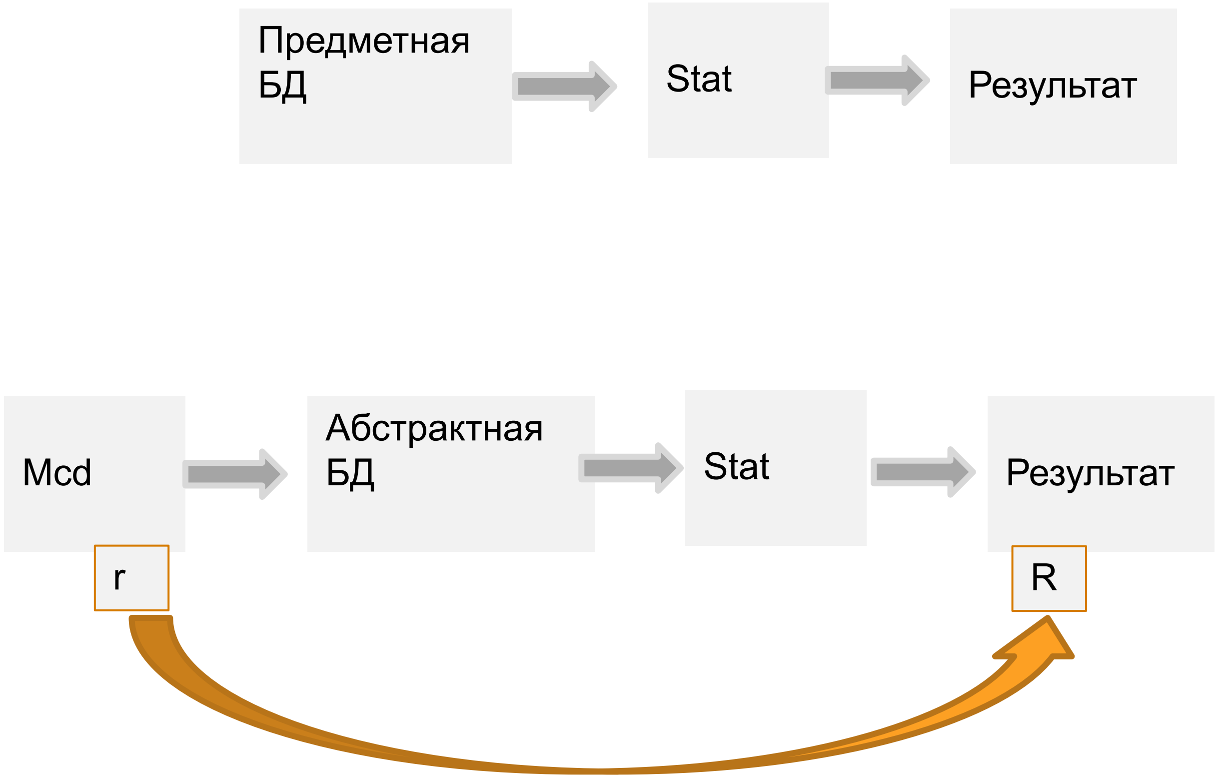
Кластерный анализ



Дискриминантный анализ



АНАЛИЗ РЕАЛЬНЫХ И МОДЕЛЬНЫХ ДАННЫХ



МАТРИЧНЫЕ ОПЕРАЦИИ В **MATHCAD**

matrix(L,N,f) - создание матрицы - **L** – число строк, **N** – число столбцов, **f** - функция **f(l,n)**

identity(n) - создание единичной матрицы размерности **n**

diag(v) - создание диагональной матрицы с элементами главной диагонали из **v**;

geninv(A) - обращение матрицы **A**, аналогично операции $\mathbf{A}^{-1}$.

rows(A) - определение числа строк матрицы **A**

cols(A) - определение числа столбцов матрицы **A**

last(v) - вычисление номера последнего элемента вектора **v**

lenght(v) - вычисление количества элементов **v**

csort(A,i), rsort(A,j) - сортировка по возрастанию элементов **i**–го столбца или строки

min(A), max(A) - минимальный и максимальный элемент матрицы

tr(A) - след квадратной матрицы **A**

rank(A) - ранг матрицы **A**

norm1(A), **norm2(A)**, **norme(A)**, **normi(A)** - нормы квадратной матрицы **A**

submatrix (M, r1, r2, c1, c2) - подматрица **M, r1** и **r2** – нижний и верхний номер строки, **c1** и **c2** - нижней и верхний номер столбца матрицы **M**

augment(A,B,…), stack(A,B,…) - слияние матриц **A, B, …** слева направо или сверху вниз

rref(A) - приведение матрицы к ступенчатому виду с единичным базисным минором

eigenvals(A) - собственные значения квадратной матрицы A

eigenvecs(A) - собственные вектора квадратной матрицы A по порядку собственных значений

eigenvec(A, l) - собственный вектор для собственного значения

lsolve(A, b) - решение СЛАУ **Ax=b**

MATLAB

STATISTICA

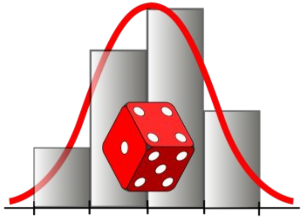
СООТВЕТСТВИЕ ТЕРМИНОЛОГИИ

Обозначение	Теория множеств	Теория вероятностей
Ω	Множеств, пространство	Пространство элементарных событий
ω	Элемент множества	Элементарное событие
A, B	Подмножество А, В	Случайное событие А, В
$A+B = A \cup B$	Объединение множеств	Сумма событий
AB	Пересечение множеств	Произведение событий
$\neg A$	Дополнение множества	Противоположное событие
A/B	Разность множеств	Разность событий
$\varnothing$	Пустое множество	Невозможное событие
$AB=\varnothing$	Множества не пересекаются	События несовместны
$A=B$	Множества равны	События равносильны
$A \subset B$	А есть подмножество В	Событие А влечет событие В

СЛУЧАЙНЫЕ ВЕЛИЧИНЫ И ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ

Пространство элементарных событий

- Любая природа
- Не обязательно числовая



Случайная величина – числовая функция на пространстве элементарных событий

Числовые характеристики случайных величин

Математическое ожидание - $\Sigma x_i p_i$

Дисперсия - $D\xi=M(\xi-M\xi)^2$

Ковариация - $\mathbf{cov}(\xi,\eta)=M(\xi-M\xi)(\eta-M\eta)$

Коэффициент корреляции $\rho_{\xi,\eta}=\dfrac{\mathbf{cov}(\xi,\eta)}{\sqrt{D\xi\cdot D\eta}}$

Свойства математического ожидания

$\xi\geq 0 \Rightarrow M\xi\geq 0$;
$Ma\xi=aM\xi$
$M(\xi+\eta)=M\xi+M\eta$;
$M(\xi\eta)=M\xi M\eta$ - для независимых случайных величин
$M(|\xi\eta|)^2\leq M\xi M\eta$ - Неравенство Коши-Буняковского

Свойства дисперсии и ковариации

$D\xi=M\xi^2-(M\xi)^2$,
$D(a+b\xi)=b^2 D\xi$
$\mathbf{cov}(\xi,\eta)=M(\xi\eta)-M(\xi)M(\eta)$

ИГРА В КОСТИ + ПЕТЕРБУРГСКИЙ ПАРАДОКС

СЛУЧАЙНЫЕ ВЕКТОРЫ

Случайный вектор

$$\xi = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \dots \\ \xi_n \end{bmatrix} \quad \xi = [\xi_1 \quad \xi_2 \quad \quad \xi_n] \quad \xi = \begin{bmatrix} \xi_{11} & \xi_{12} \\ \xi_{21} & \xi_{22} \end{bmatrix}$$

Математическое ожидание

$$\boldsymbol{M}\,\xi = \begin{bmatrix} \boldsymbol{M}\,\xi_1 \\ \boldsymbol{M}\,\xi_2 \\ \dots \\ \boldsymbol{M}\,\xi_n \end{bmatrix}$$

Дисперсионная матрица ξ - **D**ξ=$||cov(\xi_i, \xi_j)||_{ij=1..n}$

Свойства **D**ξ: **D**ξ – симметрична (**D=D**^{T})
Dξ - неотрицательно-определенная (**∀X: X^{T}DX≥**0)

D - дисперсионная матрица ⇒ **D=A^{T}A**

Представление D: **D**ξ = ***M***[(ξ-***M***ξ)(ξ-***M***ξ)^T]

Некоторые доказательства

$$\sum_{i,j} \boldsymbol{cov}(\xi_i, \xi_j) x_i x_j = \sum_{i,j} \boldsymbol{M}(\xi_i - M\xi_i)(\xi_j - M\xi_j) \cdot x_i \cdot x_j = \boldsymbol{M} \sum_{i,j} [(\xi_i - M\xi_i) \cdot x_i]^2 \geq 0$$

D=A^{T}A ⇒ D^T = (A^{T}A)^T = A^T(A^T)^T = A^{T}A = D

D=A^{T}A ⇒ ∀X: X^{T}DX = X^T(A^{T}A)X = (X^{T}A^T)(AX) = (AX)^T(AX) = Y^{T}Y = Σy$_i^2$ ≥0

ПРЕДСТАВЛЕНИЕ ДИСПЕРСИОННОЙ МАТРИЦЫ

$$\boldsymbol{M}\left[(\xi-\boldsymbol{M}\,\xi)(\xi-\boldsymbol{M}\,\xi)^T\right]=\begin{bmatrix}\xi_1-\boldsymbol{M}\,\xi_1\\ \dots\\ \xi_i-\boldsymbol{M}\,\xi_i\\ \dots\\ \xi_n-\boldsymbol{M}\,\xi_n\end{bmatrix}\begin{bmatrix}\xi_1-\boldsymbol{M}\,\xi_1 & \dots & \xi_j-\boldsymbol{M}\,\xi_j & \dots & \xi_n-\boldsymbol{M}\,\xi_n\end{bmatrix}=$$

$$=\boldsymbol{M}\begin{bmatrix}(\xi_1-\boldsymbol{M}\,\xi_1)^2 & \dots & (\xi_1-\boldsymbol{M}\,\xi_1)(\xi_j-\boldsymbol{M}\,\xi_j) & \dots & (\xi_1-\boldsymbol{M}\,\xi_1)(\xi_n-\boldsymbol{M}\,\xi_n)\\ \dots & \dots & \dots & \dots & \dots\\ (\xi_i-\boldsymbol{M}\,\xi_i)(\xi_1-\boldsymbol{M}\,\xi_1) & \dots & (\xi_i-\boldsymbol{M}\,\xi_i)(\xi_j-\boldsymbol{M}\,\xi_j) & \dots & (\xi_i-\boldsymbol{M}\,\xi_i)(\xi_n-\boldsymbol{M}\,\xi_n)\\ \dots & \dots & & \dots & \dots\\ (\xi_n-\boldsymbol{M}\,\xi_n)(\xi_1-\boldsymbol{M}\,\xi_1) & \dots & (\xi_n-\boldsymbol{M}\,\xi_n)(\xi_j-\boldsymbol{M}\,\xi_j) & \dots & (\xi_n-\boldsymbol{M}\,\xi_n)^2\end{bmatrix}=$$

$$=\begin{bmatrix}\boldsymbol{M}(\xi_1-\boldsymbol{M}\,\xi_1)^2 & \dots & \boldsymbol{M}(\xi_1-\boldsymbol{M}\,\xi_1)(\xi_j-\boldsymbol{M}\,\xi_j) & \dots & \boldsymbol{M}(\xi_1-\boldsymbol{M}\,\xi_1)(\xi_n-\boldsymbol{M}\,\xi_n)\\ \dots & \dots & \dots & \dots & \dots\\ \boldsymbol{M}(\xi_i-\boldsymbol{M}\,\xi_i)(\xi_1-\boldsymbol{M}\,\xi_1) & \dots & \boldsymbol{M}(\xi_i-\boldsymbol{M}\,\xi_i)(\xi_j-\boldsymbol{M}\,\xi_j) & \dots & \boldsymbol{M}(\xi_i-\boldsymbol{M}\,\xi_i)(\xi_n-\boldsymbol{M}\,\xi_n)\\ \dots & \dots & & \dots & \dots\\ \boldsymbol{M}(\xi_n-\boldsymbol{M}\,\xi_n)(\xi_1-\boldsymbol{M}\,\xi_1) & \dots & \boldsymbol{M}(\xi_n-\boldsymbol{M}\,\xi_n)(\xi_j-\boldsymbol{M}\,\xi_j) & \dots & \boldsymbol{M}(\xi_n-\boldsymbol{M}\,\xi_n)^2\end{bmatrix}=\|\mathbf{cov}(\xi_i,\xi_j)\|=\boldsymbol{D}\,\xi$$

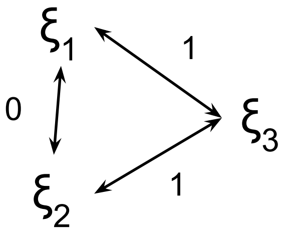
МАТРИЧНАЯ АЛГЕБРА И НЕКОТОРЫЕ УТВЕРЖДЕНИЯ

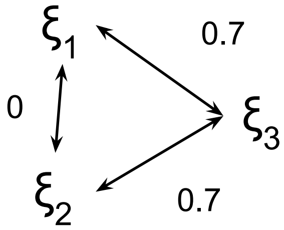
C=A*B $\quad c_{ij} = \sum_{k=1} a_{ik} b_{kj}$

$(A*B)^T = B^T*A^T$

Линейное преобразование математического ожидания случайного вектора равно математическому ожиданию линейного преобразования B (***M***ξ) = ***M*** (Bξ)

Взаимные корреляции





Критерий неотрицательно определенной матрицы

ЛИНЕЙНОЕ ПРЕОБРАЗОВАНИЕ СЛУЧАЙНОГО ВЕКТОРА

Задан вектор ξ

$$\xi = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \dots \\ \xi_n \end{bmatrix} \quad \mathbf{A} = \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_n \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ . & . & . & . \\ b_{m1} & b_{m2} & \dots & b_{mn} \end{bmatrix}$$

ξ - случайный вектор

A и **B** – числовые матрицы

$D_\xi = \boldsymbol{D}\xi = ||\boldsymbol{cov}(\xi_i, \xi_j)||_{i,j=1..n}$ - дисперсионная матрица ξ

$M_\xi = \boldsymbol{M}\xi$ - математическое ожидание ξ

Случайный вектор **η** представляет линейное преобразование вектора ξ

$$\begin{bmatrix} \eta_1 \\ \eta_2 \\ \dots \\ \eta_m \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_n \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ . & . & . & . \\ b_{m1} & b_{m2} & \dots & b_{mn} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \dots \\ \xi_n \end{bmatrix} \Rightarrow \boldsymbol{\eta = A + B\,\xi} \Rightarrow$$

Задача: найти

$\mathbf{M}_\eta$ - математическое ожидание η

$\mathbf{D}_{\eta\eta} = D\eta = ||cov(\eta_i, \eta_j)||_{i,j=1\dots m}$

Утверждение $\quad \mathbf{M}_\eta = \mathbf{A} + \mathbf{B}M_\xi \qquad \mathbf{D}_\eta = \mathbf{B}\mathbf{D}_\xi\mathbf{B}^T$

$\boldsymbol{M}\eta = \boldsymbol{M}(A+B\xi) = \boldsymbol{M}A + \boldsymbol{M}(B\xi) = A + B(\boldsymbol{M}\xi)$

$\boldsymbol{D}\eta = \boldsymbol{M}[(\eta-\boldsymbol{M}\eta)(\eta-\boldsymbol{M}\eta)^T] = \boldsymbol{M}[(A+B\xi-\boldsymbol{M}(A+B\xi))(A+B\xi-\boldsymbol{M}(A+B\xi))^T] = \boldsymbol{M}[(B\xi-\boldsymbol{M}B\xi)(B\xi-\boldsymbol{M}B\xi)^T] = =$

$\boldsymbol{M}[B(\xi-\boldsymbol{M}\xi)(B(\xi-\boldsymbol{M}\xi))^T] = \boldsymbol{M}[B(\xi-\boldsymbol{M}\xi)(\xi-\boldsymbol{M}\xi)^T B^T] = B\boldsymbol{M}[(\xi-M\xi)(\xi-M\xi)^T]B^T = B\boldsymbol{D}\xi B^T.$

ТЕОРЕМА О НОРМАЛЬНОЙ КОРРЕЛЯЦИИ

Оптимальное оценивание нормально-распределенных случайных векторов

$$\xi = (\xi_1, \ldots, \xi_m)^T \longrightarrow \eta = (\eta_1, \ldots, \eta_m)^T$$

Минимизация суммы квадратов

Совместный случайный вектор $(\xi_1, \ldots, \xi_m, \eta_1, \ldots, \eta_m)^T$

$$M\begin{bmatrix} \xi \\ \eta \end{bmatrix} = \begin{bmatrix} m_\xi \\ m_\eta \end{bmatrix}$$ - математическое ожидание совместного вектора

$$D\begin{bmatrix} \xi \\ \eta \end{bmatrix} = \begin{bmatrix} D_{\xi\xi} & D_{\xi\eta} \\ D_{\eta\xi} & D_{\eta\eta} \end{bmatrix}$$ - дисперсионная матрица совместного вектора

$$D_{\eta\eta} \equiv \left\| \mathbf{cov}(\eta_i, \eta_j) \right\| \qquad 1 \le i, j \le m$$

$$D_{\eta\xi} \equiv \left\| \mathbf{cov}(\eta_i, \xi_j) \right\| \qquad 1 \le i \le m, \ 1 \le j \le n$$

$$D_{\xi\xi} \equiv \left\| \mathbf{cov}(\xi_i, \xi_j) \right\| \qquad 1 \le i, j \le n$$

Утверждение

Оценка вектора η

$$M(\eta / \xi) = m_\eta + D_{\eta\xi} D_{\xi\xi}^{-1} (\xi - m_\xi)$$

Ошибка оценка вектора η

$$\Delta = D_{\eta\eta} - D_{\eta\xi} D_{\xi\xi}^{-1} D_{\xi\eta}$$

ТЕОРЕМА О НОРМАЛЬНОЙ КОРРЕЛЯЦИИ (ДОКАЗАТЕЛЬСТВО)

$$\zeta(\xi,\eta)=\left(\eta-m_\eta\right)-D_{\eta\xi}D_{\xi\xi}^{-1}\left(\xi-m_\xi\right)$$

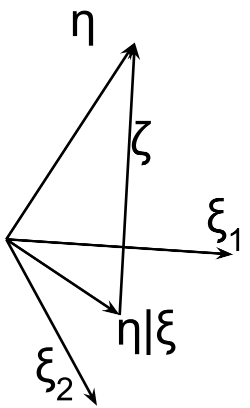
$$M\zeta=M\left(\eta-m_\eta\right)-M\left[D_{\eta\xi}D_{\xi\xi}^{-1}\left(\xi-m_\xi\right)\right]=0$$

$$D_{\zeta\xi}=M\left[\left(\zeta-m_\zeta\right)\left(\xi-m_\xi\right)^T\right]=0$$

Доказать

Вывод: ξ и ζ некоррелированы ⇒ в силу нормальности распределения - независимы

Оптимальная оценка ζ по ξ это $M\zeta=0$ - 1-ое утверждение доказано



$$D_{\zeta\zeta}=M\left[\zeta\zeta^T\right]=D_{\eta\eta}-D_{\eta\xi}D_{\xi\xi}^{-1}D_{\xi\eta}$$

Доказать

ВЫВОД ФОРМУЛ ТЕОРЕМЫ О НОРМАЛЬНОЙ КОРРЕЛЯЦИИ

$$D_{\zeta\xi} = M\left[(\zeta - m_\zeta)(\xi - m_\xi)^T\right] = M\left[\zeta(\xi - m_\xi)\right] = M\left[\left[(\eta - m_\eta) - D_{\eta\xi} D_{\xi\xi}^{-1}(\xi - m_\xi)^T\right](\xi - m_\xi)\right] =$$

$$= M\left[\left[(\eta - m_\eta)(\xi - m_\xi)\right] - D_{\eta\xi} D_{\xi\xi}^{-1}(\xi - m_\xi)^T(\xi - m_\xi)\right] = M\left[(\eta - m_\eta)(\xi - m_\xi)\right] - D_{\eta\xi} D_{\xi\xi}^{-1}\left[M(\xi - m_\xi)^T(\xi - m_\xi)\right] =$$

$$= D_{\eta\xi} - D_{\eta\xi} D_{\xi\xi}^{-1} D_{\xi\xi} = 0$$

$$D_{\zeta\zeta} = M\left[\zeta\zeta^T\right] = \left[(\eta - m_\eta) - D_{\eta\xi} D_{\xi\xi}^{-1}(\xi - m_\xi)\right]\left[(\eta - m_\eta) - D_{\eta\xi} D_{\xi\xi}^{-1}(\xi - m_\xi)\right]^T =$$

$$= M\left[\begin{array}{l} (\eta - m_\eta)(\eta - m_\eta)^T - (\eta - m_\eta)\left[D_{\eta\xi} D_{\xi\xi}^{-1}(\xi - m_\xi)\right]^T - \\ - D_{\eta\xi} D_{\xi\xi}^{-1}(\xi - m_\xi)(\eta - m_\eta)^T + D_{\eta\xi} D_{\xi\xi}^{-1}(\xi - m_\xi)\left[D_{\eta\xi} D_{\xi\xi}^{-1}(\xi - m_\xi)\right]^T \end{array}\right] =$$

$$= M\left[(\eta - m_\eta)(\eta - m_\eta)^T\right] - M(\eta - m_\eta)(\xi - m_\xi)^T D_{\xi\xi}^{-1} D_{\xi\eta} -$$

$$- M D_{\eta\xi} D_{\xi\xi}^{-1}(\xi - m_\xi)(\eta - m_\eta)^T + M D_{\eta\xi} D_{\xi\xi}^{-1}(\xi - m_\xi)(\xi - m_\xi)^T D_{\xi\xi}^{-1} D_{\xi\eta}$$

$$= D_{\eta\eta} - D_{\eta\xi} D_{\xi\xi}^{-1} D_{\xi\eta} - D_{\eta\xi} D_{\xi\xi}^{-1} D_{\xi\eta} + D_{\eta\xi} D_{\xi\xi}^{-1} D_{\xi\xi} D_{\xi\xi}^{-1} D_{\eta\xi}^T =$$

$$= D_{\eta\eta} - D_{\eta\xi} D_{\xi\xi}^{-1} D_{\xi\eta}$$

СЛУЧАЙ НЕЗАВИСИМЫХ ПРЕДИКТОРОВ И СКАЛЯРНОЙ ПРОГНОЗИРУЕМОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ

$(\xi_1,\ldots,\xi_n,\eta)^T$ (n+1)-мерный гауссовский вектор, причем ξ_1, ξ_2 … ξ_n попарно независимы *cov*(ξ_i, ξ_j)=0, а η является одномерной величиной,

НАЗВАНИЕ СЛАЙДА

НАЗВАНИЕ СЛАЙДА

MATHCAD

 - определение размеров матрицы

X_n - ввод нижнего индекса

x^{-1} - вычисление обратной матрицы

$|x|$ - определитель матрицы $|\mathbf{M}|=\det\mathbf{M}$ или длина вектора $|x|$

$\longrightarrow$- поэлементные операции с матрицами

$\mathbf{M}^{<>}$ - $\mathbf{M}^{<j>}$ j-й столбец матрицы

$\mathbf{M}^{T}$ - транспонирование матрицы

$\hat{u} \cdot \hat{e}$ - скалярное произведение векторов $x*y=\Sigma x_i y_i$

$\hat{u}\times\hat{e}$ - векторное произведение $a\times b=(a_2b_3-a_3b_2,\ a_3b_1-a_1b_3,\ a_1b_2-a_2b_1)$;

Σx - сумма компонент вектора $\Sigma x=\Sigma x_i$

m..n - диапазон изменения переменной

НАЗВАНИЕ СЛАЙДА

$$D_{\zeta\xi} = M\left[(\zeta - m_\zeta)(\xi - m_\xi)^T\right] = M\left[\zeta(\xi - m_\xi)\right] =$$
$$= M\left[\left[(\eta - m_\eta) - D_{\eta\xi} D_{\xi\xi}^{-1}(\xi - m_\xi)^T\right](\xi - m_\xi)\right] =$$
$$= M\left[\left[(\eta - m_\eta)(\xi - m_\xi)\right] - D_{\eta\xi} D_{\xi\xi}^{-1}(\xi - m_\xi)^T(\xi - m_\xi)\right] =$$
$$= M\left[(\eta - m_\eta)(\xi - m_\xi)\right] - D_{\eta\xi} D_{\xi\xi}^{-1}\left[M(\xi - m_\xi)^T(\xi - m_\xi)\right] =$$
$$= D_{\eta\xi} - D_{\eta\xi} D_{\xi\xi}^{-1} D_{\xi\xi}$$

$$D_{\zeta\zeta} = M\left[\zeta\zeta^T\right] = \left[(\eta - m_\eta) - D_{\eta\xi} D_{\xi\xi}^{-1}(\xi - m_\xi)\right]\left[(\eta - m_\eta) - D_{\eta\xi} D_{\xi\xi}^{-1}(\xi - m_\xi)\right]^T =$$
$$= M\begin{bmatrix} (\eta - m_\eta)(\eta - m_\eta)^T - (\eta - m_\eta)\left[D_{\eta\xi} D_{\xi\xi}^{-1}(\xi - m_\xi)\right]^T - \\ - D_{\eta\xi} D_{\xi\xi}^{-1}(\xi - m_\xi)(\eta - m_\eta)^T + D_{\eta\xi} D_{\xi\xi}^{-1}(\xi - m_\xi)\left[D_{\eta\xi} D_{\xi\xi}^{-1}(\xi - m_\xi)\right]^T \end{bmatrix} =$$
$$= M\left[(\eta - m_\eta)(\eta - m_\eta)^T\right] - M(\eta - m_\eta)(\xi - m_\xi)^T D_{\xi\xi}^{-1} D_{\xi\eta} -$$
$$- M D_{\eta\xi} D_{\xi\xi}^{-1}(\xi - m_\xi)(\eta - m_\eta)^T + M D_{\eta\xi} D_{\xi\xi}^{-1}(\xi - m_\xi)(\xi - m_\xi)^T D_{\xi\xi}^{-1} D_{\xi\eta}$$
$$= D_{\eta\eta} - D_{\eta\xi} D_{\xi\xi}^{-1} D_{\xi\eta} - D_{\eta\xi} D_{\xi\xi}^{-1} D_{\xi\eta} + D_{\eta\xi} D_{\xi\xi}^{-1} D_{\xi\xi} D_{\xi\xi}^{-1} D_{\eta\xi}^T =$$
$$= D_{\eta\eta} - D_{\eta\xi} D_{\xi\xi}^{-1} D_{\eta\xi}^T$$