

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ

федеральное государственное бюджетное образовательное учреждение
высшего образования

**«МОСКОВСКИЙ АВТОМОБИЛЬНО-ДОРОЖНЫЙ
ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ (МАДИ)»
ВОЛЖСКИЙ ФИЛИАЛ МАДИ**

УТВЕРЖДАЮ
Зав. кафедрой ИиТТП
Петрова А.В.

**МЕТОДИЧЕСКИЕ УКАЗАНИЯ К ЛАБОРАТОРНЫМ РАБОТАМ
ПО ДИСЦИПЛИНЕ
«СИСТЕМНЫЙ АНАЛИЗ И ПРИНЯТИЕ РЕШЕНИЙ»**

Направление подготовки
09.03.01 «Информатика и вычислительная техника»

Направленность (профиль)
***Автоматизированные системы обработки информации и
управления в отраслях транспортно-дорожного комплекса***

Квалификация
Бакалавр

Кафедра: ИиТТП

Чебоксары

Лабораторная работа 1. Метод Монте-Карло и генерация основных статистических распределений.

Программы генерации в Matcad

Примеры данного пособия сопровождаются анализом выборок сгенерированных в пакете Mathcad, которые импортированные в систему Statistica.

```
N := 100  N1 := rnorm( N + 1, 2, 1)
```

```
U1 := for j ∈ 0.. N
      | Xj ← N1j if N1j > 0
      | Xj ← 0 otherwise
      X
```

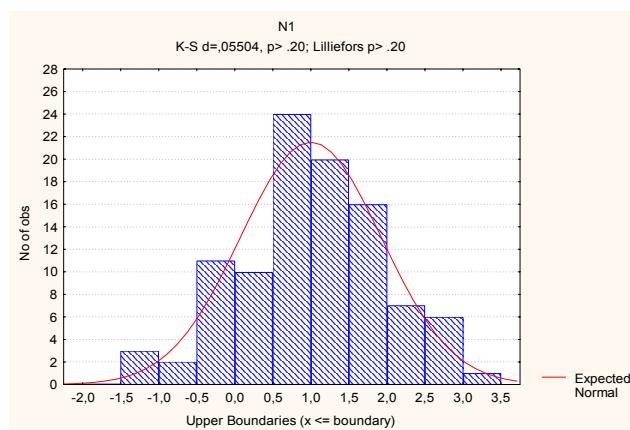
N1 =

	0
0	1.725
1	0.612
2	2.062
3	2.433
4	2.968
5	3.626

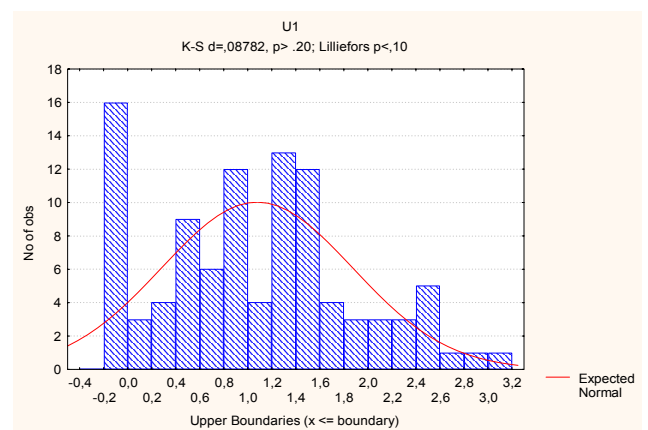
U1 =

	0
0	1.725
1	0.612
2	2.062
3	2.433
4	2.968
5	3.626

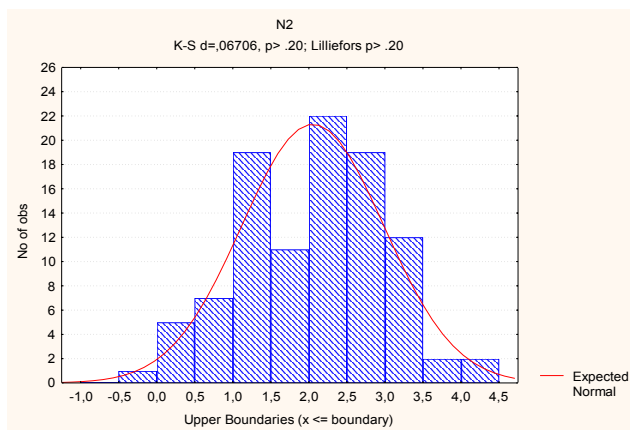
Сгенерировано четыре переменных **N1**, **U1**, **N2**, **U2**. Переменные **N1** **N2** сгенерированы по нормальному закону с параметрами (1,1) и (2, 1) соответственно. Переменные **U1** **U2** являются преобразованиями переменных **N1** **N2**, у которых все отрицательные значения заменены нулевыми.



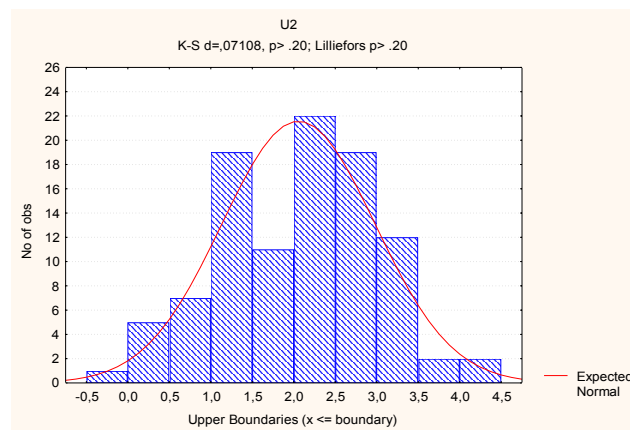
Гистограмма N1



Гистограмма U1



Гистограмма N2



Гистограмма U2

Покажем, что **N1**, **N2** соответствуют нормальному закону распределения, а **U1**, **U2** усеченному нормальному закону распределения в пункте «Функции распределения» «Подбор распределения»

Из приведенных гистограмм видно, что наборы N1 и N2 имеют распределения близкие к нормальному, так как они являются выборками из нормального распределения. Гистограмма набора U2 также близка к нормальной. Это объясняется тем, что математическое ожидание исходного (неусеченного) распределения велико и обнуляется лишь незначительная доля выборочных значений. В то же время видно, что для набора U1 гистограмма уже не похожа на гистограмму нормального распределения. Это объясняется малым значением математического ожидания исходного распределения.

Задание по построению гистограмм

1. Построить графики функции и плотности распределения для известного закона распределения с заданными параметрами (закон распределения и параметры указаны в таблице вариантов).
2. Сгенерировать выборки из заданного распределения длиной 10, 50, 100, 150, 200, 250, 300. Сохранить соответствующие выборки в файле с переменными R10, R50, ..., R300.
3. Построить гистограммы и оценить параметры сгенерированных выборок. Оценка параметров выполняется при подборе распределения и в описательной статистике. Сформировать вектор оценок параметров в зависимости от длины. (Последней строкой данных является 400, и теоретические значения параметра).
4. Построить X-Y график оценок параметров от длины выборки.
5. Дать обоснование полученных результатов и их сравнение с теоретическими.
6. Вычислить вероятность того, что случайная величина будет иметь значение, большее, чем в таблице варианта.
7. Для указанного в таблице вариантов значения функции распределения найти значение аргумента.

Содержание отчета:

- Графики функции и плотности вероятности.
- Гистограммы распределений выборок.
- Таблицы описательных статистик для выборок.
- X-Y графики оценок параметров распределений в зависимости от объема.
- Значение функции и аргумента.

Таблица вариантов:

Закон распр.	Парам.1	Парам.2	Аргумент	Функция
Нормальное	3	7	2.7	0.93

Лабораторная работа 2. Линейное преобразование случайных векторов. Генерация случайного вектора с заданной дисперсионной матрицей.

Программа генерации случайного вектора

$i := 0..2$

$\zeta^{<i>} := \text{rnorm}(100, 0, 1)$

$\zeta =$

	0	1	2
0	0.06	-2.052	0.538
1	-0.498	-1.125	-0.128
2	0.056	0.526	-0.721
3	-0.82	0.422	0.983
4	-0.047	-0.591	-0.524

$\eta := \zeta \cdot A$

$\eta =$

	0	1	2
0	-1.897	-0.772	0.379
1	-1.151	-1.018	-0.357
2	0.782	-0.136	-0.42
3	-0.541	0.378	0.146
4	-0.285	-0.701	-0.351

Для генерации трехмерного нормального вектора с вектором средних

$$m_{\xi} = (0, 0, 0)^T$$

и дисперсионной матрицей

$$D_{\xi} = \begin{bmatrix} 1 & 0.8 & 0.5 \\ 0.8 & 1 & 0 \\ 0.5 & 0 & 1 \end{bmatrix}$$

необходимо найти линейное преобразование **A**, удовлетворяющее соотношению

$$AA^T = D$$

Поиск матрицы **A** выполняется программа MathCad, приведенная на рис.13.

a11 := 1 a12 := 1 a13 := 1 a21 := 1 a22 := 1 a23 := 1 a31 := 1 a32 := 1 a33 := 1

Given

$$\begin{aligned} a_{11} \cdot a_{11} + a_{12} \cdot a_{12} + a_{13} \cdot a_{13} &= 1 \\ a_{11} \cdot a_{21} + a_{12} \cdot a_{22} + a_{13} \cdot a_{23} &= 0.8 \\ a_{11} \cdot a_{31} + a_{12} \cdot a_{32} + a_{13} \cdot a_{33} &= 0.5 \\ a_{21} \cdot a_{11} + a_{22} \cdot a_{12} + a_{23} \cdot a_{13} &= 0.8 \\ a_{21} \cdot a_{21} + a_{22} \cdot a_{22} + a_{23} \cdot a_{23} &= 1 \\ a_{21} \cdot a_{31} + a_{22} \cdot a_{32} + a_{23} \cdot a_{33} &= 0 \\ a_{31} \cdot a_{11} + a_{32} \cdot a_{12} + a_{33} \cdot a_{13} &= 0.5 \\ a_{31} \cdot a_{21} + a_{32} \cdot a_{22} + a_{33} \cdot a_{23} &= 0 \\ a_{31} \cdot a_{31} + a_{32} \cdot a_{32} + a_{33} \cdot a_{33} &= 1 \end{aligned}$$

$$\begin{bmatrix} x1 \\ x2 \\ x3 \\ x4 \\ x5 \\ x6 \\ x7 \\ x8 \\ x9 \end{bmatrix} := \text{Find}(a_{11}, a_{12}, a_{13}, a_{21}, a_{22}, a_{23}, a_{31}, a_{32}, a_{33})$$

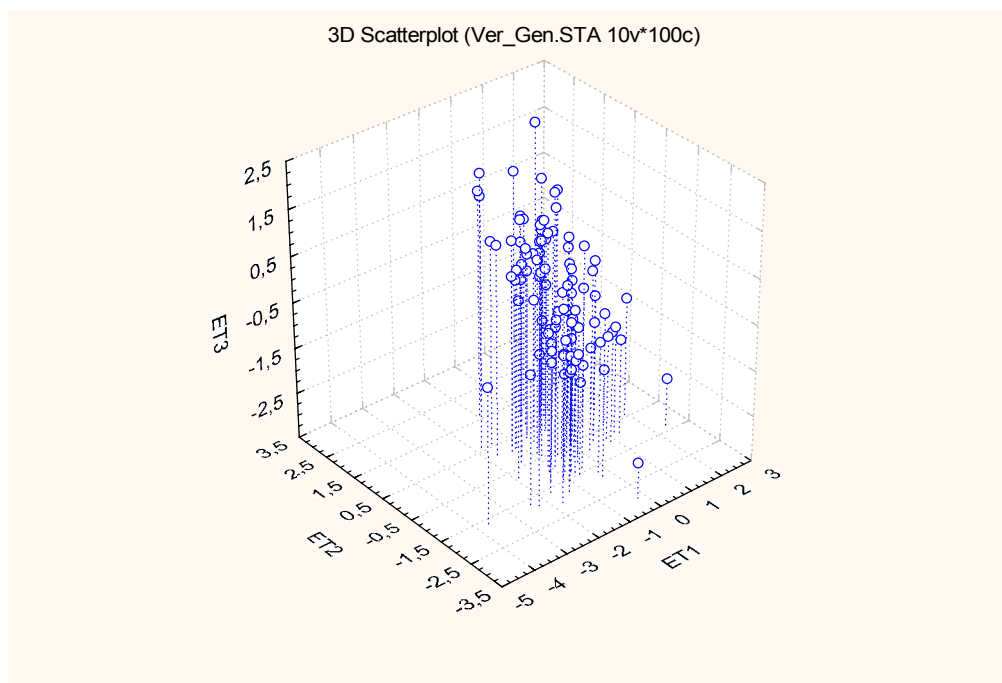
$$\begin{bmatrix} x1 \\ x2 \\ x3 \\ x4 \\ x5 \\ x6 \\ x7 \\ x8 \\ x9 \end{bmatrix} = \begin{bmatrix} 0.559 \\ 0.604 \\ 0.567 \\ 0.826 \\ 0.564 \\ -4.604 \cdot 10^{-3} \\ -0.438 \\ 0.647 \\ 0.624 \end{bmatrix}$$

$$A := \begin{bmatrix} x1 & x2 & x3 \\ x4 & x5 & x6 \\ x7 & x8 & x9 \end{bmatrix} \quad A = \begin{bmatrix} 0.559 & 0.604 & 0.567 \\ 0.826 & 0.564 & -4.604 \cdot 10^{-3} \\ -0.438 & 0.647 & 0.624 \end{bmatrix} \quad A \cdot A^T = \begin{bmatrix} 1 & 0.8 & 0.5 \\ 0.8 & 1 & 0 \\ 0.5 & 0 & 1 \end{bmatrix}$$

После нахождения матрицы линейного преобразования выполним генерацию значений вектора с независимыми компонентами и затем его линейное преобразование. Программа генерации случайного вектора и его линейного преобразования приведена на рис.14.

Графическое представление нормального вектора

После экспорта сгенерированных данных в Statistica, построим объемный график последовательности. На рис.15. приведен объемный график, полученный в результате выполнения процедуры '3D XYZ Graph > Scatterplot'.



Решение системы уравнений в матричной форме

$x1 := 1 \quad x2 := 10 \quad x3 := 10 \quad x4 := 1$

Given

$$\begin{pmatrix} x1 & x2 \\ x3 & x4 \end{pmatrix} \cdot \begin{pmatrix} x1 & x2 \\ x3 & x4 \end{pmatrix}^T = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} a1 & a2 \\ a3 & a4 \end{pmatrix} := \text{Find}(x1, x2, x3, x4)$$

$$B := \begin{pmatrix} a1 & a2 \\ a3 & a4 \end{pmatrix} \quad \underline{C} := B \cdot B^T \quad C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$B = \begin{pmatrix} 0.16 & 0.987 \\ 0.987 & -0.16 \end{pmatrix}$$

$$\begin{aligned}
 &XX := \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix} \\
 &\text{Given} \quad XX \cdot XX^T = \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix} \\
 &XX := \text{Find}(XX) \\
 &XX = \begin{pmatrix} -0.302 & 0.953 \\ 0.873 & 0.487 \end{pmatrix} \quad XX \cdot XX^T = \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}
 \end{aligned}$$

Предварительно задаются начальные значения $a_{31} := 1 \quad a_{32} := 1 \quad a_{33} := 1$ для организации итерационной числовой процедуры расчета корней системы. Оператор **Given** определяет начало области задания системы нелинейных уравнений, а оператор **Find** конец области. При этом вектору в левой части присваиваются значения рассчитанных корней системы уравнений.

Лабораторная работа 3. Метод прогнозирования случайных векторов на основе теоремы о нормальной корреляции

Частный случай теоремы о нормальной корреляции

$$D\xi\xi = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, a \quad D\eta\xi = (0.7 \ 0.7)$$

$$XY0^{(i)} := \text{morm}(N, 0, 1) \quad XY0 =$$

	1	2	3
1	-3.029	0.611	1.16
2	0.366	1.317	...

$$XY := XY0 \cdot A^T \quad XY =$$

	1	2	3
1	-2.934	0.814	-1.664
2	0.024	2.156	...

$$D\xi\xi := \text{submatrix}(D, 1, 2, 1, 2) \quad D\xi\xi = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$D\eta\xi := \text{submatrix}(D, 3, 3, 1, 2) \quad D\eta\xi = (0.7 \ 0.7)$$

$$L := D\eta\xi \cdot D\xi\xi^{-1} \quad L = (0.7 \ 0.7)$$

$$X := \text{submatrix}(XY, 1, N, 1, 2) \quad Y_{pr} := X \cdot L^T \quad \Delta := XY^{(3)} - Y_{pr}$$

$$XY^{(4)} := Y_{pr} \quad XY^{(5)} := \Delta$$

$$XY =$$

	1	2	3	4	5
1	-2.934	0.814	-1.664	-1.484	-0.18
2	0.024	2.156	1.532	1.526	...

Переменная	Описательные статистики (Таблица данных5)						
	N набл.	Среднее	Медиана	Минимум	Максим.	Дисперсия	Ст.откл.
x1	100	-0,202	-0,306	-2,895	2,101	0,882	0,939
x2	100	-0,102	-0,092	-2,987	2,437	1,076	1,037
y	100	-0,202	-0,264	-2,774	2,411	1,069	1,034
y_p	100	-0,213	-0,360	-2,722	2,548	1,041	1,020
d_y	100	0,011	0,033	-0,324	0,382	0,022	0,149

Переменная	Корреляции (НормКорр)				
	x1	x2	y	y_p	d_y
x1	1,000	0,086	0,685	0,705	-0,076
x2	0,086	1,000	0,770	0,767	0,092
y	0,685	0,770	1,000	0,990	0,161
y_p	0,705	0,767	0,990	1,000	0,017
d_y	-0,076	0,092	0,161	0,017	1,000

Программа генерации модели прогноза для пятимерной совместной модели

ORIGIN := 1 N := 100

i := 1..5 m := 1..4 j := 1..4 A_{i,j} := rnd(1)

$$A = \begin{pmatrix} 1.268 \times 10^{-3} & 0.193 & 0.585 & 0.35 \\ 0.823 & 0.174 & 0.71 & 0.304 \\ 0.091 & 0.147 & 0.989 & 0.119 \\ 8.923 \times 10^{-3} & 0.532 & 0.602 & 0.166 \\ 0.451 & 0.057 & 0.783 & 0.52 \end{pmatrix} \quad D := A \cdot A^T$$

$$D = \begin{pmatrix} 0.502 & 0.557 & 0.649 & 0.513 & 0.652 \\ 0.557 & 1.305 & 0.839 & 0.578 & 1.095 \\ 0.649 & 0.839 & 1.021 & 0.694 & 0.886 \\ 0.513 & 0.578 & 0.694 & 0.673 & 0.592 \\ 0.652 & 1.095 & 0.886 & 0.592 & 1.09 \end{pmatrix}$$

j1 := 1..5

$$R_{i,j1} := \frac{D_{i,j1}}{\sqrt{D_{i,i}} \cdot \sqrt{D_{j1,j1}}}$$

$$R = \begin{pmatrix} 1 & 0.688 & 0.905 & 0.883 & 0.881 \\ 0.688 & 1 & 0.727 & 0.617 & 0.918 \\ 0.905 & 0.727 & 1 & 0.837 & 0.839 \\ 0.883 & 0.617 & 0.837 & 1 & 0.692 \\ 0.881 & 0.918 & 0.839 & 0.692 & 1 \end{pmatrix}$$

XY0^(m) := morm(N, 0, 1) XY := XY0 · A^T

	1	2	3	4	5
1	-1.668	-2.257	-2.912	-1.965	-2.228
2	0.482	0.292	0.638	1.115	0.164
3	-0.604	-1.367	-0.938	-0.294	-1.336
4	-0.446	1.021	-0.067	0.064	...

Dξξ := submatrix(D, 1, 3, 1, 3)

Dηξ := submatrix(D, 4, 5, 1, 3)

Dηη := submatrix(D, 3, 5, 3, 5) Bprog := Dηξ · Dξξ⁻¹ = $\begin{pmatrix} 0.806 & -0.018 & 0.182 \\ 0.84 & 0.565 & -0.131 \end{pmatrix}$

xx := submatrix(XY, 1, N, 1, 3) yy := xx · Bprog^T

XY⁽⁶⁾ := yy⁽¹⁾ XY⁽⁷⁾ := yy⁽²⁾ XY⁽⁸⁾ := XY⁽⁶⁾ - XY⁽⁴⁾ XY⁽⁹⁾ := XY⁽⁷⁾ - XY⁽⁷⁾

	1	2	3	4	5
1	-1.668	-2.257	-2.912	-1.965	-2.228
2	0.482	0.292	0.638	1.115	...

Переменная	Описательные статистики (НормКорр)				
	Среднее	Минимум	Максим.	Дисперсия	Ст.откл.
x1	-0,181	-1,857	1,393	0,466	0,683
x2	-0,279	-3,052	2,185	1,302	1,141
x3	-0,252	-2,854	2,354	0,892	0,944
y1	-0,130	-2,117	1,885	0,524	0,724
y2	-0,307	-2,831	2,398	1,125	1,061
y1_p	-0,187	-1,962	1,410	0,483	0,695
y2_p	-0,277	-2,912	2,106	1,003	1,002
d_y1	-0,056	-0,838	0,897	0,135	0,367
d_y2	0,030	-0,470	0,439	0,037	0,192

Переменная	Корреляции (НормКорр)								
	Отмеченные корреляции значимы на уровне $p < 0,05000$ N=100 (Построчное удаление ПД)								
	x1	x2	x3	y1	y2	y1_p	y2_p	d_y1	d_y2
x1	1,000	0,661	0,899	0,868	0,869	0,995	0,887	0,171	-0,171
x2	0,661	1,000	0,745	0,533	0,921	0,678	0,931	0,233	-0,233
x3	0,899	0,745	1,000	0,788	0,863	0,937	0,872	0,221	-0,221
y1	0,868	0,533	0,788	1,000	0,642	0,867	0,743	-0,331	0,331
y2	0,869	0,921	0,863	0,642	1,000	0,874	0,984	0,390	-0,389
y1_p	0,995	0,678	0,937	0,867	0,874	1,000	0,891	0,183	-0,183
y2_p	0,887	0,931	0,872	0,743	0,984	0,891	1,000	0,221	-0,220
d_y1	0,171	0,233	0,221	-0,331	0,390	0,183	0,221	1,000	-1,000
d_y2	-0,171	-0,233	-0,221	0,331	-0,389	-0,183	-0,220	-1,000	1,000

Задание по прогнозу нормально распределенной случайной величины

1. Сформировать данные математического ожидания и дисперсионной матрицы, указанных в таблице вариантов в переменных **M** и **D** соответственно.
2. Сгенерировать трехмерный нормальный вектор с заданной дисперсионной матрицей и математическим ожиданием. Переименуйте столбцы сгенерированной матрицы в Z, X и Y соответственно.
3. Построить объемный X-Y-Z график сгенерированных данных. Сохранить график.
4. Построить теоретическую плоскость прогноза. Сохранить график.
5. Наложить график с плоскостью на график с выборочными точками.
6. Вычислить вектор прогноза ZP по значениям X и Y.
7. Найти вектор ошибки DZ=Z-ZP. Построить гистограмму вектора ошибки. Оценки параметров подобранного распределения.

Содержание отчета:

- Расчетные значения коэффициентов прогноза и ошибки.
- Объемный график точек и поверхности прогноза.
- Гистограмма ошибки и оценка параметров ошибки.
- Обоснование полученных результатов.

Таблица вариантов:

1	Мат.	Дисп		
1	1	0.8	0.5	
2	0.8	1	0	
3	0.5	0	1	

Лабораторная работа 4. Генерация данных по модели простой регрессии. Взаимосвязь дисперсии ошибки и коэффициента корреляции простой регрессии

Для простой регрессии для генерации данных воспользуемся классической моделью

$$y_i = a + bx_i + \varepsilon_i$$

где y - зависимая переменная;

x - независимая переменная;

ε - ошибка.

Пусть область определения ограничена $[0; 10]$ и значения x равномерно распределены по всей области $0 \leq x \leq 10$.

Пусть $a=1$; $b=1$; $M\varepsilon_i=0$; $M\varepsilon_i=1$.

Этот вектор содержит 101 значение, равномерно распределенные на интервале $[0, 10]$.

Для генерации модели простой регрессии используем MathCad-программу (REG1), приведенную на рис.21.

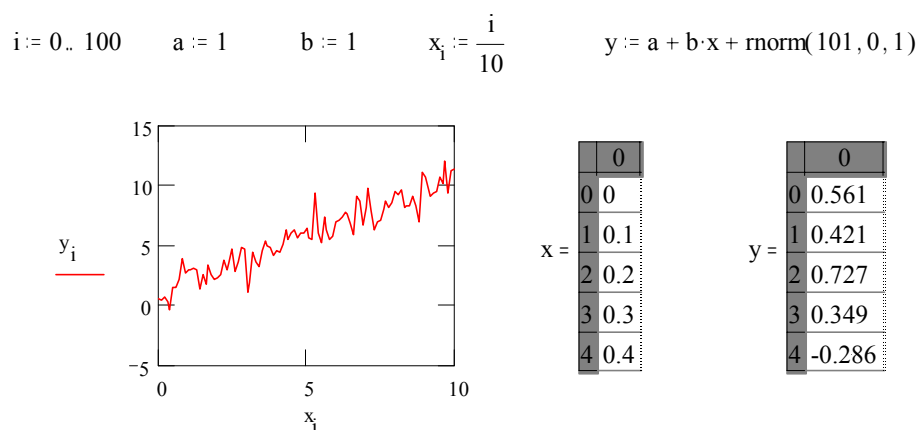


Рис.21. Генерация модели простой регрессии.

Таблицы полученных значений x и y экспортируем в Statistica (файл REG1). В качестве наименований переменных зададим X и YL .

Задание:

1. Сгенерировать нормально распределенный вектор ошибки E с параметрами и объемом (M , D , N), указанными в таблице вариантов. Построить гистограмму E и оценить параметры выборки.
2. Сформировать вектор независимой переменной X , значения которой равномерно распределены на интервале с границами A и B .
3. Вычислить значения вектора Y по соотношению $Y=a+bX+E$ где a и b заданы в соответствии с вариантом.

4. Выполнить процедуру простой линейной регрессии Y по X , и дать прогноз величины Y для $X=(A+B)/2$.
5. Сохранить остатки в переменной S .
6. Построить гистограмму остатков и оценить параметры распределения остатков.
7. Дать теоретическое обоснование полученных результатов.

Содержание отчета:

- График регрессии.
- Таблица дисперсионного анализа для простой регрессии.
- График регрессии с прогнозом.
- График остатков.
- Гистограмма остатков с оценкой параметров.
- Гистограмма E с оценок параметров.

Таблица вариантов имеет вид:

Вар.	N	D	A	B	a	b
1	60	5	3	25	2	6

При генерации варианта выдается окно, содержащее значения элементов таблицы вариантов .

- Таблица оценок параметров и анализа дисперсий.
- График остатков.
- Гистограмма остатков с оценкой параметров.
- Гистограмма E с оценок параметров.

Таблица вариантов:

Вар.	N	D	B1	B2	a	b	c
1	60	5	3	25	2		
2							

Лабораторная работа 6. Генерация данных по моделям нелинейной регрессии

Общая схема модели нелинейной регрессии имеет вид

$$y_i = f(\beta, x_i) + \varepsilon_i$$

где f - произвольная функция;

β - параметры регрессии;

ε - ошибки.

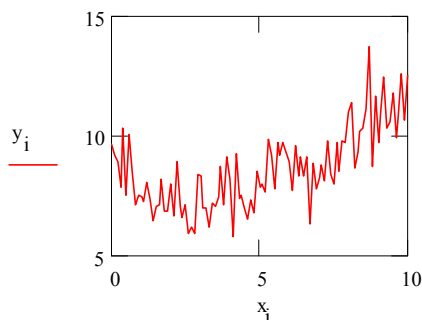
В качестве исследуемой функции возьмем $y = 20 \frac{e^{0.2 \cdot x}}{x + 2}$ на интервале $0 \leq x \leq 10$, на которую наложен шум ε .

В качестве искоемых параметров регрессии будем использовать $y = 20 \frac{e^{\beta_1 \cdot x}}{x + \beta_2}$

где теоретические значения параметров регрессии соответственно равны $\beta_1 = 0.2$, $\beta_2 = 2$

Для генерации вектора используем MathCad-программу (REGN), приведенную на рис.22.

$$i := 0..100 \quad x_i := \frac{i}{10} \quad y_i := 20 \cdot \frac{\exp(0.2 \cdot x_i)}{x_i + 2} \quad y := y + \text{rnorm}(101, 0, 1)$$



$$x =$$

0
0 0
1 0.1
2 0.2
3 0.3
4 0.4

$$y =$$

0
0 9.67
1 9.17
2 8.935
3 7.839
4 10.364

Из рисунка виден явно нелинейный характер модели, что делает невозможным использование простой и множественной регрессии для оценки параметров регрессии.

Пошаговая регрессия

Для генерации будем использовать факторную модель многомерного анализа.

Переменные **n1, n2, e1, e2, e3, e4, e5, e6** генерируются в пункте "Генерация случайных чисел"

$$X_i = a_{i1}n_1 + a_{i2}n_2 + b e_i \quad \text{или в векторной форме } \mathbf{X} = \mathbf{A}\mathbf{n} + \mathbf{b}\mathbf{e}$$

В данном случае для преобразования возьмем матрицу

$$\mathbf{A} = \begin{bmatrix} -5 & -1 & -9 & 5 & 1 & 9 \\ 5 & 9 & 1 & 5 & 9 & 1 \end{bmatrix}^T \quad b = 3$$

Для генерации данных используем MathCad-программу (REGS), приведенную на рис.24.

$$\text{ORIGIN} := 1 \quad iv := 1..2 \quad ie := 1..6 \quad A := \begin{bmatrix} -5 & -1 & -9 & 5 & 1 & 9 \\ 5 & 9 & 1 & 5 & 9 & 1 \end{bmatrix} \quad b := 3$$

$$VM^{<iv>} := \text{rnorm}(100, 0, 1) \quad EM^{<ie>} := \text{rnorm}(100, 0, 1) \quad XM := VM \cdot A + b \cdot EM$$

$$VM =$$

	1	2
1	-0.439	0.438
2	-0.679	1.76
3	-0.473	-0.647
4	-0.951	0.551
5	-1.686	1.546
6	0.044	-1.822
7	-0.121	1.308
8	0.556	-1.049
9	2.192	0.459
10	0.809	-1.052
11	0.985	-1.242

$$EM =$$

	1	2	3	4	5	6
1	-0.627	0.06	-2.052	0.538	-0.026	-0.885
2	-0.553	-0.498	-1.125	-0.128	1.76	0.159
3	-0.346	0.056	0.526	-0.721	-1.507	-1.804
4	-0.464	-0.82	0.422	0.983	1.65	-1.167

$$XM =$$

	1	2	3	4	5	6
1	2.502	4.558	-1.766	1.609	3.423	-6.167
2	10.539	15.028	4.501	5.021	20.443	-3.878
3	-1.908	-5.183	5.19	-7.766	-10.82	-10.318
4	6.119	3.451	10.379	0.946	8.958	-11.512
5	16.914	15.462	14.943	-2.268	12.871	-11.949

$$i := 1..6 \quad j := 1..6$$

$$D := A^T \cdot A \quad DK_{i,j} := \frac{D_{i,j}}{\sqrt{D_{i,i} \cdot D_{j,j}}}$$

$$DK =$$

1	0.781	0.781	0	0.625	-0.625
0.781	1	0.22	0.625	0.976	0
0.781	0.22	1	-0.625	0	-0.976
0	0.625	-0.625	1	0.781	0.781
0.625	0.976	0	0.781	1	0.22
-0.625	0	-0.976	0.781	0.22	1

$$i := 1..100 \quad Y_i := \sum_{j=1}^6 XM_{i,j} + \text{rnorm}(100, 0, 2)_1$$

$$Y =$$

	1
1	3.883
2	50.126
3	-31.136

На последнем рисунке также приведено теоретическое значение корреляционной матрицы (DK) для приведенной модели. Экспериментальный анализ этой зависимости проводится в разделе 'Пошаговая регрессия'.

Задание – нелинейная регрессия

А. Сгенерировать модель данных для нелинейной регрессии.

1. Сгенерировать нормально распределенный вектор ошибки **E** с параметрами и объемом (M, D, N), указанными в таблице вариантов. Построить гистограмму **E** и оценить параметры выборки.

2. Сформировать вектор независимой переменной **X1** значения которого равномерно распределены в области (A,B) с границами, указанными в таблице вариантов.

3. Вычислить значения вектора **Y** по соотношению $Y = f(b, X) + E$, где **f** и **b** заданы в соответствии с вариантом.

4. Сгенерировать вектор ошибки **E1**, с неравномерной дисперсией.

5. Сформировать $Y1 = f(b, X) + E1$.

В. Провести анализ данных на основании нелинейной регрессии для функции Y.

1. Оценить параметры регрессии при различных начальных параметрах регрессии, заданных в таблице.
2. Построить таблицу анализа дисперсий.
3. Вывести график модели и график нормальной вероятности.
4. Получить график остатков, сохранить остатки и построить гистограмму остатков с оценкой параметров распределения.
5. Сравнить остатки с истинами.

С. Дать теоретическое обоснование полученных результатов.

Содержание отчета:

- График модели.
- Таблица оценок параметров и анализа дисперсий.
- График остатков.
- Гистограмма остатков с оценкой параметров.
- Гистограмма E и E1 с оценкой параметров.

Таблица вариантов:

Вар.	F	D	A	B	b1	b2	b начальное
1	$20 \cdot e^{(b1 \cdot x)/(b2+x)}$	2	3	25	2	4	$b1-0.5; b2-1.5$
2							

Задание - Пошаговая регрессия

Математическая постановка задачи:

Пусть ξ результирующая векторная случайная величина $\xi = (\xi_1, \xi_2, \dots, \xi_6)^T$, которая задается линейным преобразованием $\xi_i = a_{i1}v_1 + a_{i2}v_2 + \varepsilon_i$.

В векторной форме это преобразование имеет вид $\xi = A \cdot v + b \cdot \varepsilon$

В нашем случае для преобразования возьмем матрицу

$$A = \begin{bmatrix} -5 & -1 & -9 & 5 & 1 & 9 \\ 5 & 9 & 1 & 5 & 9 & 1 \end{bmatrix}^T \quad b = 3.$$

В соответствии с этим выборочные значения величин ξ_i преобразуются по правилу

$$X_i = a_{i1}n_1 + a_{i2}n_2 + 3 \cdot e_i,$$

где X_i - вектор выборочных значений случайной величины ξ_i .

Используя лемму о линейном преобразовании случайных величин и в силу независимости v_i и ε_i , дисперсионная матрица вектора x будет равна

$$D\xi = \begin{bmatrix} -5 & 5 \\ -1 & 9 \\ -9 & 1 \\ 5 & 5 \\ 1 & 9 \\ 9 & 1 \end{bmatrix} \cdot \begin{bmatrix} -5 & 5 \\ -1 & 9 \\ -9 & 1 \\ 5 & 5 \\ 1 & 9 \\ 9 & 1 \end{bmatrix}^T + \begin{bmatrix} 3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 \end{bmatrix}^2$$

Дисперсия случайной величины ξ_i соответственно равна

$$D\xi_i = \sqrt{a_{i1}^2 + a_{i2}^2 + 3^2}.$$

В результате корреляционная матрица вектора с точностью до второго знака может быть записана

$$D\xi = \begin{bmatrix} 59 & 45 & 50 & 0 & 40 & -40 \\ & 91 & 18 & 40 & 80 & 0 \\ & & 91 & -40 & 0 & 80 \\ & & & 59 & 50 & 50 \\ & & & & 91 & 18 \\ & & & & & 91 \end{bmatrix}$$

В силу симметрии выписана лишь наддиагональная часть матрицы.

А. Сгенерировать модель данных по схеме факторного анализа.

1. Сгенерировать выборки нормально распределенных случайных величин $N_1, N_2, E_1, E_2, E_3, E_4, E_5$ и E_6 . Для всех выборок математическое ожидание равно 0, дисперсия - 1, а длина выборки 100. При генерации каждого вектора данных необходимо изменять начальные значения дачика генерации.

2. Сформировать значения векторов независимых переменных $X_1, \dots, X_6$ на основании соотношения $X_i = a_{i1} \cdot N_1 + a_{i2} \cdot N_2 + a_{i3} \cdot E_i$ где коэффициенты a_{ji} берутся из таблицы вариантов.

3. Вычислить значения зависимой переменной на основании соотношения $Y = \sum c_i \cdot X_i + c_7 \cdot E_7$.

В. Выполнить анализ данных на основании метода пошаговой регрессии.

1. В качестве зависимой переменной указать Y .
2. В качестве независимых переменных указать $X_1, \dots, X_6$.
3. Задавая статистики F-включения и F-исключения добиться того, чтобы в результирующей группе параметров осталось три независимых переменных.
4. Для выбранной группы параметров вычислить статистики пошаговой регрессии.
5. На основе множественной регрессии для выбранной группы получить оценки параметров регрессии таблицу дисперсионного анализа.

С. Дать теоретическое обоснование полученных результатов.

Содержание отчета:

- Экран ввода пошаговой регрессии.
- Таблица статистик пошаговой регрессии для полной группы параметров.
- Таблица статистик для выбранной из трех элементов группы.
- Таблица оценок параметров регрессии и анализа дисперсий.

Таблица вариантов:

В таблице вариантов исходные данных для генерации модели данных указаны в соответствии со структурой, приведенной ниже:

Вар	a11	a12	a13	a14	a15	a16	
	a21	a22	a23	a24	a25	a26	
	a31	a32	a33	a34	a35	a36	
	c1	c2	c3	c4	c5	c6	c7

Лабораторная работа 7. Факторное планирование эксперимента.

Оптимальность планов. Полные и дробные факторные планы $2m-n$. Трехуровневые планы.

В задачах регрессионного и дисперсионного анализа предполагается, что исследователь имеет дело с уже полученными экспериментальными данными. Такие методы анализа относятся к пассивному эксперименту. В данном разделе рассматривается вопрос организации активного эксперимента, т.е. экспериментатор до проведения эксперимента выбирает значения управляемых параметров, при которых он будет проводить измерения.

Программа генерации плана

```
MF := iw ← 1
      for i1 ∈ -1, 1
        for i2 ∈ -1, 1
          for i3 ∈ -1, 1
            MFWiw,1 ← i1
            MFWiw,2 ← i2
            MFWiw,3 ← i3
            MFWiw,4 ← 4 + 2·i1 + 4·i2 + 6·i3 + norm(1, 0, 1)1
            iw ← iw + 1
      MFW
```

$$MF = \begin{bmatrix} -1 & -1 & -1 & -7.532 \\ -1 & -1 & 1 & 3.2 \\ -1 & 1 & -1 & -0.247 \\ -1 & 1 & 1 & 12.487 \\ 1 & -1 & -1 & -2.369 \\ 1 & -1 & 1 & 8.346 \\ 1 & 1 & -1 & 4.828 \\ 1 & 1 & 1 & 16.625 \end{bmatrix}$$

Построим регрессионные зависимости для двух планов

- В меню Statistics выберите Multiple Regression.
- В первом случае в качестве зависимой переменной указывается Y_1 , в качестве независимой - X_1 . Во втором случае зависимая и независимая переменные соответственно представляют Y_2 и X_2 .
- В окне Multiple Regression Results нажмите кнопку Summary: Regression results (табл.5.1.).
- В закладке Advanced нажмите кнопку ANOVA(Overall goodness of fit) (табл.5.2.).
- В окне Multiple Regression Results нажмите кнопку OK.
- В закладке Scatterplots нажмите кнопку Predicted vs. observed.

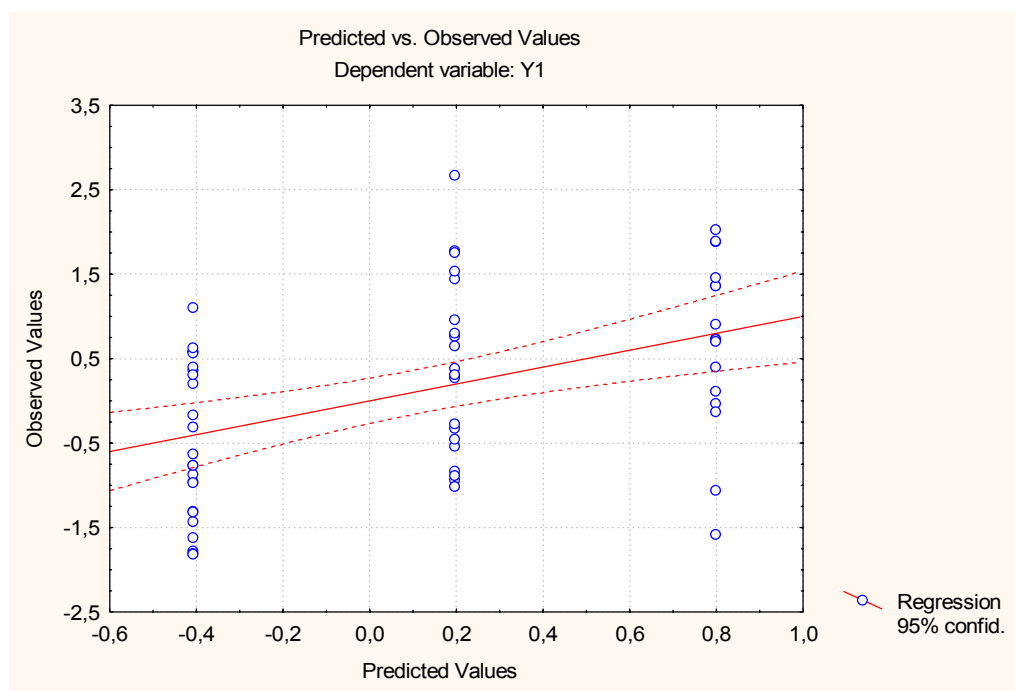
На рисунке приведен график наблюдаемых значений отклика относительно его прогноза для регрессионной зависимости Y_1 по X_1 .

В таблице приведены численные значения оценок параметров регрессии и их стандартных ошибок.

Оценки параметров регрессии для Y_1 по X_1

	B	St. Err. of B	t(58)	p-level
Intercept	,196041	,131192	1,494302	,140518
X1	,603183	,160677	3,754005	,000405

График прогноза регрессии Y_1 по X_1



Оценки коэффициентов корреляции для приведенной модели соответственно равны $R = ,44212965$ $R^2 = ,19547863$ $R^2 = ,18160757$ (Adjusted), а значение F-отношения $F = 14,093$ при уровне значимости $p < ,00040$. В табл.2 значения статистик дисперсионного анализа регрессионной модели, которые дает информацию о степени адекватности линейной модели.

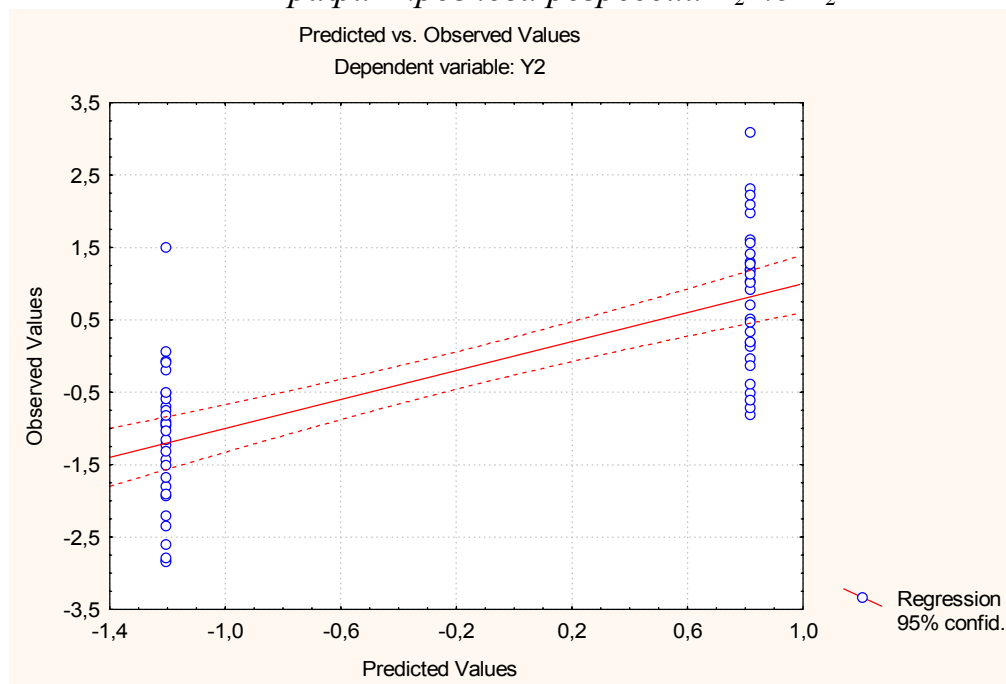
Таблица дисперсионного анализа модели Y_1 по X_1

	Sums of Squares	df	Mean Squares	F	p-level
Regress.	14,16511	1	14,16511	14,09255	,000405
Residual	58,29864	58	1,00515		
Total	72,46375				

Так, значение статистики F-отношения равное 14 (это не достаточно большое значение) говорит о том, что линейная модель может быть принята с большим сомнением.

На рисунке приведен график наблюдаемых значений отклика относительно предсказанных значений для регрессии Y_2 по X_2 .

График прогноза регрессии Y_2 по X_2



В таблице приведены оценки параметров модели. Значение свободного члена равно - 0.19 (вместо 0 по теоретической модели) и наклона - 1.01 (вместо 1). Для данного объема выборки это достаточно точные оценки. Кроме того, коэффициент корреляции этой модели равен 0.72, что существенно выше коэффициента корреляции первой модели (0.44).

Оценки параметров регрессии для Y_1 по X_2

	B	St. Err. of B	t(58)	p-level
Intercept	-,193667	,127877	-1,51448	,135335
X2	1,011667	,127877	7,91127	,000020

Оценки коэффициентов корреляции для приведенной модели соответственно равны $R = ,72043355$, $R^2 = ,51902450$, $R^2 = ,51073182$ (Adjusted), а значение F-отношения $F = 62,588$ при уровне значимости $p < ,00020$.

В таблице приведены статистики дисперсионного анализа регрессионной модели. Значение F-отношения в последней модели существенно выше и равно 62.5 в отличие от 14 в первой модели.

Таблица дисперсионного анализа модели Y_1 по X_1

	Sums of Squares	df	Mean Squares	F	p-level
--	-----------------	----	--------------	---	---------

Regress.	61,4082	1	61,40817	62,58826	,000210
Residual	56,9064	58	,98115		
Total	118,3146				

На основании полученных данных можно сделать вывод о том, что модель 2 (Y_2 по X_2) является более точной, так для этой модели коэффициент корреляции и значение F-отношения превышают соответствующие значения первой модели.

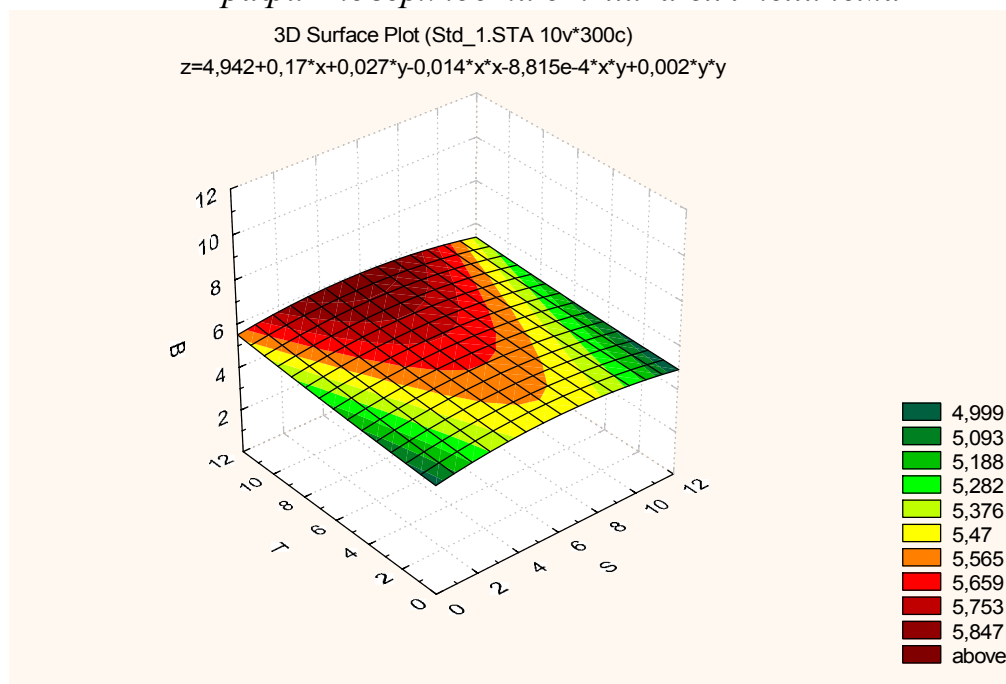
Графическое представление функции отклика выполняется в меню Graphs -> 3D XYZ Graphs -> Surface Plots...

Указывается поверхности и ее параметры. Полиномиальные зависимости XY -полинома предполагают вычисление коэффициентов β_i и β_{ij} . на основании регрессионных соотношений. Строится поверхность для полинома для подогнанных данных, которые являются исходными для аппроксимации. На рисунке представлен график поверхности полинома.

Модели факторных планов - это частные случаи общей линейной регрессионной модели. Вектор параметров и содержит суммарное среднее, главные эффекты и взаимодействия. Матрица плана F независимых переменных состоит из значений +1 и -1. Эксперимент планируется - это значит, что F выбирается так, чтобы оценки имели некоторые желательные свойства.

Формирование активных экспериментов не всегда удается реализовать в практике экономической деятельности предприятий. Эти эксперименты могут достаточно дорого стоить.

График поверхности отклика для полинома



Поэтому в данной главе воспользуемся модельными данными. Пусть регрессионная зависимость определяется уравнением $y=x$. Далее рассматриваются следующие наборы значений независимой переменной:

-1, -1,...-1, 0, 0, ..., 0, 1, 1, ... , 1
 -1, -1,...-1, 1, ... , 1

В первом наборе каждое значение повторяется по 30 раз, во втором - по 20. Генерируются две независимых выборки случайных величин из нормального распределения. Значения зависимой переменной являются суммой независимой переменной и случайной выборки. Для формирования значений независимых переменных X_1 и X_2 и зависимой Y используются операторы пакета *MathCad*, приведенные на рисунке.

Генерация переменных регрессионных моделей

```
ORIGIN := 1
i := 1..20      x1i := -1   i := 21..40      x1i := 0       i := 41..60      x1i := -1
i := 1..30      x2i := -1   i := 31..60      x2i := 1
i := 1..60
y1i := x1i + norm(1, 0, 1)i      y2i := x2i + norm(1, 0, 1)i
```

 $x1 =$

	1
1	-1
2	-1
3	-1
4	-1
5	-1

 $y1 =$

	1
1	-0.394
2	-0.29
3	-0.81
4	0.549
5	-0.607

 $x2 =$

	1
1	-1
2	-1
3	-1
4	-1
5	-1

 $y2 =$

	1
1	-0.963
2	-1.434
3	-1.512
4	-1.934
5	0.062

Переменные модели, сформированные на экране в таблице соответствующих переменных, затем необходимо экспортировать в пакет Statistica.

Сгенерируем трехфакторную модель для полного факторного эксперимента 2^3 . Имена переменных содержащих значения уровней фактора будут соответственно A , B и C . Стандартный порядок комбинаций уровней факторов предполагает следующее задание значений переменных:

A	-1	1	-1	1	-1	1	-1	1
B	-1	-1	1	1	-1	-1	1	1
C	-1	-1	-1	-1	1	1	1	1

Математическое ожидание отклика формируется на основе соотношения:

$$y = \beta_0 x_0 + \sum_{i=1}^3 \beta_i x_i + \sum_{i,j=1}^3 \beta_{ij} x_i x_j + \beta_{123} x_1 x_2 x_3.$$

Причем, в нашем случае будем считать, что есть всего одно взаимодействие β_{12}

β_0	β_1	β_2	β_3	β_{12}
4	2	4	6	7

Поэтому математическое ожидание отклика MF определяется:

$$MF = 4 + 2*A + 4*B + 6*C + 7*A*B$$

Случайный отклик F является суммой среднего MF и ошибки E. Вектор данных E формируется процедурой *"rnorm"*, имеет нормальное распределение и длину равную 8.

Лабораторная работа 8. Факторный анализ.

Генерации данных по модели факторного анализа

Формирование для моделей основано на линейном преобразовании случайных величин. Выборки независимых величин могут быть получены в результате выполнения процедуры "Генерация случайных чисел".

Пусть все случайные величины $\{v_i\}_{i=1}^2, \{\varepsilon_i\}_{i=1}^6$ имеют стандартное нормальное распределение с нулевым математическим ожиданием и единичной дисперсией.

Представим эти величины в виде векторов $v = (v_1, v_2)^T$ и $\varepsilon = (\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5, \varepsilon_6)^T$

Соответствующие выборки в пакете имеют идентификаторы n1, n2, e1, e2, e3, e4, e5, e6. Длины всех векторов (длины выборок) равны 100.

Пусть ξ результирующая векторная случайная величина $\xi = (\xi_1, \xi_2, \dots, \xi_6)^T$, которая задается линейным преобразованием $\xi_i = a_{i1}v_1 + a_{i2}v_2 + \varepsilon_i$.

В векторной форме это преобразование имеет вид $\xi = A \cdot v + b \cdot \varepsilon$

В нашем случае для преобразования возьмем матрицу

$$A = \begin{bmatrix} -5 & 5 \\ -1 & 9 \\ -9 & 1 \\ 5 & 5 \\ 1 & 9 \\ 9 & 1 \end{bmatrix} \quad b=3$$

В соответствии с этим выборочные значения величин ξ_i преобразуются по правилу

$$X_i = a_{i1}n_1 + a_{i2}n_2 + 3 \cdot e_i,$$

где X_i - вектор выборочных значений случайной величины ξ_i .

Используя лемму о линейном преобразовании случайных величин и в силу независимости v_i и ε_i , дисперсионная матрица вектора x будет равна

$$D_{\xi} = \begin{bmatrix} -5 & 5 \\ -1 & 9 \\ -9 & 1 \\ 5 & 5 \\ 1 & 9 \\ 9 & 1 \end{bmatrix} \cdot \begin{bmatrix} -5 & 5 \\ -1 & 9 \\ -9 & 1 \\ 5 & 5 \\ 1 & 9 \\ 9 & 1 \end{bmatrix}^T + \begin{bmatrix} 3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 \end{bmatrix}^2$$

Дисперсия случайной величины ξ_i соответственно равна

$$D\xi_i = \sqrt{a_{i1}^2 + a_{i2}^2 + 3^2}.$$

В результате корреляционная матрица вектора с точностью до второго знака может быть записана

$$D\xi = \begin{bmatrix} 59 & 45 & 50 & 0 & 40 & -40 \\ & 91 & 18 & 40 & 80 & 0 \\ & & 91 & -40 & 0 & 80 \\ & & & 59 & 50 & 50 \\ & & & & 91 & 18 \\ & & & & & 91 \end{bmatrix}$$

В силу симметрии выписана лишь наддиагональная часть матрицы.

Statistica. Вычисление общих и специфических факторов, а также факторных нагрузок выполняется в пункте "Factor Analysis". В качестве исследуемых переменных выступают X1, X2, X3, X4, X5, и X6.

При формировании этих величин в алгоритме генерации участвовало две общих величины n1 и n2. Специфическими факторами были e1, e2, e3, e4, e5, e6. Таким образом можно предположить, что данные, сгенерированные по модели факторного анализа величины, и данные, которые обработаны по факторной модели, будут очень близки.

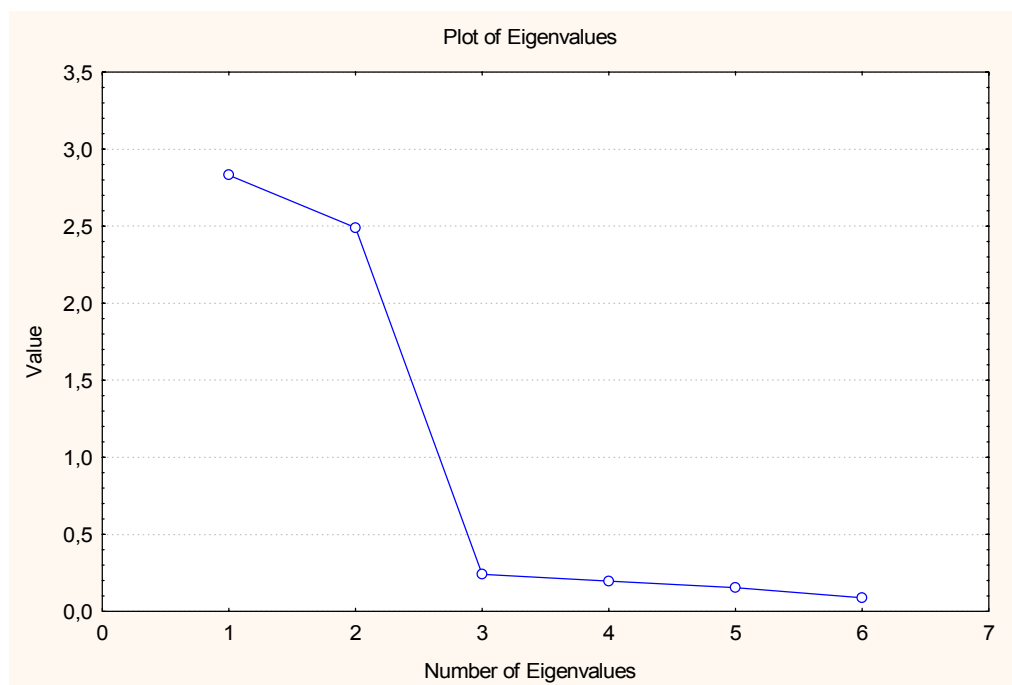
В таблице приведены результаты расчета процедуры факторного анализа.

Eigenvalues (fac1.sta) Extraction: Principal components

	Eigenval	% total Variance	Cumul.Eigenval	Cumul. %
1	2,830849	47,18081	2,830849	47,1808
2	2,490405	41,50675	5,321254	88,6876
3	,240472	4,00787	5,561726	92,6954
4	,196347	3,27245	5,758073	95,9679
5	,153829	2,56382	5,911902	98,5317
6	,088098	1,46830	6,000000	100,0000

Таблица показывает, что 88.7% накопленного процента дисперсии ложится на два фактора. При этом собственные числа для этих факторов существенно больше чем остальные 4. Это действительно говорит о совпадении модельных и статистических результатов. График собственных значений приведен на рисунке.

График собственных значений



Из графика также хорошо видно, что значимыми являются всего два фактора.

Весовые коэффициенты далеки от тех, которые использовались для генерации. Однако используя алгоритмы вращения, можно добиться близости значений. В таблице приведены значения коэффициентов корреляций факторов и анализируемых переменных факторов.

Factor Loadings (Quartimax normalized) (fac1.sta)

Extraction: Principal components (Marked loadings are > ,700000)

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
X1	,671430	-,611238	,022894	,002222	,416655	,037946
X2	,962894	-,127547	,011945	,034887	,004962	,234908
X3	,055768	-,959986	,011316	,270939	-,031691	,027874
X4	,706953	,497236	-,502444	-,005422	-,012892	,018120
X5	,922004	,186691	,090421	-,032808	-,100144	-,309473
X6	,168890	,937661	,016385	,300088	,030345	,032014
Expl.Var	2,759493	2,472758	,261689	,165789	,185746	,154525
Prp.Totl	,459915	,412126	,043615	,027631	,030958	,025754

На рисунке приведен график зависимости переменных от двух первых факторов.

График факторов

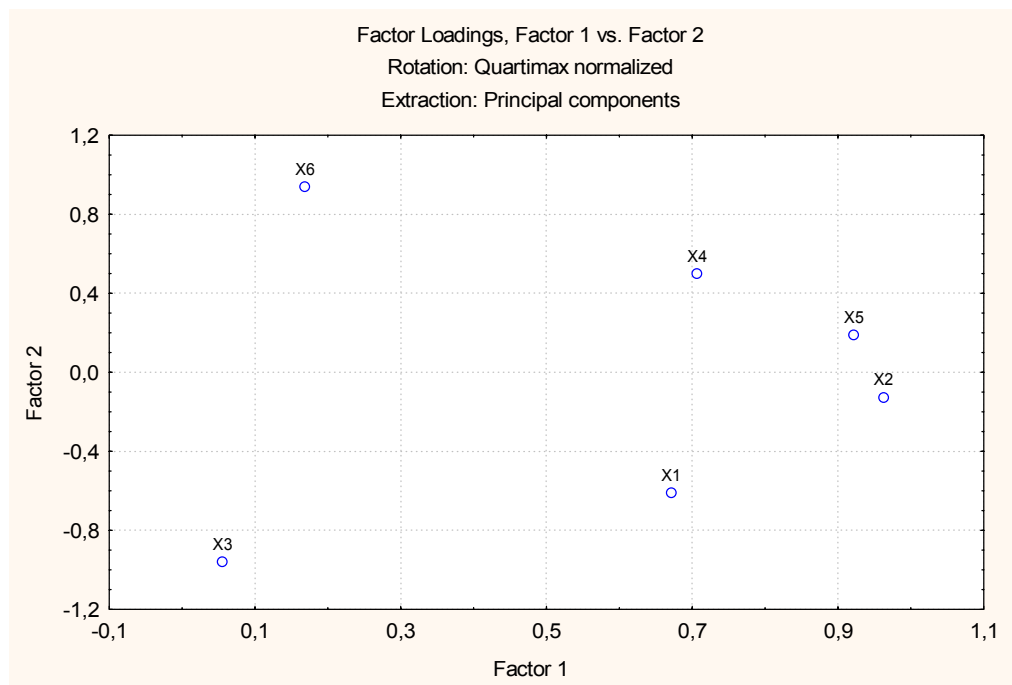


График показывает, что все переменные на основании первых двух факторов хорошо разнесены и в связи с этим могут хорошо идентифицироваться.

Задание - факторный анализ

А. Сгенерировать модель данных по схеме факторного анализа.

1. Сгенерировать выборки нормально распределенных случайных величин N_1 , N_2 , E_1 , E_2 , E_3 , E_4 , E_5 и E_6 . Для всех выборок математическое ожидание равно 0, дисперсия - 1, а длина выборки 100. При генерации каждого вектора данных необходимо изменять начальные значения датчика генерации.

2. Сформировать значения векторов независимых переменных $X_1, \dots, X_6$ на основании соотношения $X_i = a_{1i} \cdot N_1 + a_{2i} \cdot N_2 + a_{3i} \cdot E_i$, где коэффициенты a_{ji} берутся из таблицы вариантов.

В1. Выполнить процедуру факторного анализа.

1. Вычислить общности и коэффициенты факторов.
2. Построить график нагрузок и график меток факторов.
3. Сохранить общности.
4. Выполнить процедуру вращения. Найти матрицу перехода. Путем нескольких вращений добиться наибольшего соответствия коэффициентов общностей с теоретическими. Для вращения матрицы использовать различные методы.

С. Дать теоретическое обоснование полученных результатов.

Содержание отчета:

- Экран ввода главных компонент.
- Графики весов и рассеяния.
- Таблица накопленного процента.
- Таблица общностей факторного анализа.
- График нагрузок и меток факторов.

Таблица вариантов:

В таблице вариантов исходные данных для генерации модели данных указаны в соответствии со структурой, приведенной ниже:

Вар	a11	a12	a13	a14	a15	a16
	a21	a22	a23	a24	a25	a26
	b1	b2	b3	b4	b5	b6

Лабораторная работа 9. Кластерный анализ.

Генерации данных по модели кластерного анализа

Для методов дискриминантного и кластерного анализа будем использовать следующую модель.

$$f1 = k1 + n1; \quad f2 = k2 + n2,$$

где $n1, n2$ - выборки случайных величин v_1 и v_2 длиной 100, записанные в виде вектор-столбца;

$k1, k2$ - два вектора длины 100, которые формируются в пакете по алгоритму в системе MathCad.

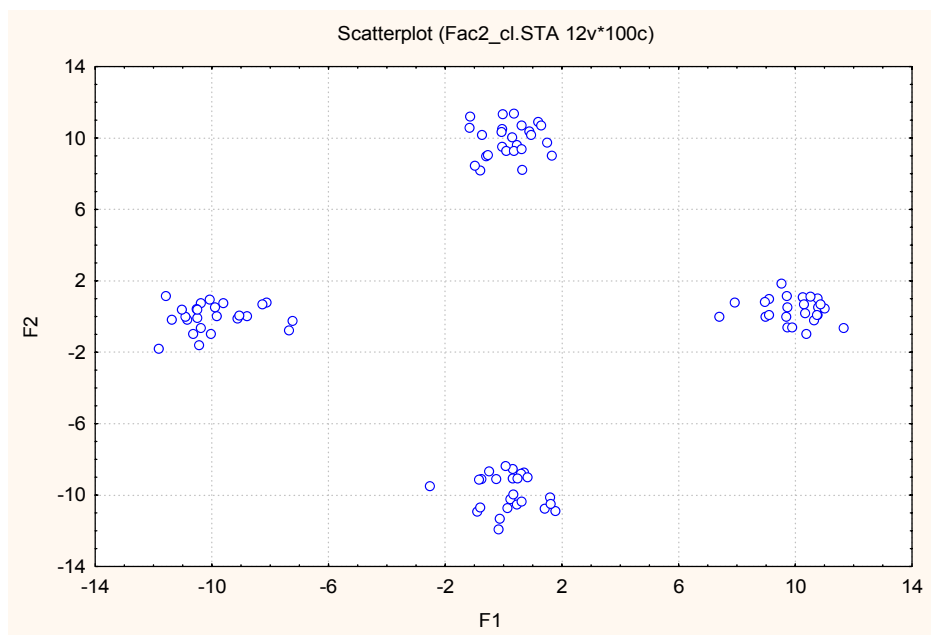
$$\begin{array}{lll} i := 1..25 & KM_{i,1} := 10 & KM_{i,2} := 0 \\ i := 26..50 & KM_{i,1} := 0 & KM_{i,2} := 10 \\ i := 51..75 & KM_{i,1} := -10 & KM_{i,2} := 0 \\ i := 76..100 & KM_{i,1} := 0 & KM_{i,2} := -10 \end{array}$$

Они моделируют смещение математических ожиданий для каждой 25 выборочных значений. Вектора выборочных значений случайной величины z

$$Z_i = a_{i1}f1 + a_{i2}f2 + 12 \cdot e_i.$$

X-Y-График величин $f1$ и $f2$ приведен на рисунке.

График значений генерирующих величин.



В результате сгенерированные данные дают возможность проводить на них статистический анализ, причем все теоретические характеристики считаются полностью известными.

Задание – кластерный анализ

А. Сгенерировать модель данных для кластеризации.

1. Сформировать две компоненты f_1 и f_2 с повторяющимися значениями из 16 значений, по 4 в каждой группе. Центры групп указаны в таблице вариантов.
2. Сгенерировать выборки нормально распределенных случайных величин $N_1, N_2, E_1, E_2, E_3, E_4, E_5$ и E_6 . Для всех выборок математическое ожидание равно 0, дисперсия - 1, а длина выборки 16. При генерации каждого вектора данных необходимо изменять начальные значения датчика генерации.
3. На центры групп наложить независимый шум N_1 и N_2 . $f_1=f_1 + N_1, f_2=f_2 + N_2$.
4. Сформировать значения векторов независимых переменных $Z_1, \dots, Z_6$ на основании соотношения $Z_i=a_{1i}*f_1 + a_{2i}*f_2 + a_{3i}*E_i$, где коэффициенты a_{ji} берутся из таблицы вариантов.

В. Выполнить процедуру кластерного анализа.

1. Ввести в качестве факторов кластеризации две компоненты вектора данных Z .
2. Получить таблицу кластеризации и проверить соответствие вычисленного разбиения с истинным.
3. Если разбиение не соответствует истинному, использовать для кластеризации другие возможные методы (А-средние, С-центр, F-дальние, Медиана, N-ближние, S-средние).
4. Если разбиение для указанных в п.2.3. методов не соответствует, выбрать другую пару компонент вектора Z и повторить кластерный анализ заново.
5. Построить график кластеров.

С. Дать теоретическое обоснование полученных результатов.

Содержание отчета:

- Экран ввода параметров кластерного анализа.
- Таблица кластеризации для первой пары компонент Z .
- График кластеров для первой пары Z .
- Таблица кластеризации для последней пары компонент Z .
- График кластеров для последней пары Z .

Таблица вариантов:

В таблице вариантов исходные данные для генерации модели данных указаны в соответствии со структурой, приведенной ниже:

Таблица коэффициентов

Таблица центров

Var	a11	a12	a13	a14	a15	a16	f11,f21	f12,f22
	a21	a22	a23	a24	a25	a26	f13,f23	f14,f24
	b1	b2	b3	b4	b5	b6		

Лабораторная работа 10. Дисперсионный анализ.

Моделирование данных для моделей дисперсионного анализа

Сгенерируем данные по схеме двухфакторного анализа. Каждый из факторов **A** и **B** будет иметь по три уровня, обозначенные соответственно **1 2 3**

Будем считать, что среднее значение отклика равно **2**.

На каждой комбинации уровней имеется 10 наблюдений.

Влияние каждого взаимодействия уровней факторов сведено в таблицу

$$\begin{bmatrix} 2 & 7 & 4 \\ 3 & 10 & 7 \\ 7 & 12 & 2 \end{bmatrix}$$

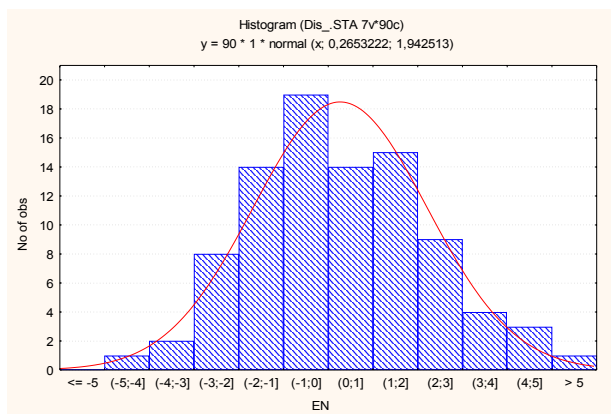
На рисунке приведена MathCad-программа генерации указанных функций.

Программа генерации модельных данных.

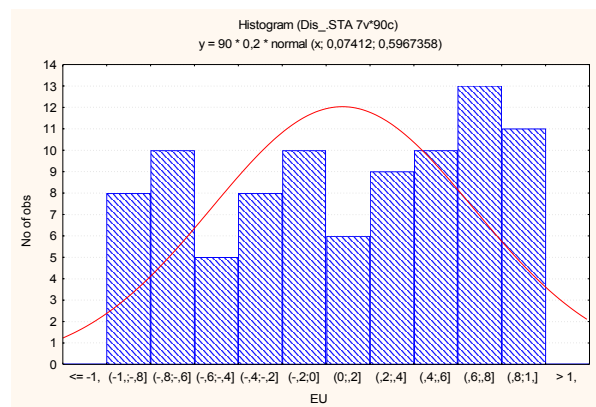
ORIGIN := 1	$A := \begin{bmatrix} 2 & 7 & 4 \\ 3 & 10 & 7 \\ 7 & 12 & 2 \end{bmatrix}$																																																																																																																																							
F_P := iw← 1 for i ∈ 1.. 3 for j ∈ 1.. 3 for k ∈ 1.. 10 FW_{1,iw}← i FW_{2,iw}← j FW_{3,iw}← rnorm(1, 0, 2)₁ FW_{4,iw}← runif(1, -1, 1)₁ FW_{5,iw}← 2 + A_{i,j} + FW_{3,iw} FW_{6,iw}← 2 + A_{i,j} + FW_{4,iw} FW_{7,iw}← iw iw← iw + 1 FW F_P^T = <table><tr><td></td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td></tr><tr><td>1</td><td>1</td><td>1</td><td>2.309</td><td>0.603</td><td>6.309</td><td>4.603</td><td>1</td></tr><tr><td>2</td><td>1</td><td>1</td><td>-3.053</td><td>-0.876</td><td>0.947</td><td>3.124</td><td>2</td></tr><tr><td>3</td><td>1</td><td>1</td><td>2.232</td><td>-0.663</td><td>6.232</td><td>3.337</td><td>3</td></tr><tr><td>4</td><td>1</td><td>1</td><td>1.994</td><td>0.3</td><td>5.994</td><td>4.3</td><td>4</td></tr><tr><td>5</td><td>1</td><td>1</td><td>1.311</td><td>0.753</td><td>5.311</td><td>4.753</td><td>5</td></tr><tr><td>6</td><td>1</td><td>1</td><td>3.351</td><td>0.696</td><td>7.351</td><td>4.696</td><td>6</td></tr><tr><td>7</td><td>1</td><td>1</td><td>-0.714</td><td>0.44</td><td>3.286</td><td>4.44</td><td>7</td></tr><tr><td>8</td><td>1</td><td>1</td><td>-1.875</td><td>0.631</td><td>2.125</td><td>4.631</td><td>8</td></tr><tr><td>9</td><td>1</td><td>1</td><td>0.366</td><td>0.308</td><td>4.366</td><td>4.308</td><td>9</td></tr><tr><td>10</td><td>1</td><td>1</td><td>-1.722</td><td>0.855</td><td>2.278</td><td>4.855</td><td>10</td></tr><tr><td>11</td><td>1</td><td>2</td><td>-1.455</td><td>0.45</td><td>7.545</td><td>9.45</td><td>11</td></tr><tr><td>12</td><td>1</td><td>2</td><td>4.182</td><td>-0.295</td><td>13.182</td><td>8.705</td><td>12</td></tr><tr><td>13</td><td>1</td><td>2</td><td>-0.333</td><td>-0.621</td><td>8.667</td><td>8.379</td><td>13</td></tr><tr><td>14</td><td>1</td><td>2</td><td>1.466</td><td>-0.284</td><td>10.466</td><td>8.716</td><td>14</td></tr><tr><td>15</td><td>1</td><td>2</td><td>0.281</td><td>0.095</td><td>9.281</td><td>9.095</td><td>15</td></tr><tr><td>16</td><td>1</td><td>2</td><td>1.119</td><td>0.33</td><td>10.119</td><td>9.33</td><td>16</td></tr></table>		1	2	3	4	5	6	7	1	1	1	2.309	0.603	6.309	4.603	1	2	1	1	-3.053	-0.876	0.947	3.124	2	3	1	1	2.232	-0.663	6.232	3.337	3	4	1	1	1.994	0.3	5.994	4.3	4	5	1	1	1.311	0.753	5.311	4.753	5	6	1	1	3.351	0.696	7.351	4.696	6	7	1	1	-0.714	0.44	3.286	4.44	7	8	1	1	-1.875	0.631	2.125	4.631	8	9	1	1	0.366	0.308	4.366	4.308	9	10	1	1	-1.722	0.855	2.278	4.855	10	11	1	2	-1.455	0.45	7.545	9.45	11	12	1	2	4.182	-0.295	13.182	8.705	12	13	1	2	-0.333	-0.621	8.667	8.379	13	14	1	2	1.466	-0.284	10.466	8.716	14	15	1	2	0.281	0.095	9.281	9.095	15	16	1	2	1.119	0.33	10.119	9.33	16
	1	2	3	4	5	6	7																																																																																																																																	
1	1	1	2.309	0.603	6.309	4.603	1																																																																																																																																	
2	1	1	-3.053	-0.876	0.947	3.124	2																																																																																																																																	
3	1	1	2.232	-0.663	6.232	3.337	3																																																																																																																																	
4	1	1	1.994	0.3	5.994	4.3	4																																																																																																																																	
5	1	1	1.311	0.753	5.311	4.753	5																																																																																																																																	
6	1	1	3.351	0.696	7.351	4.696	6																																																																																																																																	
7	1	1	-0.714	0.44	3.286	4.44	7																																																																																																																																	
8	1	1	-1.875	0.631	2.125	4.631	8																																																																																																																																	
9	1	1	0.366	0.308	4.366	4.308	9																																																																																																																																	
10	1	1	-1.722	0.855	2.278	4.855	10																																																																																																																																	
11	1	2	-1.455	0.45	7.545	9.45	11																																																																																																																																	
12	1	2	4.182	-0.295	13.182	8.705	12																																																																																																																																	
13	1	2	-0.333	-0.621	8.667	8.379	13																																																																																																																																	
14	1	2	1.466	-0.284	10.466	8.716	14																																																																																																																																	
15	1	2	0.281	0.095	9.281	9.095	15																																																																																																																																	
16	1	2	1.119	0.33	10.119	9.33	16																																																																																																																																	

Вектор данных **MF** представляет среднее значение функционалов.

Каждая из выборок имеет объем 90.



а) Нормальная ошибка



б) Равномерная ошибка

Выборки сохраним в переменных **N90** и **R90** соответственно. Далее выполним пункт описания статистик для сравнительного анализа выборок. В таблица приведены описательные статистики ошибок.

Описательные статистики выборок Descriptive Statistics (dis_.sta)

	Valid N	Mean	Median	Minimum	Maximum	Std.Dev.
EN	90	,265322	,225500	-4,28000	5,005000	1,942513
EU	90	,074120	,122000	-,99800	,992000	,596736

Из таблицы видно, что средние значения выборок близки к нулю, а дисперсии практически одинаковы, 5.059 и 5.057 соответственно.

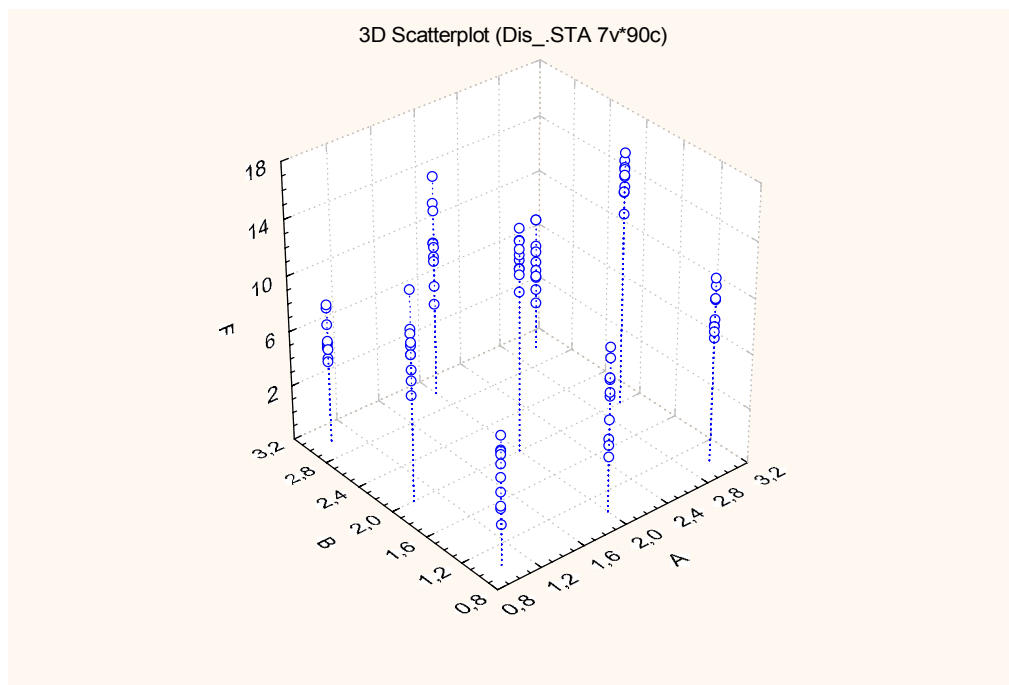
Функционалы **F** и **F1** формируются по правилу

$$F = MF + N90$$

$$F1 = MF + R90$$

На рисунке приведен объемный график значений функционала **F** на различных уровнях факторов **A** и **B**.

Объемный график **F** по **A** и **B**.



Последний рисунок дает наглядное представление о различии значений функционала на различных комбинациях уровней.

Задание - Дисперсионный анализ

A. Сгенерировать модель данных по схеме двухфакторного дисперсионного анализа.

1. Сформировать значения векторов независимых факторов А и В, каждый из которых может находиться на трех уровнях, обозначенных числами 1, 2, и 3.
2. Сгенерировать выборку нормально распределенной случайной величины N1. Для всех выборок математическое ожидание равно 0, дисперсия - D, а длина выборки 9*Р.
3. Ввести среднее значение отклика МУ по каждой комбинации факторов.
4. Сформировать повторяющиеся наблюдения. Повторение каждой комбинации факторов равно значению, указанному в таблице Р. Вычислить значения зависимой переменной на основании соотношения $Y = MY + N1$.

В. Выполнить анализ данных на основании метода дисперсионного анализа.

1. Выполнить процедуру однофакторного анализа, построить таблицу дисперсий и график средних по А и В.
2. Выполнить процедуру двухфакторного анализа, построить таблицу дисперсий, графики средних и таблицы множественных сравнений о разнородности групп.
3. Выполнить процедуры рангового анализа по Краскалу и Фридману. Вычислить ранги подгрупп.

С. Дать теоретическое обоснование полученных результатов.

Содержание отчета:

- Экран ввода и таблица дисперсионного анализа.
- Графики средних $\bar{Y}$ по А и $\bar{Y}$ по В.
- Экран ввода и таблица дисперсий $\bar{Y}$ по А,В.
- Множественный график средних.
- Таблицы множественных сравнений $\bar{Y}$ по А, $\bar{Y}$ по В, $\bar{Y}$ по А,В.

- Таблица ранговых критериев подгрупп.

Таблица вариантов:

Вар	a11	a12	a13
	a21	a22	a23
	a31	a32	a33

Лабораторная работа 11. Модель оптимизации и стохастической аппроксимации.

В качестве алгоритмов управления используются алгоритмы стохастической аппроксимации.

Пусть X векторная переменная в R^N , для которой выполняются условия:

1. Каждому значению X соответствует случайная величина Y с математическим ожиданием $MY(X)$.
2. Будем полагать, что $MY(X)$ имеет единственный максимум в точке X^* (в основном, без ограничения общности будем полагать $X^*=0$).
3. $MY(X)$ - является непрерывной функцией с непрерывными первыми и вторыми производными. Обозначим их $B(X)$ и $A(X)$ соответственно.
4. Вторые частные производные $\partial^2 MY / \partial x_i \partial x_j$ ограничены на всей области.
5. Пусть $\{a_k\}$ и $\{c_k\}$ две последовательности положительных чисел, удовлетворяющие условиям:

$$\text{а) } \lim_{k \rightarrow \infty} c_k = 0, \text{ б) } \sum_1^\infty a_k = \infty, \text{ в) } \sum_1^\infty a_k \cdot c_k < \infty, \text{ г) } \sum_1^\infty \left(\frac{a_k}{c_k} \right)^2 < \infty \quad (1)$$

6. Строится рекуррентная последовательность случайных векторов X :

$$\text{Переход по градиенту: } X_{k+1} = X_k + a_k \frac{\Delta Y_k(c_k)}{c_k}.$$

$$\text{Переход по знаку приращения: } X_{k+1} = X_k + a_k \cdot \text{sign}(\Delta Y_k). \quad (2)$$

7. Вектор ΔY_k определяется на основании реализации случайных величин в локальной окрестности точки X_k в соответствии с одним из планов Π_1 , Π_2 или Π_3 :

Π_1 - центральный план:

$$\Pi_1 = [X_k, X_k + c_k E_1, \dots, X_k + c_k E_i, \dots, X_k + c_k E_N]^T; \quad (3)$$

Π_2 - симметричный план:

$$\Pi_2 = [X_k + c_k E_1, X_k - c_k E_1, \dots, X_k + c_k E_N, X_k - c_k E_N]^T. \quad (4)$$

Π_3 - план с центральной точкой:

$$\Pi_3 = [X_k, X_k + c_k E_1, X_k - c_k E_1, \dots, X_k + c_k E_N, X_k - c_k E_N]^T. \quad (5)$$

$$\text{где } E_i = \begin{bmatrix} 0 & 0 & \dots & 1 & \dots & 0 \\ 1 & 2 & & i & & N \end{bmatrix}^T.$$

8. Дисперсия оценки σ_k^2 в каждой точке плана X_k :

$$\sigma_k^2 \leq \sigma^2 < \infty. \quad (6)$$

При выполнении перечисленных условий на основании теоремы последовательность X_k с вероятностью 1 сходится к нулю.

Класс последовательностей $\{a_k\}$ и $\{c_k\}$ может быть весьма разнообразный, но в практике имитации достаточно рассмотреть степенные зависимости:

$$a_k = \frac{a_0}{k^{ap}}; c_k = \frac{c_0}{k^{cp}}. \quad (7)$$

Для определенных последовательностей a_k и c_k условия 5 дают следующие ограничения на показатели степеней:

$$а) ap > 0, б) ap < 1, в) cp + ap > 1, г) ap - cp > 0.5. \quad (8)$$

После преобразований получим:

$$1 \geq ap > 0.75 \text{ и } ap - 0.5 > cp > 1 - ap. \quad (9)$$

Для определения координат вектора ΔY_k необходимо произвести L замеров значения Y . Величина L равна:

$N+1$ – для плана П1,
 $2N$ – для плана П2,
 $2N+1$ – для плана П3.

$$\xi_{x_k}(t)$$

Значение параметра в окрестности точки X_k , соответствующее l -му замеру, будем обозначать как $X_{k,l}$. Таким образом, для определения координат вектора ΔY_k необходимо построить семейство процессов $(\xi_{X_{k,l}}(t | s_k))_{l=1}^L$

В результате проведенной выше формализации алгоритм функционирования управляемой имитационной модели представляет собой следующую последовательность действий:

1. Начальная настройка модели и выбор начального значения управляемого параметра $X_0, k=0$.
2. При заданном значении параметра X_k в его локальной окрестности в соответствие с одним из планов Π_i ($i=1,2,3$) генерируются выборочные траектории $(\xi_{X_{k,l}}(t/s_k))_{l=1}^L$ длительностью T каждая из общего начального состояния s_k .
3. Вычисляются среднеинтегральные оценки основного процесса для всех $l=1...L$ при общем начальном состоянии s_k :

$$\zeta_{X_{k,l}}(T) = \frac{1}{T} \int_0^T \xi_{X_{k,l}}(t | s_k) dt \quad (10)$$

4. Вычисляются оценки координат вектора β_k - градиента функции $Y(X_k)$.

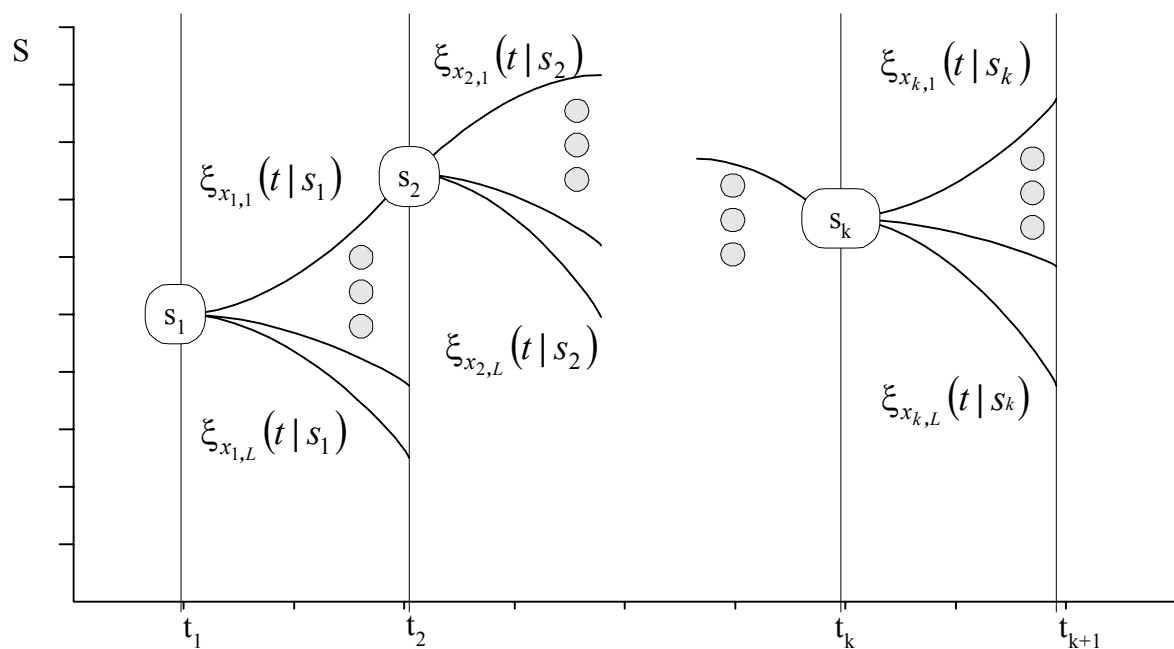
5. Выбирается очередное значение управляемого параметра:

$$X_{k+1} = X_k + a_k \beta_k \quad (11)$$

6. Задается начальное состояние s_{k+1} следующего интервала управления, равное конечному состоянию одного из процессов предыдущего шага.

7. В соответствии с выбранным критерием останова выполняется переход к пункту 2, либо к окончанию моделирования.

Управляемый имитационный процесс



Таким образом, для решения поставленной задачи оптимизации необходимо знать поведение:

- основных процессов $\xi(t/s_j)$
 - характеристик $\zeta(T/s_j)$ основного процесса
 - процесса управления $\eta(t)$
- управляемого имитационного процесса.

Лабораторная работа 12. Модель стохастического сетевого планирования.

В задачах сетевого планирования рассматривается оргграф $G(V, A)$, где V – множество вершин, а A – множество дуг, в котором выделены две вершины: s (источник) и t (сток). Из источника должны только выходить дуги, а в сток – только входить. Оргграф должен быть ациклическим и со взвешенными дугами. Вершины называются событиями, дуги – работами; вес дуги – это время выполнения данной работы. Источник s – это событие-начало работы над проектом; сток t – это событие-окончание работы. Функция `grPERT` решает стандартную задачу сетевого планирования. Синтаксис вызова:

`[CrP, Ts, Td] = grPERT(E)` – для ациклического оргграфа E находит критический путь `CrP`, время наступления каждого события `Ts` и запас времени для каждой работы `Td`.

Входной параметр `E(m, 2)` или `E(m, 3)` – список дуг оргграфа и их веса. Первый и второй элементы каждой строки – это номера вершин дуги (её начало и конец); третий (если он есть) – вес дуги (ограничение на пропускную способность); `m` – количество дуг. Если веса не заданы, они считаются единичными.

Выходной параметр `CrP` – критический путь: вектор-строка с номерами вершин, входящих в него.

Выходной параметр `Ts(1, n)` – время начала каждого события.

Выходной параметр `Td(m, 1)` – запас времени для каждой работы.

Пример обращения (скопируйте этот фрагмент в командное окно или редактор MATLAB):

```
clear all % очистили память
V=[1 1;0 0;1 0;1 -1;2 1;2 0;2 -1;3 1;...
  4 0;3 -1;3 0]; % координаты вершин-событий
E=[2 1 5;2 3 5;2 4 5;1 3 2;3 4 2;1 5 3;...
  1 6 2;3 6 5;3 7 2;4 7 3;5 6 1;6 7 1;...
  5 8 5;6 8 2;6 11 3;7 11 2;7 10 3;8 11 2;...
  11 10 2;8 9 5;11 9 4;10 9 4]; % дуги-работы и их веса-времена
grPlot(V,E,'d','%d','%d');
```

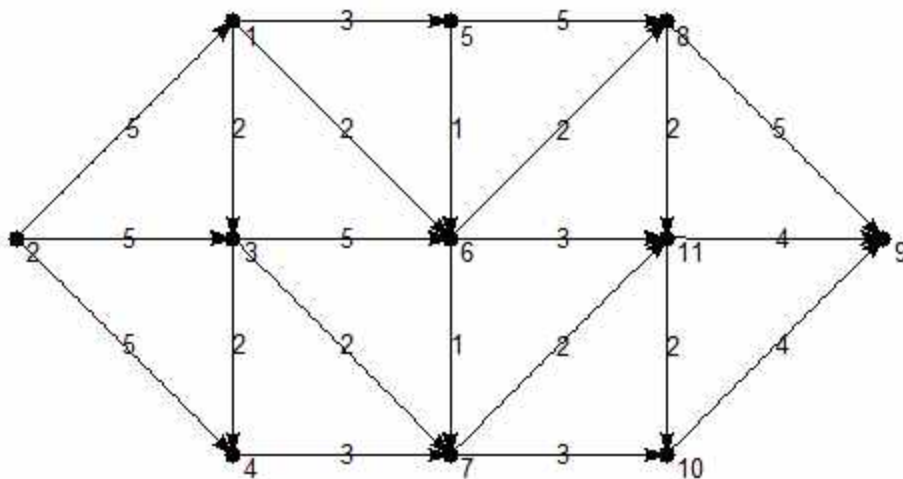
```

set(get(gcf,'CurrentAxes'),...
    'FontName','Times New Roman Cyr','FontSize',10) % установили шрифт
title('\bfСхема проекта')
[CrP,Ts,Td]=grPERT(E); % решаем задачу
grPlot([V Ts'],[CrP(1:end-1);CrP(2:end)],'d','%d',"");
set(get(gcf,'CurrentAxes'),...
    'FontName','Times New Roman Cyr','FontSize',10) % установили шрифт
title('\bfКритический путь и момент каждого события')
grPlot([V Ts'],[E(:,1:2) Td],'d','%d','%d');
set(get(gcf,'CurrentAxes'),...
    'FontName','Times New Roman Cyr','FontSize',10) % установили шрифт

    title('\bfМоменты времени событий и возможные
задержки времени работ')

```

Схема проекта



Лабораторная работа 13. Модель управления запасами и планирование эксперимента.

Построение и анализ дробных факторных планов $2^{k-(n-k)}$

На практике достаточно часто ограничиваются рассмотрением двух уровней факторов, что дает возможность оценить вклад каждого на изменение исследуемого функционала [25, 121]. Для решения этих задач в основном используются дробные факторные планы, которые направлены на оценку главных эффектов и не рассматривают эффекты взаимодействия высших порядков. Далее в работе проведен анализ и получены оценки влияния следующих факторов:

s - цена хранения одной позиции за единицу времени (30 - 70);

g - стоимость поставки партии $g=(700 - 1300)$;

h - штраф за отсутствие деталей (200 - 400);

p - параметр интенсивности запросов на определенную деталь (0.2 - 0.6).

В соответствии с дробным планом, сформированным в пакете Statistica 10 была рассмотрена задача анализа влияния указанных переменных на издержки системы управления запасами. Значения +1, -1 соответствуют нижнему и верхнему уровню фактора, а столбец F издержкам, рассчитанным в программном модуле.

Таблица факторного плана

№	s	g	h	p	F
1	-1	-1	-1	-1	142 120
2	1	-1	-1	1	1 575 620
3	-1	1	-1	1	1 577 780
4	1	1	-1	-1	225 350
5	-1	-1	1	1	4 623 990
6	1	-1	1	-1	340 460
7	-1	1	1	-1	333 730
8	1	1	1	1	4 584 510

В результате анализ этого плана получены количественные оценки влияния факторов. Уже из таблицы видно, что наиболее значимым является интенсивность потребности деталей.

Оценка эффектов влияния факторов на функционал

Фактор	Оценки эффектов; R-кв.=,83602;Скор.,61738 (План2n) 2**(4-1) план; Остаточн.SS=137809E7 ЗП F									
	Эффект	Ст.Ош.	t(3)	p	-95, % Дов.Пред	+95, % Дов.Пред	Коэф.	Ст.Ош. Коэф.	-95, % Дов.Пред	+95, % Дов.Пред
Сред/Св.член	1675445	415043,6	4,036793	0,027344	354591	2996299	1675445	415043,6	354591	2996299
(1)s	12080	830087,2	0,014553	0,989303	-2629628	2653788	6040	415043,6	-1314814	1326894
(2)mu	2830060	830087,2	3,409353	0,042171	188352	5471768	1415030	415043,6	94176	2735884
(3)g	9795	830087,2	0,011800	0,991326	-2631913	2651503	4897	415043,6	-1315956	1325751
(4)h	1590455	830087,2	1,916010	0,151230	-1051253	4232163	795227	415043,6	-525626	2116081

Однако в программной реализации факторного планирования пакета Statistica 10 имеется множество форм представления результатов [11, 111]. Так, диаграмма Парето представляет достаточно информативную форму визуализации результатов эксперимента, на которой в виде столбцов отображаются оценки эффектов факторов. Значения проранжированы и отображаются по мере убывания

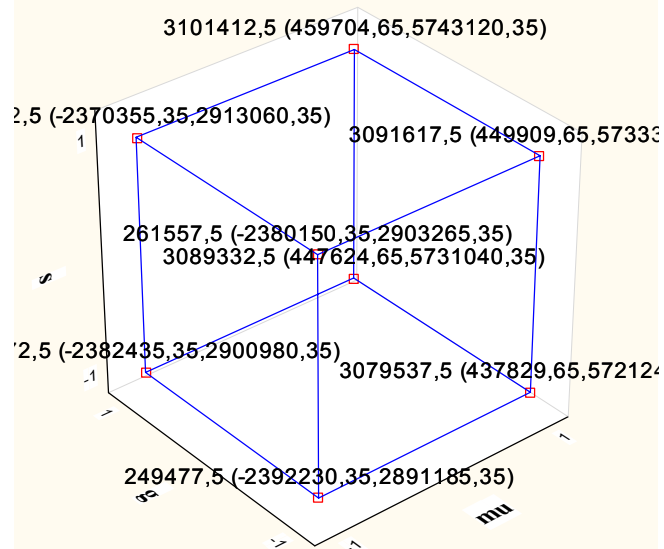
Диаграмма Парето оценки эффектов факторов



Кубические и квадратичные диаграммы используются для конечного представления прогнозируемых значений зависимой переменной на верхних и нижних уровнях факторов [11]. Квадратичная диаграмма отображает прогнозируемые значения для двух исходных факторов, кубическая диаграмма - для трех.

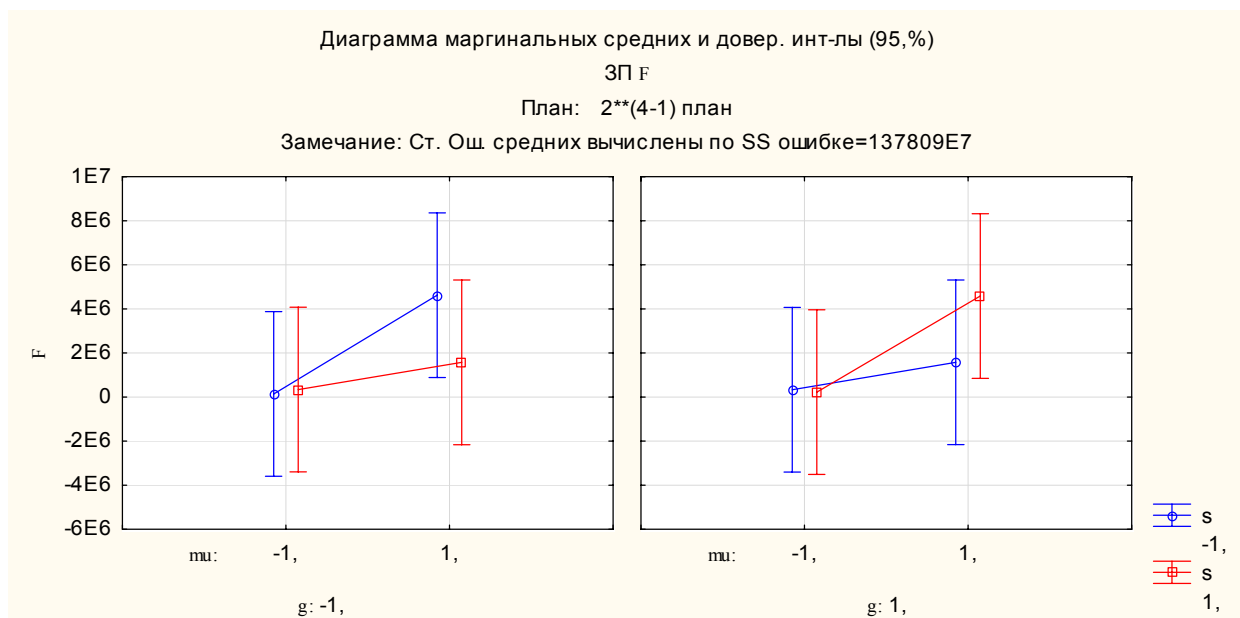
Кубическая диаграмма представления прогнозов

Предсказанные сред. значения: F
 $2^{**}(4-1)$ план; Остаточн. SS=137809E7
 Модель включает: Гл. эффекты
 (95,% доверительные интервалы приводятся в скобках)



Диаграммы взаимодействий служат для визуализации средних значений [11]. Она представляет классическую диаграмму взаимодействий с указанием средних точек, между которыми построена линия. Диаграмма полезна при наличии в модели эффектов взаимодействия.

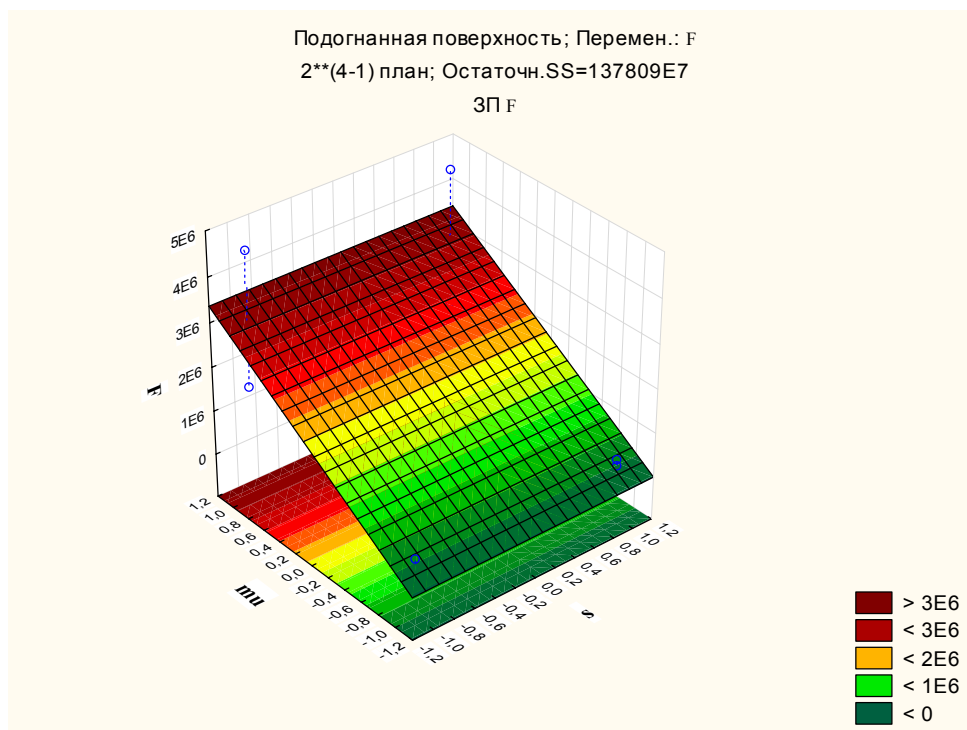
Диаграммы средних значений затрат от четырех параметров



Графики остатков [11] дает визуальную оценку точности аппроксимации реальных данных полученной линейной моделью. Распределение остатков может зависеть от абсолютных значений. Такая модель представления используется также для формирования вероятностных графиков и гистограмм остатков. Диаграммы поверхности и контурные диаграммы используются для непрерывных факторов плана и применяются, когда полезно посмотреть на график поверхности исследуемой функции от значений факторов.

Данные диаграммы могут быть весьма полезными для общего анализа исследуемой зависимости. В целом планы 2^{k-p} больше всего применяются на практике. При этом достаточно эффективно может быть получена оценка вклада каждого фактора на основании небольшого количества экспериментов. Однако, логика проведения испытаний очень проста - всего по два уровня для каждого фактора.

Диаграммы поверхности аппроксимации



Построение и анализ планов Бокса-Бенкена $3^{**}(k-p)$

Для нелинейных зависимостей влияния факторов необходимо по крайней мере три уровня каждого фактора, что позволит дать оценку квадратичных эффектов [118]. В целом, нелинейность в практических задачах нелинейность проявляется, когда параметры системы лежат достаточно близко к оптимальному значению. Кроме того, даже некоторые категориальные фактор могут иметь более двух уровней, например, варианты стратегий управления запасами. Общие принцип построения дробных трехуровневых факторных планов $3^{**}(k-p)$ аналогичен принципам построения $2^{**}(k-p)$ планов. Это план формируется на основании полного факторного плана 3^2 (факторы A и B). Третий фактор C рассчитывается на основании соотношения $C=1-\text{mod}_3(A+B)$. Символ $\text{mod}_3(x)$ представляет остаток от деления на 3. Фундаментальное тождество определяется как $0=\text{mod}_3(A+B+C)$. Так, поскольку влияние максимального и минимального объема запасов влияет явно нелинейно, то для этих факторов, а также случайного времени исполнения заказа построим факторный план Бокса-Бенкена $3^{**}(3-1)$:

Z_{max} – максимальный объем запаса на складе (10 - 30);

Z_{min} – минимальный объем запаса на складе (4 – 8);

T_{zak} – время доставки партии (3 – 5);

Значения -1, 1 и 0 определяют нижний уровень, верхний и среднюю точку соответственно. Столбец F содержит рассчитанные значения издержек. В результате анализ плана формируются оценки эффектов. Оценки главных эффектов помечаются буквой L вместе с наименованием фактора и показывают разницу между средними значениями на нижнем и верхнем уровне фактора. Оценка нелинейного эффекта (Q) определяет разность между усредненным значением в центральной точке и линейной комбинацией значений на нижнем и верхнем уровне фактора.

Трехуровневый план Бокса-Бенкена

№	Zmin	Zmax	Tzak	F
1	-1	-1	-1	4 568 950
2	-1	0	1	667 200
3	-1	1	0	657 550
4	0	-1	1	295 350
5	0	0	0	402 500
6	0	1	-1	218 000
7	1	-1	0	193 050
8	1	0	-1	399 900
9	1	1	1	353 250

Оценка эффектов

Фактор	Оценки эффектов; R-кв.=,76002;Скор.,04007 (План3п) 3 3-х уровневые ф, 1 Блоки; Остаточн.SS=188242E7 ЗП F										
	Эффект	Ст.Ош.	t(2)	p	-95,% Дов.Пред	+95,% Дов.Пред	Коэф.	Ст.Ош.	-95,% Дов.Пред	+95,% Дов.Пред	
								Коэф.			
Сред/Св.член	861750	457339	1,88427	0,200206	-1106020	2829520	861750	457338,7	-1106020	2829520	
(1)Zmin(L)	-1649167	1120246	-1,47215	0,278846	-6469198	3170865	-824583	560123,2	-3234599	1585432	
Zmin(K)	-834700	970162	-0,86037	0,480253	-5008970	3339570	-417350	485080,9	-2504485	1669785	
(2)Zmax(L)	-1276183	1120246	-1,13920	0,372680	-6096215	3543848	-638092	560123,2	-3048107	1771924	
Zmax(K)	-557825	970162	-0,57498	0,623366	-4732095	3616445	-278913	485080,9	-2366047	1808222	
(3)Tzak(L)	-1290350	1120246	-1,15184	0,368484	-6110381	3529681	-645175	560123,2	-3055191	1764841	
Tzak(K)	-666075	970162	-0,68656	0,563273	-4840345	3508195	-333038	485080,9	-2420172	1754097	

Диаграмма Парето интерпретируется также, как и для двухуровневого плана.

Диаграмма Парето трехуровневого плана



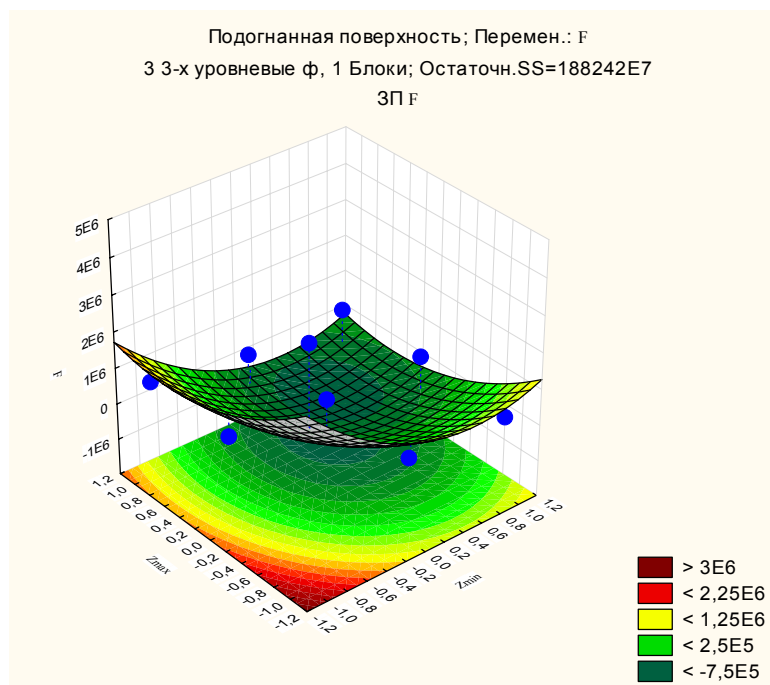
Из графика прогнозируемые и наблюдаемые видно, что явно выпадает одна точка и остальные также не достаточно близки к прямой. Это может говорить о неправильно подобранных границах варьирования факторов.

График прогнозируемых и наблюдаемых значений



Для исследуемой системы факторов полезен график поверхности, который некоторым образом отображает модель прогноза для исследуемого функционала.

График поверхности отклика функционала



Таким образом из графика поверхности видно, что существуют оптимальные значения $Z_{\max}$ и $Z_{\min}$, соответственно максимального и минимального значения запаса детали на складе.

Диаграммы взаимодействия средних также показывают на возможный выброс значения функционала, который может быть связан с достаточно большим ростом на соответствующем интервале. Поэтому наверно необходимо некоторым образом изменить границы вариации факторов.

Процедура анализа трехуровневых планов достаточно близка к процедуре анализа двухуровневых, но в данном случае для всех эффектов здесь имеется возможность анализа помимо линейности еще и квадратичной нелинейности.

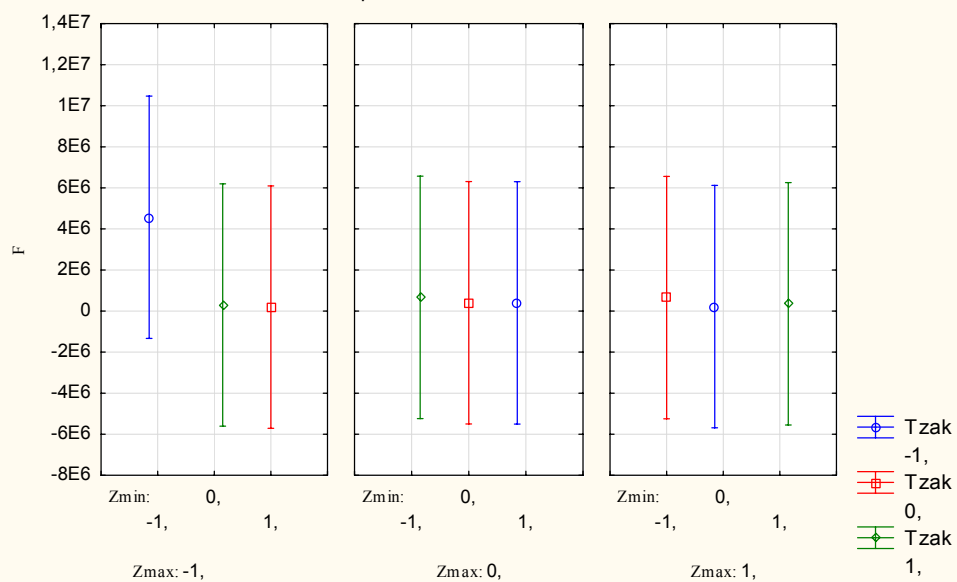
Диаграмм условных средних значений

Диаграмма маргинальных средних и довер. инт-лы (95,%)

ЗП F

План: 3 3-х уровневые ф, 1 Блоки , 9 опыт

Замечание: Ст. Ош. средних вычислены по SS ошибке=188242E7



Лабораторная работа 14. Операции над нечеткими переменными.

Построение функций принадлежности

Функция принадлежности определяет «смысл» термина. Эти функции были уже построены ранее из некоторых интуитивных соображений.

При введении некоторой функции необходимо определение ее зависимости от контекста задачи. Например «уровень знаний» = «хорошо» для одного преподавателя может существенно отличаться от мнения другого. Или в ситуации, когда идет речь просто о получении оценки или при выборе тестируемых для проведения ответственных работ. Житейский пример, «рост» (просто) и «рост» баскетболиста имеет совершенно различные функции. Все зависит от контекста и условий постановки задачи.

Обозначение для дискретных переменных

$$A = \{(a_i | \mu_i)\} \quad A = \{(a_1 | \mu_1), (a_2 | \mu_2), \dots, (a_n | \mu_n)\}$$

Например, «уровень знаний» в десятибалльной шкале

$$X - \text{число} \quad U = \{1, 2, \dots, 10\}$$

$$\langle \text{плохой} \rangle = \{(1 | 0.4), (2 | 0.5), (3 | 1), (4 | 0.4)\}$$

$$\langle \text{небольшой} \rangle = \{(3 | 0.5), (4 | 0.5), (5 | 1), (6 | 1), (7 | 0.8), (8 | 0.5)\}$$

$$\langle \text{очень хороший} \rangle = \{(6 | 0.4), (7 | 0.6), (8 | 0.8), (9 | 0.9), (10 | 1)\}$$

Следует различать значения термов <неплохой> и <не плохой>. Первый является базовым термом, в то время, как второй - булевым преобразованием базового термина «плохой»

Для определения непрерывных функций принадлежности используются аналитические выражения (как и для распределения вероятностей).

Например, семантическое правило задания «сложности» теста

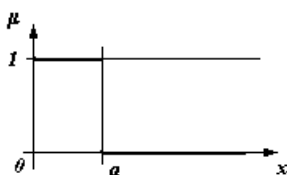
$$\mu(S=\text{легкий}) = \exp(-kx^2)$$

$$\mu(S=\text{сложный}) = \exp(-k(x-15)^2)$$

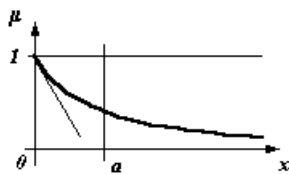
с областью определения [0,15].

В таблице приведены основные классы функций принадлежности используемые на практике

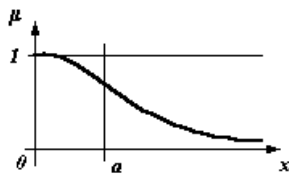
<величина x мала>



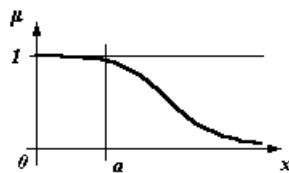
$$\mu(x) = \begin{cases} 1 & 0 \leq x \leq a \\ 0 & a < x \end{cases}$$



$$\mu(x) = \exp(-kx) \quad k > 0$$

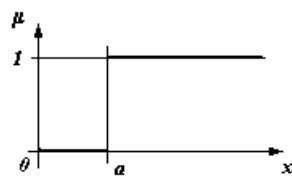


$$\mu(x) = \exp(-kx^2) \quad k > 0$$

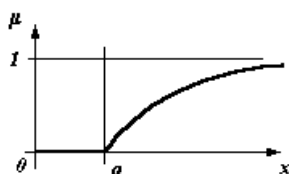


$$\mu(x) = \frac{1}{1 + k \cdot x^2} \quad k > 0$$

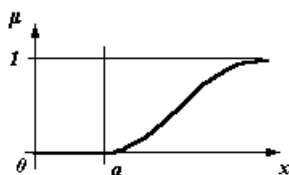
<величина x большая>



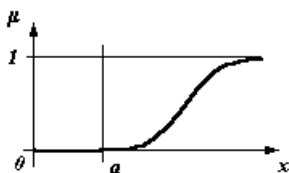
$$\mu(x) = \begin{cases} 0 & 0 \leq x \leq a \\ 1 & a < x \end{cases}$$



$$\mu(x) = \begin{cases} 0 & 0 \leq x \leq a \\ 1 - \exp(-k(x-a)) & a < x \end{cases} \quad k > 0$$



$$\mu(x) = \begin{cases} 0 & 0 \leq x \leq a \\ 1 - \exp(-k(x-a)^2) & a < x \end{cases} \quad k > 0$$



$$\mu(x) = \begin{cases} 0 & 0 \leq x \leq a \\ \frac{k(x-a)^2}{1 + k(x-a)^2} & a < x \end{cases}$$

Этот класс функций принадлежности может быть расширен по усмотрению конкретного эксперта при решении конкретной задачи

Операции над нечеткими множествами

При манипуляции с нечеткими понятиями возникает необходимость выполнения некоторых операций, таких как сложение (например, времени выполнения) или умножение (например, сложности задач на их количество). Для реализации этих операций введены элементарные операции над нечеткими переменными

Принадлежность множеств может быть собственной и несобственной

$$A \subset B \Leftrightarrow \forall x \in E \mu_A(x) \leq \mu_B(x)$$

$$A \subset\subset B \Leftrightarrow A \subset B \exists x \in E \mu_A(x) < \mu_B(x)$$

Равенство переменных рассматривается как двусторонняя принадлежность

$$A = B \Leftrightarrow A \subset B \wedge B \subset A$$

Дополнение переменной приводит к дополнению до 1 функции принадлежности

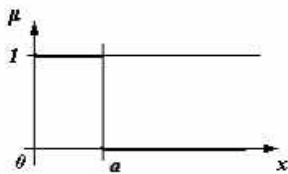
$$\overline{A} \Leftrightarrow \forall x \mu_{\overline{A}}(x) = 1 - \mu_A(x) \quad \overline{\overline{A}} = A$$

Пересечение и дополнение переменных определяется через минимум и максимум функций принадлежности соответственно

$$A \cap B \Leftrightarrow \forall x \in E \mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x))$$

$$A \cup B \Leftrightarrow \forall x \in E \mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x))$$

Принадлежность



Объединение



Пересечение



Разность определяется через пересечение функций

$$A - B = A \cap \bar{B} \quad A - B \neq B - A$$

Кроме того вводятся некоторые дополнительные операции, согласованные с обычными теоретико-множественными, такие как

$$\text{дизъюнктивная сумма } A \oplus B = (A \cap \bar{B}) \cup (\bar{A} \cap B)$$

$$\text{Алгебраическое произведение} \quad A \bullet B \Leftrightarrow \forall x \in E \quad \mu_{A \bullet B}(x) = \mu_A(x) \cdot \mu_B(x)$$

$$\text{Алгебраическая сумма} \quad A \oplus B \Leftrightarrow \forall x \in E \quad \mu_{A \oplus B}(x) = \mu_A(x) + \mu_B(x) - \mu_A(x) \cdot \mu_B(x)$$

В результате определив класс функций принадлежности введенные операции позволяют вычислять смысл всех элементов терм-множества, образующего исследуемый вид. Приведенные выше операции не представляют технических сложностей.

Свойства дистрибутивности $R1 \cap R2$

R1	$\cap$	(R2	$\cup$	R3)	$= (R1$	$\cap$	R2)	$\cup$	(R1	$\cap$	R3
	$\cup$		$\cap$			$\cup$		$\cap$			
	$\bullet$		$\cup$			$\bullet$		$\cup$			
	$\bullet$		$\cap$			$\bullet$		$\cap$			

Для некоторых операций свойства нечетких множеств не соответствуют обычным, так $A \cap \bar{A} \neq 0$ и $A \cup \bar{A} \neq E$.

Рассматривая нечеткие множества, а не переменные, расстояние или мера различия между ними определяется как

$$d(A, B) = \sum_{i=1}^N (\mu_A(x_i) - \mu_B(x_i)) \quad \delta(A, B) = \frac{d(A, B)}{N},$$

что является частным случаем евклидова расстояния

$$e(A, B) = \sqrt{\sum_{i=1}^N (\mu_A(x_i) - \mu_B(x_i))^2} \quad 0 \leq e(A, B) \leq \sqrt{N}$$

Принцип обобщения

Для определения более сложных операций вводится понятие принцип обобщения операций. Для любой бинарной операции определенной над обычными переменными возможно обобщение этой операции и на нечеткие переменные.

Пусть '*' - произвольная бинарная операция над двумя переменными, тогда функция принадлежности результирующей будет равна

$$A * B = \sum_{i,j} (\mu_i \wedge \nu_j) \cdot (u_i * v_j) = \bigvee_{i,j} (\min(\mu_A(x_i), \mu_B(x_j)); (x_i, x_j))$$

Это правило естественным образом может быть расширено и на непрерывные переменные.

Это правило можно использовать, например при оценки уровня знаний, когда определено, что тестируемый имеет «немного» знаний, и после следующего теста считают, что он повысил свой уровень еще на «немного». Понятие «немного» является значением лингвистической переменной «уровня знаний». Пусть его функция принадлежности равна

$\langle \text{немного} \rangle = \{(1|0.4), (2|0.5), (3|1), (4|0.4)\}$

Тогда для суммы, реализуя принцип обобщения получим

$\langle \text{немного} \rangle + \langle \text{немного} \rangle$

	1 0.4	2 0.5	3 1	4 0.4
1 0.4	2 0.4	3 0.4	4 0.4	5 0.4
2 0.5	3 0.4	4 0.5	5 0.5	6 0.4
3 1	4 0.4	5 0.5	6 1	7 0.4
4 0.5	5 0.4	6 0.5	7 0.5	8 0.4

Таким образом, результат сложения дает значение с функцией принадлежности.

$\langle \text{немного} \rangle + \langle \text{немного} \rangle = \{(2|0.4), (3|0.4), (4|0.5), (5|0.5), (6|1), (7|0.5), (8|0.4)\}$

Аналогично выполняются любые арифметические операции.

Лабораторная работа 15. Операции над нечеткими отношениями. Нечеткое сетевое планирование.

Лингвистические неопределенности.

Как отмечалось выше, часто используются термины

<более менее>, <очень> и другие.

При построении формальных процедур каждый из термов должен определять некоторое правило преобразования функции принадлежности того терма, к которому он применяется.

Одним из правил раскрытия неопределенности <более менее> является простое извлечение квадратного корня

$$\langle \text{более менее } X \rangle = \text{DIL} \langle X \rangle \quad \mu_{\langle \text{более менее } X \rangle} = \mu_{\langle X \rangle}^{0.5}$$

Другое раскрытие неопределенности <более менее> заключается в определении ядра преобразования и распространения его на всю область определения переменной. Пусть при моделировании временного фактора, используются термы «давно», «недавно» и др. Время исчисление пусть определяется годами.

$$\langle \text{недавно} \rangle = \{(95 | 0.7), (96 | 0.8), (97 | 1)\}$$

$$\langle \text{более менее недавно} \rangle = F(\langle \text{более менее} \rangle; K)$$

$$F - \text{оператор } F(A, K) = \int_U \mu_A(u) K(u) du$$

K - ядро преобразования (например)

$$K\{97\} = \{(97 | 1), (96 | 0.9)\}$$

$$K\{96\} = \{(96 | 1), (95 | 0.9)\}$$

$$K\{95\} = \{(95 | 1), (94 | 0.9)\}$$

$$\langle \text{более менее недавно} \rangle =$$

$$= 1 \cdot \{(97 | 1), (96 | 0.9)\} \cup 0.8 \cdot \{(96 | 1), (95 | 0.9)\} \cup 0.7 \cdot \{(95 | 1), (94 | 0.9)\} =$$

$$= \{(97 | 1) \cup (96 | 0.9) \cup (96 | 0.8) \cup (95 | 0.72) \cup (95 | 0.7) \cup (94 | 0.49)\} =$$

$$= \{(97 | 1) \cup (96 | 0.9) \cup (95 | 0.72) \cup (94 | 0.49)\}$$

Это отличается от

$$\langle \text{более менее недавно} \rangle = \{(95 | 0.7), (96 | 0.8), (97 | 1)\}^{0.5} =$$

$$= \{(97 | 1), (96 | 0.9), (95 | 0.84)\}$$

Для формализации преобразования неопределенности «очень» можно использовать простое возведение в квадрат.

Структурированная и булева переменная.

Структурированной лингвистической переменной называется переменная, у которой $T(L)$ можно задать алгоритмически.

Булевой лингвистической переменной называется переменная, у которой термы L являются булевыми выражениями.

T_p - первичные термы (<хорошо> | <плохо> | ...)

T_b - булевы термы (<и> | <или> | <не> | ...)

T_l - лингвистические неопределенности (<очень> | <более менее> | ...)

Построим функцию принадлежности для терма

<не очень молодой и не очень старый>

Последовательность построения результирующей функции принадлежности включает шаги

$M_{\langle \text{очень плохой} \rangle} = \text{CON } M_{\langle \text{плохой} \rangle}$

$M_{\langle \text{не очень плохой} \rangle} = \neg M_{\langle \text{очень плохой} \rangle}$

$M_{\langle \text{очень хороший} \rangle} = \text{CON } M_{\langle \text{хороший} \rangle}$

$M_{\langle \text{не очень хороший} \rangle} = \neg M_{\langle \text{очень хороший} \rangle}$

$M_{\langle \text{не очень молодой и не очень старый} \rangle} =$

$= M_{\langle \text{не очень хороший} \rangle} \wedge M_{\langle \text{очень хороший} \rangle}$

При более сложных преобразованиях и определениях булевых операций требуется построение грамматического дерева разбора.

Основные понятия и определения нечетких отношений

Определение нечеткого отношения строится на понятиях теории графов и лингвистических переменных. Пусть x и y две нечеткие переменные

$$x \in E_1 \quad y \in E_2$$

Подмножество $G: \forall (x, y) \in E_1 \times E_2 : \mu_G(x, y) \in M$ является нечетким графом,

где M - функция принадлежности, определенная на декартовом произведении.

$$G \{ (x_i, y_i) \mid \mu_{ij} \}_{ij=1..N}$$

Использование нечеткого отношения всегда проявляется в реальных ситуациях, когда выполняется некоторый сравнительный анализ. Например, сравнивается «уровень знаний» двух тестируемых. В этом случае термины «выше», «ниже», «много выше» являются термами этого нечеткого отношения.

Пусть $E_1 = E_2 = R \quad R = (-\infty, \infty) \quad y < x \quad \text{в } R^2$

$$\mu_{R^2} = \begin{cases} 0 & y \geq x \\ 1 & y < x \\ 1 + \frac{1}{(x-y)^2} & \end{cases}$$

В этом случае определена функция принадлежности отношения «много выше» - <<.

Для нечетких отношений, как и для нечетких переменных определены подобные теоретико-множественные операции. операции

$$\begin{aligned} \bigvee_x &\Leftrightarrow \max_x & \bigwedge_x &\Leftrightarrow \min_x \\ \mu_1(x) = \bigvee_y \mu(x, y) &\Leftrightarrow \mu_1(x) = \max_y \mu(x, y) \\ \mu_2(x) = \bigwedge_y \mu(x, y) &\Leftrightarrow \mu_2(x) = \min_y \mu(x, y) \end{aligned}$$

Проекция нечеткого отношения $\mu_R^1(x) = \bigvee_y \mu_R(x, y)$

Глобальная проекция $h(R) = \bigvee_x \bigvee_y \mu_R(x, y) = \bigvee_y \bigvee_x \mu_R(x, y)$

Принадлежность

$$R_1 \text{ содержит } R_2 \Leftrightarrow \forall (x, y) \in E_1 \times E_2: \mu_{R_1}(x, y) \leq \mu_{R_2}(x, y)$$

Пусть имеются два мнения преподавателей о понимании термина «много выше». Каждый определен своей функцией принадлежности. Тогда

$$\begin{aligned} \mu_{R_1}(x, y) &= \begin{cases} 0 & y - x < 0 \\ 1 - \exp(-k_1(y - x)^2) & y - x \geq 0 \end{cases} \\ \mu_{R_2}(x, y) &= \begin{cases} 0 & y - x < 0 \\ 1 - \exp(-k_2(y - x)^2) & y - x \geq 0 \end{cases} \end{aligned}$$

Одно отношение будет принадлежать другому, если

$$k_1 > k_2 \Rightarrow R_2 \text{ содержит } R_1$$

Аналогично можно ввести отношения <очень близкие> или <очень различные>.

Алгебраическое произведение

$$\mu_{R_1 \cdot R_2}(x, y) = \mu_{R_1}(x, y) \cdot \mu_{R_2}(x, y)$$

Дополнение

$$\forall (x, y) \in E_1 \times E_2: \mu_R(x, y) = 1 - \mu_R(x, y)$$

Дизъюнктивная сумма

$$R_1 \oplus R_2 = (R_1 \text{ I } \bar{R}_2) \cup (\bar{R}_1 \text{ I } R_2)$$

Одной из основных операций нечетких отношений, которая его принципиально отличает от нечеткой переменной является композиция (аналогия матричного произведения, и произведение линейных операторов).

$$(\max\cdot)-\text{композиция} \quad \mu_{R_1 \bullet R_2}(x,z) = \bigcup_y [\mu_{R_1}(x,y) \cdot \mu_{R_2}(y,z)]$$

$$(\max\cdot)-\text{композиция} \quad \mu_{R_1 \bullet R_2}(x,z) = \bigcup_y [\mu_{R_1}(x,y) \cdot \mu_{R_2}(y,z)]$$

Подмножество α -уровня

$$G_\alpha = \{(x,y) \mid \mu_R(x,y) \geq \alpha\}$$

$$\text{Свойства } \alpha_1 \geq \alpha_2 \Rightarrow G_{\alpha_1} \subset G_{\alpha_2}$$

$$\text{Теорема о декомпозиции} \quad \underline{R} = \bigcup_{\alpha} \alpha \cdot \underline{R}_\alpha \quad 0 < \alpha \leq 1$$

Свойство max-min-композиции

$$(\underline{R}_3 \bullet \underline{R}_2) \bullet \underline{R}_1 = \underline{R}_3 \bullet (\underline{R}_2 \bullet \underline{R}_1) - \text{ассоциативность}$$

Свойства нечетких отношений

$$1. \text{ Симметрия } \forall (x,y) \in E \times E \quad (\mu_R(x,y) = \mu) \Rightarrow (\mu_R(y,x) = \mu)$$

$$2. \text{ Рефлексивность } \forall (x,x) \in E \times E \quad \mu_R(x,x) = 1$$

$$3. \text{ Транзитивность } \forall (x,y), (x,z), (y,z) \in E \times E$$

$$\mu_R(x,z) \geq \max_y [\min (\mu_R(x,y), \mu_R(y,z))]$$

Свойства выполняются для всех видов композиций

Транзитивное замыкание отношения сводится к алгоритму

$$\underline{R}^2 = \underline{R} \bullet \underline{R} \quad \mu_R(x,z) \geq \max_y [\min (\mu_R(x,y), \mu_R(y,z))]$$

$$\underline{R} \bullet \underline{R} \subset \underline{R}$$

$$\text{Tr} \underline{R} = \underline{R} \cup \underline{R}^2 \cup \underline{R}^3 \cup \dots$$

Алгоритм дискретной композиции

Для случая дискретных переменных правило композиции переходит в матричные преобразования.

Пусть A - вектор степеней принадлежности

F- нечеткое отношение

$$A = \begin{bmatrix} 0.5 \\ 1 \\ 0.7 \\ 0.1 \end{bmatrix}^T \cdot F = \begin{bmatrix} 1 & 0.5 & 0.2 & 0.1 \\ 0.7 & 1 & 0.9 & 0.3 \\ 0.4 & 0.6 & 1 & 0.7 \\ 0.3 & 0.5 & 0.7 & 1 \end{bmatrix}; \begin{bmatrix} 0.5 \\ 1 \\ 0.7 \\ 0.1 \end{bmatrix}^T \cdot \begin{bmatrix} 1 & 0.5 & 0.2 & 0.1 \\ 0.7 & 1 & 0.9 & 0.3 \\ 0.4 & 0.6 & 1 & 0.7 \\ 0.3 & 0.5 & 0.7 & 1 \end{bmatrix} = \begin{bmatrix} 0.7 \\ 1 \\ 0.7 \\ 0.7 \end{bmatrix}^T$$

Определение - композиционное правило вывода

$R(u)=A \quad R(u,v)=F$ имеет вид $R(v)=A \bullet F$

Далее определим размытые множества подблоков и компонент

Пример на принятие решений

A - «малый» = $\{(1|1), (2|0.6), (3|0.2)\}$

$$F - \text{«примерно равны»} = F = \begin{bmatrix} (1,1)|1 & (1,2)|0.5 & (1,3)|0 & (1,4)|0 \\ (2,1)|0.5 & (2,2)|1 & (2,3)|0.5 & (2,4)|0 \\ (3,1)|0 & (3,2)|0.5 & (3,3)|1 & (3,4)|0.5 \\ (4,1)|0 & (4,2)|0 & (4,3)|0.5 & (4,4)|1 \end{bmatrix}$$

«результат» = «малый» $\bullet$ «примерно равны»

$$\begin{bmatrix} 1 & 0.6 & 0.2 & 0 \end{bmatrix} \begin{bmatrix} (1,1)|1 & (1,2)|0.5 & (1,3)|0 & (1,4)|0 \\ (2,1)|0.5 & (2,2)|1 & (2,3)|0.5 & (2,4)|0 \\ (3,1)|0 & (3,2)|0.5 & (3,3)|1 & (3,4)|0.5 \\ (4,1)|0 & (4,2)|0 & (4,3)|0.5 & (4,4)|1 \end{bmatrix} = \begin{bmatrix} 1 & 0.6 & 0.5 & 0.2 \end{bmatrix}$$

Результат можно аппроксимировать «более менее малый»

где «более менее» оператор с ядром $K(i)=\{(i|1), (i+1|0.7)\}$

u - «малый»

u и v - «примерно равны»

v - «более менее малый»

Приближенный вывод

Правило MODUS PONENS как частный случай композиционного правила вывода

$$\begin{array}{ccc} \text{MODUS PONENS} & \underline{X \Rightarrow Y, X} & \text{MODUS TOLLENS} \quad \underline{X \Rightarrow Y, \neg Y} \\ & \Delta Y & \Delta \neg X \end{array}$$

Если A , то $B \quad A \Rightarrow B$ Размытые цели, ограничения и решения.

Алгоритм композиции в непрерывном случае

$$R_1 \subset X \times Y, R_2 \subset X \times Y$$

$$(\max\text{-min})\text{-композиция} \quad \mu_{R_1 \bullet R_2}(x,z) = \max_y [\min (\mu_{R_1}(x,y), \mu_{R_2}(y,z))]$$

Пример:

$$\mu_{R_1}(x,y) = e^{-k(x-y)^2}, k \geq 1 \quad \mu_{R_2}(x,y) = e^{-k(y-z)^2}, k \geq 1$$

В результате получаем:

$$\mu_{R1 \bullet R2}(a, b) = \max_y \left[\min(e^{-k(x-y)^2}, e^{-k(x-y)^2}) \right],$$

$$\mu_{R1 \bullet R2}(a, b) = e^{-k(a - \frac{a+b}{2})^2} = e^{-k \left[\frac{a-b}{2} \right]^2},$$

$$\mu_{R1 \bullet R2}(x, z) = e^{-k(a - \frac{a+b}{2})^2} = e^{-k \left[\frac{x-z}{2} \right]^2}.$$

Лабораторная работа 16. Процедуры многокритериальной оптимизации

Программа расчета по методу идеальной точки

Пусть $\Omega \subset \mathbb{R}^m$, $y, y' \in \Omega$

Положим $y_i^* = \max y_i$

В результате получим точку $y^* = (y_1^*, y_2^*, \dots, y_m^*)$ как решение обычной задачи одномерной оптимизации. В предположении замкнутости Ω , такая точка существует.

Однако идеальная точка $y^*(\Omega) = (y_1^*, y_2^*, \dots, y_m^*)$ как правило не принадлежит множеству Ω . При этом справедливо соотношение

$$\forall y \in \Omega \quad y^* P y$$

Функция выбора в методе идеальной точки задается уравнением:

$$C_x(y) = \arg \min \rho(y, y^*)$$

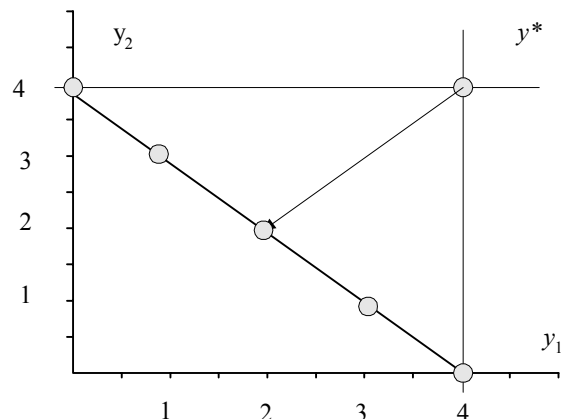
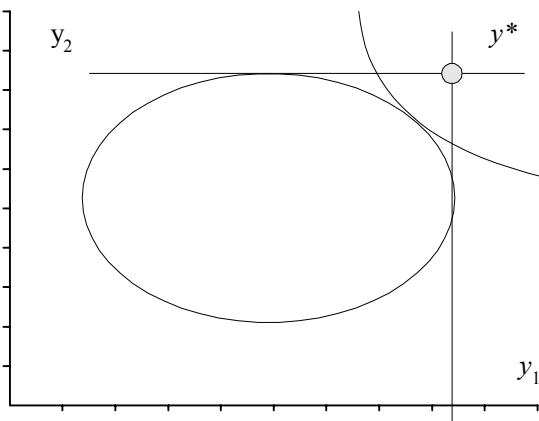
Где ρ - некоторая метрика

$$\rho(a, b) = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \quad \text{— Евклидова метрика}$$

$$\text{Для метрики } \rho(a, b) = \sum_{i=1}^n |a_i - b_i| \quad \text{функция обозначается } C_x^1,$$

$$\text{Для метрики } \rho(a, b) = \max_i |a_i - b_i| \quad \text{функция обозначается } C_x^\infty$$

На рис. представлена иллюстрация решения по методу идеальной точки



Если множество Ω - выпукло, то решение $C_X(Y)$ состоит из одной точки. Пусть имеются две точки y_1 и y_2 на равном расстоянии ρ_2 от идеальной, тогда в силу выпуклости Ω ($\lambda y_1 + (1-\lambda)y_2 \in \Omega$) и теоремы Пифагора средняя точка при $\lambda=0.5$, что и доказывает утверждение единственности.

Для этого метода также справедливо условие принадлежности решения множеству Парето

$$C_X(\Omega) \in P(\Omega)$$

Для множества решений в метриках ρ_1 и ρ_∞ условия единственности не выполняются

Программа расчета по методу выбора с учетом числа доминируемых критериев

$$\Omega \subset \mathbf{R}^m, y, y' \in \Omega$$

Определяется число критериев, по которому один вариант превосходит другой

$$\text{Card } I = \{i: y_i > y'_i\}$$

C^k – функция выбора определяется по правилу определения числа доминируемых критериев:

$q(y, y')$ – число критериев по которым $y > y'$, т.е. $y_i > y'_i$

$$x \in \Omega \quad Q_Y(y) = \max_{y' \in Y} q(y, y')$$

$Q_Y(y)$ – доминирующий показатель варианта y при предъявлении Y .

В результате функция выбора определяется соотношением:

$$C^k(Y) = \left\{ y \in Y \mid Q_Y(y) = \min_{y'' \in Y} Q_Y(y'') \right\}$$

Пусть множество предъявления $\Omega = \{y_1, y_2, y_3, y_4, y_5\}$

$y_1 = \{1, 1, 5\}$, $y_2 = \{3, 2, 4\}$, $y_3 = \{4, 3, 2\}$, $y_4 = \{7, 0, 1\}$, $y_5 = \{2, 8, 0\}$.

Матрица $A = \|a_{ij}\|$ $a_{ij} = q(y_i, y_j)$ – i -ая альтернатива уступает j -ой по a_{ij} критериям

$$A = \begin{vmatrix} 0 & 2 & 2 & 1 & 2 \\ 1 & 0 & 2 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 2 & 2 & 2 & 0 & 1 \\ 1 & 2 & 2 & 2 & 0 \end{vmatrix} \quad \begin{bmatrix} 2 \\ 2 \\ 1 \\ 2 \\ 2 \end{bmatrix} \quad i = \arg \min_{y_i \in Y} Q_\Omega(y_i)$$

В результате, в качестве решения выбираем y_3 .

Лабораторная работа 17. Процедуры экспертного оценивания.

Программа расчета коэффициента конкордации

Задача состоит в сопоставлении оцениваемой системе одной перестановки. В данном случае процедура характеризуется параметрами:

Ω представляет множество всех перестановок объектов;

Ω_{Σ} - Ω ;

L – эксперты изолированы;

Q – обратная связь отсутствует.

Отображение определяется следующим образом. Результаты опроса экспертов сводятся в таблицу (табл. __). В i-ой строке стоят места (ранги), данные i-ым экспертом ранжируемым объектам. В (N+1)-й строке стоят суммы рангов, полученных объектами от экспертов.

Эксперты	Объекты			
	1	2	...	n
1	r_{11}	r_{12}	...	r_{1n}
2	r_{21}	r_{22}	...	r_{2n}
	...	...	...	...
N	r_{N1}	r_{N2}	...	r_{Nn}
Σ ранг	r_1	r_2	...	r_n

Все n объектов упорядочиваются в соответствии с величиной r_{ij} , порядок последовательности определяется на основании:

$$r_s = \sum_{j=1}^N r_{sj}.$$

Рассмотрим два крайних случая. Первый случай: ранжирование всех экспертов совпадают. Каждый объект получил от всех экспертов одинаковый ранг, который для j-го объекта равен r_j/N . Второй случай: полная несогласованность экспертов. Будем понимать под несогласованностью противоположность ранжировок, даваемых экспертами.

В силу () получаем

$$\sum_{j=1}^N r_{ij} = \frac{n(n+1)}{2},$$

что приводит к справедливости выражения

$$\sum_{j=1}^n r_i = \sum_{i=1}^N \sum_{j=1}^n r_{ij} = \frac{N \cdot n(n+1)}{2}.$$

За средний ранг принимаем величину

$$r_{icp} = \frac{1}{n} \sum_{j=1}^n r_{ij} = \frac{N \cdot (n+1)}{2},$$

а за степень согласованности мнений – сумму квадратов отклонений r_j от среднего значения r_{icp} .

Степень согласованности определяется коэффициентом конкордации W (по аналогии с

определением дисперсии распределения $\sum_{j=1}^N p_i (r_{ij} - \bar{r}_j)^2$)

$$W = \frac{12 \cdot \sum_{j=1}^N \left(r_i - \frac{1}{2} N(n+1) \right)^2}{N^2 (n^3 - n)}$$

В случае нестроого ранжирования коэффициент конкордации равен

$$W = \frac{12 \cdot \sum_{j=1}^N \left(r_i - \frac{1}{2} N(n+1) \right)^2}{N^2 (n^3 - n) - N \sum_{i=1}^N \sum_{j=1}^{K_i} (r_{ij}^3 - r_{ij})}.$$

Представление функции выбора

$$\beta(X) = \langle \beta_1(X), \beta_2(X), \dots, \beta_N(X) \rangle$$

$$\beta_i(X) = 1 \text{ при } x_i \in X, \beta_i(X) = 0 \text{ при } x_i \notin X$$

Рассматривается семейство булевых функций $f_i(\dots)$ $i=1..N$

$$f_1(v_2, v_3, \dots, v_N), \dots, f_i(v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_N), \dots, f_N(v_1, v_2, \dots, v_{N-1})$$

$$\beta_i(X) \wedge f_i(\beta(X)) = 1 \Leftrightarrow x_i \in C(X)$$

Логической формой функции выбора называется семейство функций

$$\langle f_1, \dots, f_N \rangle$$

Пример:

X	C(X)	$\beta(X)$	$\beta(C(X))$
X_1	X_1	1000	1000

X_2	X_2	0100	0100
X_3	X_3	0010	0010
X_4	X_4	0001	0001
X_1, X_2	X_1	1100	1000
X_1, X_3	X_1	1010	1000
X_1, X_4	X_1	1001	1000
X_2, X_3	X_2	0110	0100
X_2, X_4	X_2	0101	0100
X_3, X_4	X_3	0011	0010
X_1, X_2, X_3	X_1, X_2	1110	1100
X_1, X_2, X_4	X_1, X_2	1101	1100
X_1, X_3, X_4	X_1, X_3	1011	1010
X_2, X_3, X_4	X_2, X_3	0111	0110
X_1, X_2, X_3, X_4	X_1	1111	1000

β_2	β_3	β_4	f_1		β_1	β_3	β_4	f_2
0	0	0	1		0	0	0	1
0	0	1	1		0	0	1	1
0	1	0	1		0	1	0	1
0	1	1	1		0	1	1	1
1	0	0	1		1	0	0	0
1	0	1	1		1	0	1	1
1	1	0	1		1	1	0	1
1	1	1	1		1	1	1	0

$$f_1 \equiv 1$$

$$f_2 = \neg\beta_1 \vee \neg\beta_3\beta_4 \vee \beta_3\neg\beta_4$$

β_1	β_2	β_4	f_3		β_1	β_2	β_3	f_4
0	0	0	1		0	0	0	1
0	0	1	1		0	0	1	0

0	1	0	0	0	1	0	0
0	1	1	1	0	1	1	0
1	0	0	0	1	0	0	0
1	0	1	1	1	0	1	0
1	1	0	0	1	1	0	0
1	1	1	0	1	1	1	0

$$f_3 = \neg\beta_1 \neg\beta_2 \vee \neg\beta_2 \beta_4 \vee \neg\beta_1 \beta_4$$

$$f_4 = \neg\beta_1 \neg\beta_2 \neg\beta_3$$

Список использованных источников

1. Фаддеев М.А. Элементарная обработка результатов эксперимента: Учебное пособие (Учебно-научный и инновационный комплекс "Новые многофункциональные материалы и нанотехнологий"). - Нижний Новгород: Нижегородский госуниверситет, 2010. - 122 с. - <http://window.edu.ru/resource/847/79847>в НТБ нет
2. Жиляков Е.Г., Перлов Ю.М., Ревтова Е.П. Основы эконометрического анализа данных: учебное пособие. - Белгород: Белгородский гос. ун-т, 2004. - 104 с. - <http://window.edu.ru/resource/893/62893>в НТБ нет
3. Горицкий Ю.А., Перцов Е.Е. Практикум по статистике с пакетами StatGraphics, Statistica, SPSS Московский энергетический институт. Образовательный математический сайт Exponenta.ru <http://www.exponenta.ru/educat/systemat/goritskii/lr.asp>в НТБ нет
4. Зайчикова Н.А. Математико-статистические методы анализа данных: Учебно-методическая разработка. - Самара: Самарский институт (филиал) РГТЭУ, 2010. - 61 с. - <http://window.edu.ru/resource/054/69054>в НТБ нет
5. Горицкий Ю.А. Практикум по статистике с пакетом Statistica Московский энергетический институт. Образовательный математический сайт Exponenta.ru <http://www.exponenta.ru/educat/systemat/goritskii/part2/lr.asp>в НТБ нет
6. Володин И.Н. Лекции по теории статистических выводов: Учебное пособие. - Казань: Казанский (Приволжский) федеральный университет, 2010. - 174 с. - <http://window.edu.ru/resource/071/76071>,в НТБ нет