МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
«Курганский государственный университет»

Кафедра «Инноватика и менеджмент качества»

ИНЖЕНЕРНОЕ И УПРАВЛЕНЧЕСКОЕ МОДЕЛИРОВАНИЕ В КОМПЬЮТЕРНОЙ СИСТЕМЕ

Лабораторный практикум для студентов, обучающихся по направлениям 221700.62 «Стандартизация и метрология» и 222000.62 «Инноватика»

Курган 2013

Кафедра: «Инноватика и менеджмент качества»
Дисциплина: «Инженерное и управленческое моделирование в компьютерной системе» (направления 221700.62, 222000.62)

Составил: канд. техн. наук, доцент В.Е. Овсянников

Утверждены на заседании кафедры «5» марта 2013 г.
Рекомендованы методическим советом университета «12» марта 2013 г.

Содержание

Введение

Целью освоения дисциплины «Инженерное и управленческое моделирование в компьютерной системе» является формирование способностей и готовности специалиста применять аппарат математического моделирования и средства его компьютерной реализации при решении задач инженерной практики.

Задачами освоения дисциплины «Инженерное и управленческое моделирование в компьютерной системе» являются ознакомление студентов с основными методами построения математических моделей линейных и нелинейных объектов; изучение основ нейросетевых технологий; формирование навыков программной реализации математических моделей и обработки информации в компьютерной системе.

В результате изучения базовой части дисциплины студент должен:

- владеть знаниями содержания и способов использования компьютерных и информационных технологий;
- обладать умениями применять компьютерную технику и информационные технологии в своей профессиональной деятельности;
- владеть средствами компьютерной техники и информационных технологий.

В ходе освоения дисциплины «Инженерное и управленческое моделирование в компьютерной системе» предполагается выполнение комплекса лабораторных работ. Для студентов очной формы обучения предусмотрено 8 лабораторных работ, охватывающих вопросы математической статистики, корреляционного и регрессионного анализа данных и моделирования искусственных нейронных сетей. Студенты заочной формы обучения выполняют 7 лабораторных работ.

1 РУКОВОДСТВО ПО ВЫБОРУ ЗАДАНИЯ И ПОДГОТОВКЕ ИСХОДНЫХ ДАННЫХ

Лабораторные работы выполняются в соответствии с вариантом. Номер варианта соответствует двум последним цифрам зачетной книжки студента. Массив исходных данных формируется посредством прибавления номера варианта к исходным данным, приведенным в таблице 1.

Таблица 1 - Исходные данные

X1	X2	Y
800+№	772.5+№	257.5+№
942.5+№	847.5+№	282.5+№
972.5+№	945+№	315+№
1025+№	1050+№	350+№

Продолжение таблицы 1

X1	X2	Y
665+№	525+№	175+№
1077.5+№	900+№	300+№
1107.5+№	1215+№	405+№
830+№	825+№	275+№
1017+№	967+№	322+№
1175+№	1207.5+№	402+№
440+№	1125+№	425+№
1287+№	1275+№	335+№
980+№	1005+№	365+№
1077.5+№	1095+№	442+№
1235+№	1327.5+№	475+№
1415+№	1425+№	320+№
1055+№	960+№	380+№
1167.5+№	1140+№	237.5+№
927.5+№	712.5+№	325+№
1182.5+№	975+№	437.5+№
1317.5+№	1312.5+№	247.5+№
905+№	697.5+№	232.5+№
717.5+№	1117.5+№	372.5+№
1205+№	975+№	325+№
1077.5+№	922.5+№	307.5+№
777.5+№	600+№	200+№
1047.5+№	937.5+№	312.5+№
1212.5+№	1027.5+№	342.5+№
1415+№	375+№	325+№
1152.5+№	975+№	290+№
845+№	870+№	387.5+№

Примечания: № - номер варианта.

Исходные данные оформляются в виде текстовых документов формата txt в программе блокнот (рисунок 1).

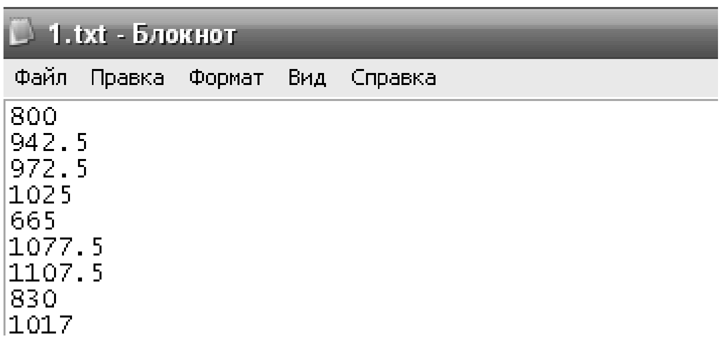


Рисунок 1- Формат исходных данных

<u>Считывание исходных данных</u> осуществляется при помощи встроенного компонента MathCAD «Импорт данных» (Data import wizard). Процесс считывания исходных данных представлен на рисунках 2 и 3:

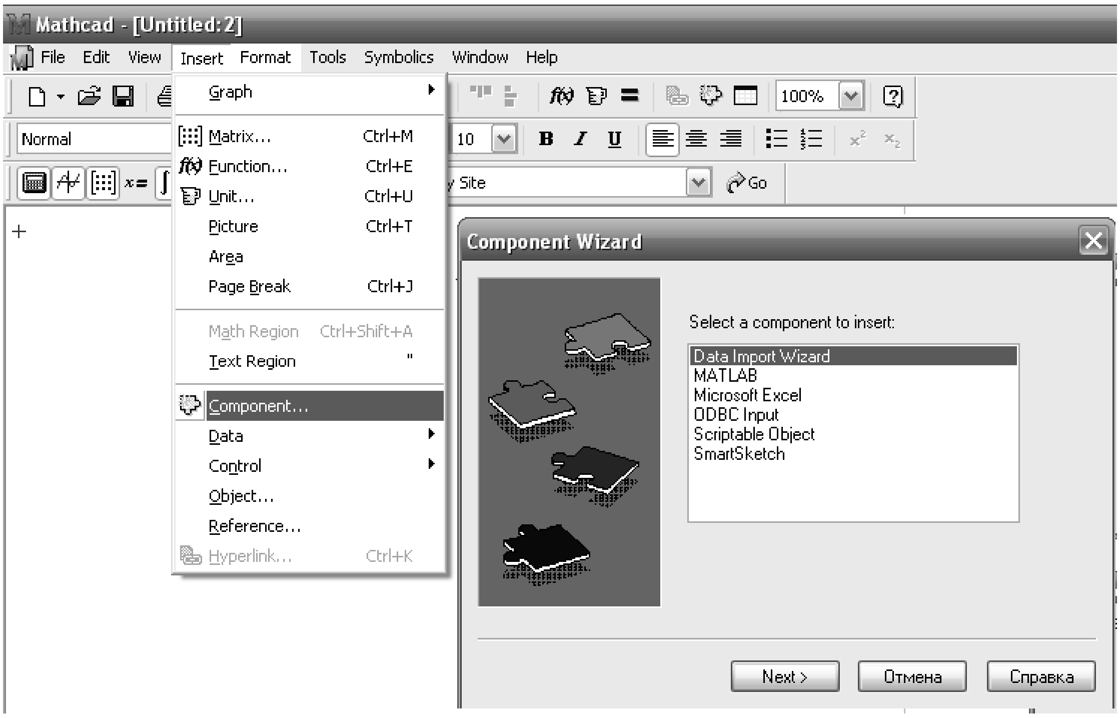


Рисунок 2 - Вставка компонента в документ MathCAD

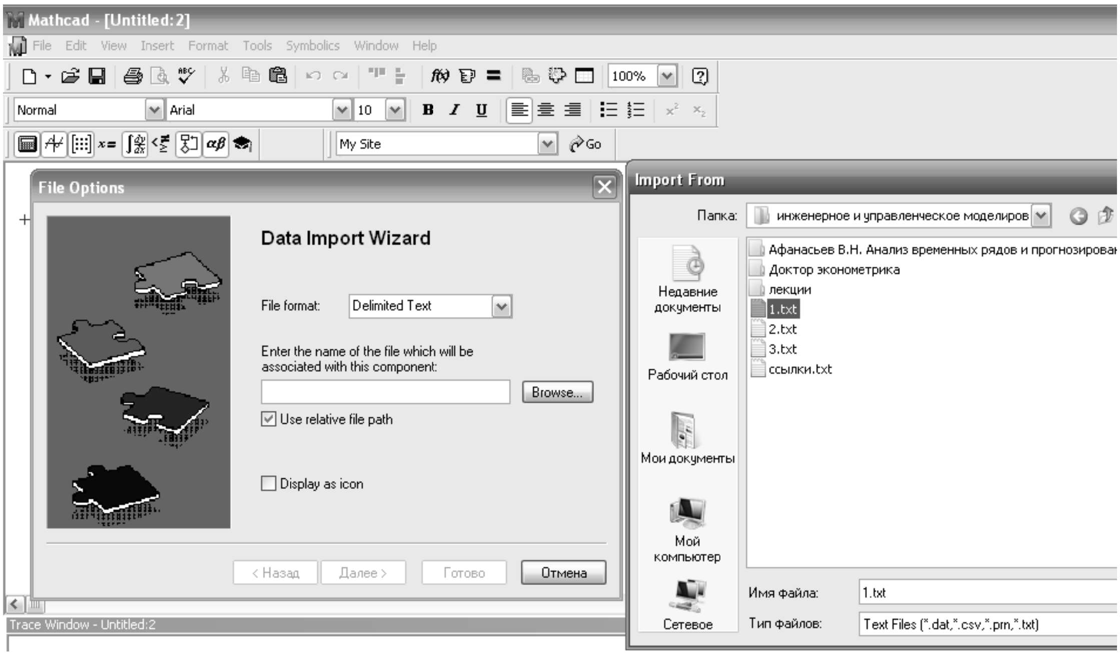


Рисунок 3 - Считывание данных

2 СОДЕРЖАНИЕ ЛАБОРАТОРНЫХ РАБОТ

2.1 Лабораторная работа №1

Тема работы: определение численных характеристик выборки.

Цель работы: ознакомиться с методикой первичной обработки данных выборки, научиться выполнять группировку данных, представлять выборку в графическом виде и вычислять ее численные характеристики с использованием средств MathCAD.

Теоретическая часть

Предварительная статистическая обработка опытных данных начинается обычно с того, что их располагают в порядке возрастания (неубывания). Упорядоченная таким образом выборка называется вариационным рядом, а сама процедура упорядочения – *ранжированием* (или сортировкой) опытных данных.

Формой графического представления эмпирических данных является *гистограмма и полигон*.

Весь интервал $[x_{\min}, x_{\max}]$, в котором заключены элементы выборки, разбивается на ряд частичных интервалов (a_i, b_i) длины h и подсчитывается число элементов выборки n_i, попавших в i-й интервал $i = 1,2,\ldots,m$. Параллельно вычисляется и относительная частота $w_i = n_i / n$. При графическом изображении гистограммы и полигона каждый интервал удобнее представлять не двумя границами a_i и b_i, а одним значением $\bar{x}_i = a_i + h/2$ - серединой интервала.

Гистограммой частот называется ступенчатая фигура, состоящая из прямоугольников, основаниями которых служат частичные интервалы длины h, а высоты равны отношению n_i / h (плотность частоты).

Полигон частот – это ломаная линия, получающаяся при соединении точек с координатами $(\bar{x}_i,\ n_i / h)$, т.е. соединяются середины верхних сторон прямоугольников гистограммы.

Доверительная область для функции распределения, соответствующая уровню доверия $p = 1 - \alpha$, определяется неравенствами

$$\hat{F}_n(x) - \frac{z_\alpha}{\sqrt{n}} \le F(x) \le \hat{F}_n(x) + \frac{z_\alpha}{\sqrt{n}}, \qquad (1)$$

где z_α - корень уравнения,

$K(z_\alpha) = 1 - \alpha$,

$K(z)$ - функция (распределения) Колмогорова, определяемая абсолютно сходящимся функциональным рядом $(z > 0)$.

Числовые характеристики выборки вычисляются по следующим формулам:

- эмпирическое (выборочное) среднее

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i\ ; \qquad (2)$$

- выборочная дисперсия

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \equiv \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \bar{x}^2, \quad \text{или} \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2; \qquad (3)$$

- стандартное отклонение

$$S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}, \quad \text{или} \quad s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}; \qquad (4)$$

- размах выборки

$$R = x_{\max} - x_{\min}; \qquad (5)$$

- эмпирический центральный момент k-го порядка

$$\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^k, \quad \text{или} \quad \hat{\mu}_k = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^k; \qquad (6)$$

- эмпирические коэффициенты асимметрии и эксцесса

$$\hat{A} = \frac{\hat{\mu}_3}{\hat{\mu}_2^{3/2}}, \qquad \hat{E} = \frac{\hat{\mu}_4}{\hat{\mu}_2^2} - 3. \qquad (7)$$

Задание: определить численные характеристики выборки, построить гистограмму и полигон частот, вычислить доверительную область.

Пример выполнения лабораторной работы №1 в среде MathCAD

Считывание исходных данных из файла на жестком диске

X :=

	0
0	800
1	942.5
2	972.5
3	$1.025\cdot10^3$
4	665
5	$1.077\cdot10^3$

Определение числа элементов в выборке:

n := rows(X)

Вычисление размаха выборки:

Xmax := max(X) Xmax = 1.415×10^3

Xmin := min(X) Xmin = 440

Сортировка данных:

X := sort(X)

Определение числа интервалов группировки:

m := ceil(5·log(n)) ceil(1 + 1.44·ln(n)) = 6 ceil(1 + 3.32·log(n)) = 6

Вычисление длины группировки интервала:

$\Delta := \frac{\text{ceil(Xmax)} - \text{floor(Xmin)}}{\text{m}}$ Δ = 121.875

Определение параметров группировки данных:

$i := 0..\, m-1$

$a_i := \mathrm{floor}(Xmin) + \Delta \cdot i \qquad b_i := a_i + \Delta$

$x_i := \dfrac{a_i + b_i}{2}$

$\min(a) = 440 \qquad \max(a) = 1.293\times 10^{3} \quad j := 0..\, m \quad X1_j := \mathrm{floor}(Xmin) + \Delta \cdot j$

Определение вектора относительных частот:

$f := \dfrac{\mathrm{hist}(X1, X)}{n\cdot\Delta} \qquad \sum_{j=0}^{m-1} \left(f_j \cdot \Delta\right) = 1$

$$X1 = \begin{pmatrix} 440 \\ 561.875 \\ 683.75 \\ 805.625 \\ 927.5 \\ 1.049\times 10^{3} \\ 1.171\times 10^{3} \\ 1.293\times 10^{3} \\ 1.415\times 10^{3} \end{pmatrix} \quad x = \begin{pmatrix} 500.938 \\ 622.813 \\ 744.688 \\ 866.563 \\ 988.438 \\ 1.11\times 10^{3} \\ 1.232\times 10^{3} \\ 1.354\times 10^{3} \end{pmatrix} \quad f = \begin{pmatrix} 2.647\times 10^{-4} \\ 2.647\times 10^{-4} \\ 7.94\times 10^{-4} \\ 7.94\times 10^{-4} \\ 1.853\times 10^{-3} \\ 1.853\times 10^{-3} \\ 1.588\times 10^{-3} \\ 7.94\times 10^{-4} \end{pmatrix}$$

Построение гистрограммы и полигона частот средствами MathCAD:

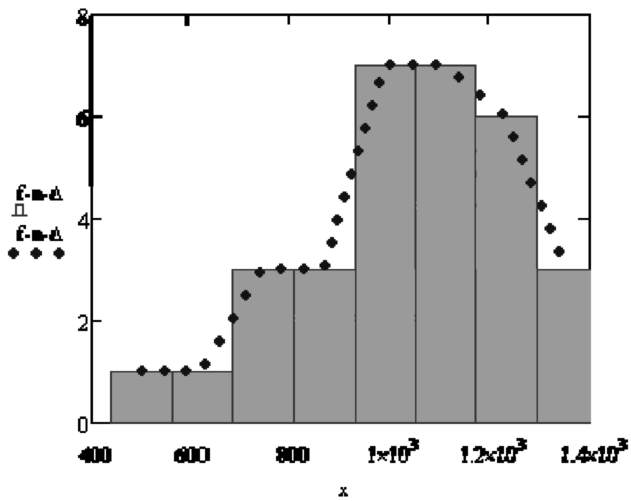


Рисунок 4 - Гистрограмма и полигон частот

Определение вектора относительных накопленных частот:

$f := \dfrac{\mathrm{hist}(X1, X)}{n} \qquad k := 0..\, m-1$

$$F_k := \sum_{i=0}^{k} f_i$$

$$F = \begin{pmatrix} 0.032 \\ 0.065 \\ 0.161 \\ 0.258 \\ 0.484 \\ 0.71 \\ 0.903 \\ 1 \end{pmatrix}$$

Построение полигона относительных накопленных частот:

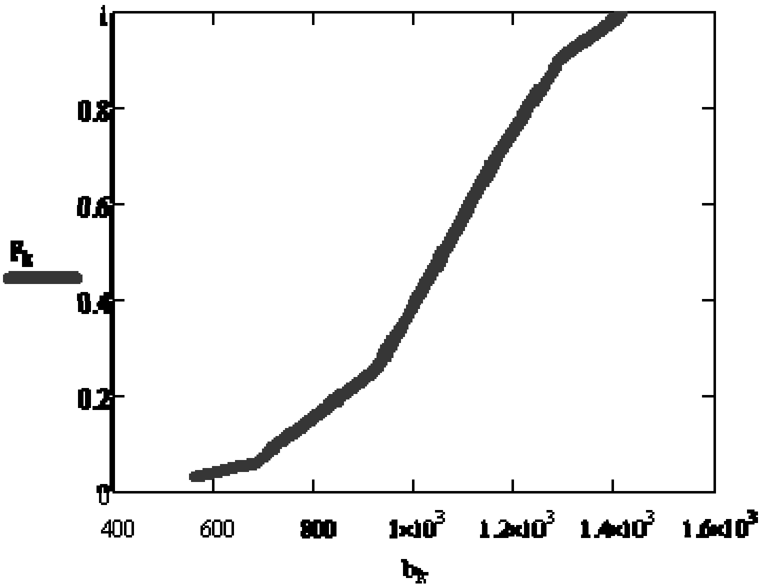


Рисунок 5 - Полигон относительных накопленных частот

Определение числовых характеристик выборки:
Выборочное среднее:

$$Xm := \text{mean}(X) \qquad Xm1 := \frac{1}{n} \cdot \sum_{i=0}^{n-1} X_i$$

$$Xm = 1.034 \times 10^3 \qquad Xm1 = 1.034 \times 10^3$$

Дисперсия и среднеквадратическое отклонение:

$$dx := \text{var}(X) \qquad dx1 := \frac{1}{n} \cdot \sum_{i=0}^{n-1} \left(X_i - Xm\right)^2$$

$$dx = 4.684 \times 10^4 \qquad dx1 = 4.684 \times 10^4$$

$$\text{stdev}(X) = 216.436 \qquad \sqrt{dx1} = 216.436 \qquad DX := \text{Var}(X)$$

$$DX1 := \frac{1}{n-1}\cdot\sum_{i=0}^{n-1}\left(X_i - Xm\right)^2$$

$DX = 4.841 \times 10^4$ $DX1 = 4.841\times 10^4$ $Stdev(X) = 220.01$ $\sqrt{DX1} = 220.01$

Медиана:

$median(X) = 1.055\times 10^3$

Моменты 3 и 4 порядка, асимметрия и эксцесс:

$$\mu 3 := \frac{1}{n}\cdot\sum_{i=0}^{n-1}\left(X_i - Xm\right)^3 \qquad \mu 4 := \frac{1}{n}\cdot\sum_{i=0}^{n-1}\left(X_i - Xm\right)^4$$

$$A := \frac{\mu 3}{Stdev(X)^3} \qquad skew(X) = -0.54 \qquad A = -0.489$$

$$E := \frac{\mu 4}{Stdev(X)^4} - 3 \qquad kurt(X) = 0.524 \qquad E = 0.049$$

Несмещенные оценки асимметрии и эксцесса:

$$A := n\cdot\frac{\sum_{i=0}^{n-1}\left(X_i - Xm\right)^3}{(n-1)\cdot(n-2)\cdot DX1^{1.5}}$$

$$E := \frac{\left[n\cdot(n+1)\cdot\sum_{i=0}^{n-1}\left(X_i - Xm\right)^4\right] - 3\cdot\left[\sum_{i=0}^{n-1}\left(X_i - Xm\right)^2\right]^2\cdot(n-1)}{(n-1)\cdot(n-2)\cdot(n-3)\cdot DX1^2}$$

$A = -0.54$ $E = 0.524$

Множитель C(n), входящий в оценку несмещенного стандартного отклонения:

$$C(n) := \sqrt{\frac{n-1}{2}}\cdot\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}$$

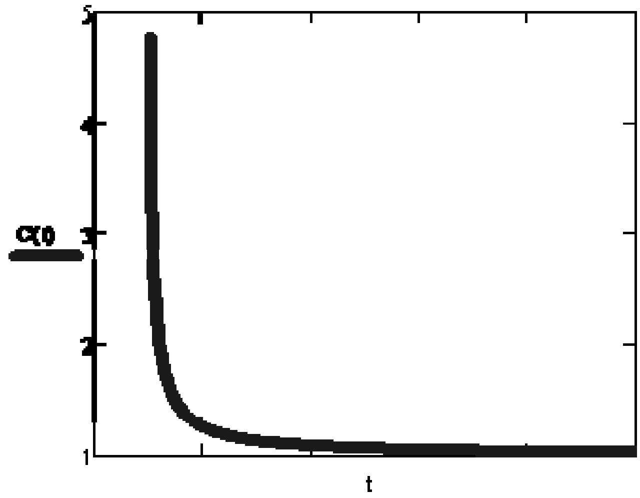


Рисунок 6 - Функциональная зависимость C(t)

Функция Колмогорова и ее график:

$$K(z,n) := \begin{cases} 0 & \text{if } z \le 0.3 \\ \left[1 - 2\cdot\sum_{k=1}^{n}\left[(-1)^{k+1}\cdot e^{-2\cdot k^2\cdot z^2}\right]\right] & \text{otherwise}\end{cases}$$

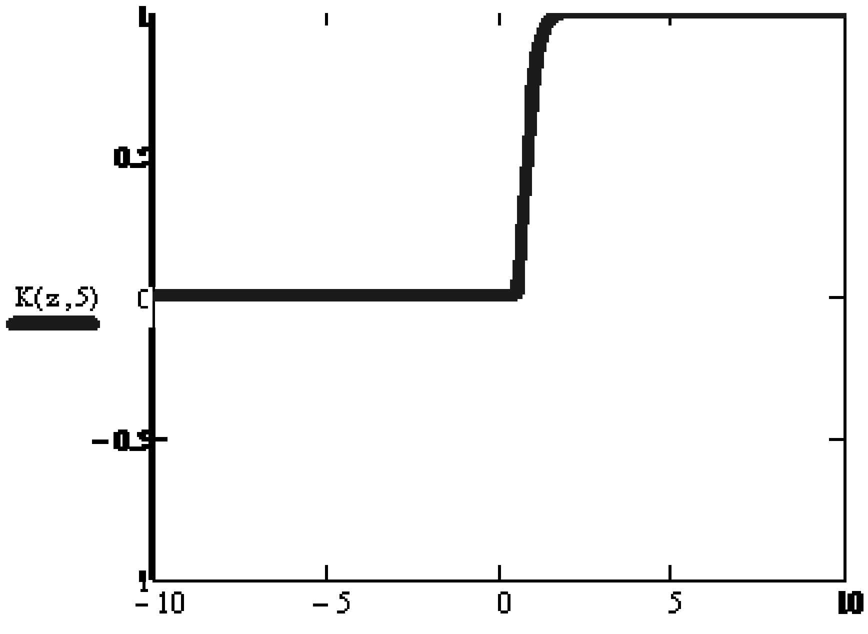


Рисунок 7 - График функции Колмогорова

Вычисление корней функции Колмогорова:

$u := 1 \qquad \alpha := 0.05$

Given

$K(u,5) = 1 - \alpha$

$z := \mathrm{Find}(u) \qquad z = 1.358$

Определение границ доверительной области для функции распределения:

$$k := 0..m-1$$

$$F1_k := F_k - \frac{z}{\sqrt{n}} \qquad F1_k := \begin{cases} 0 & \text{if } F1_k < 0 \\ F1_k & \text{otherwise} \end{cases}$$

$$F2_k := F_k + \frac{z}{\sqrt{n}} \qquad F2_k := \begin{cases} 0 & \text{if } F2_k < 0 \\ F2_k & \text{otherwise} \end{cases}$$

Доверительная область и эмпирическая функция распределения:

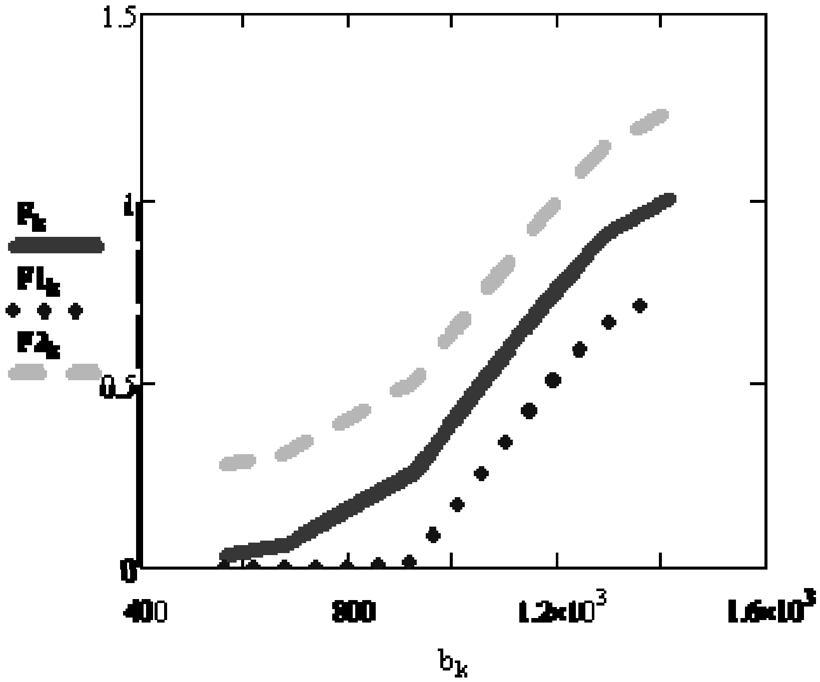


Рисунок 8 - Доверительная область функции распределения

2.2 Лабораторная работа №2

Тема работы: точечные и интервальные оценки параметров нормально распределенной случайной величины.

Цель работы: ознакомиться с методикой определения точечных и интервальных оценок выборочного среднего и дисперсии случайной величины.

Теоретическая часть

Случайный интервал, являющийся доверительной оценкой величины, a с надежностью 1-α:

$$\left(\bar{x} - u_q \frac{\sigma}{\sqrt{n}},\ \ \bar{x} + u_q \frac{\sigma}{\sqrt{n}} \right), \tag{8}$$

где $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ - точечная оценка измеряемой величины, u_q - квантиль стандартного нормального распределения уровня $q = 1 - \alpha / 2$, определяемый из решения уравнения

$$\Phi(u_q) = 1 - \alpha / 2, \tag{9}$$

$\Phi(x)=\frac{1}{\sigma\sqrt{2\pi}}\int\limits_{-\infty}^{x}\exp\left(-\frac{t^2}{2}\right)dt$ - функция стандартного нормального распределения, α - уровень значимости.

Если точность измерений не известна, то сначала вычисляется оценка среднеквадратичной ошибки

$$s=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}\,, \tag{10}$$

а затем строится доверительный интервал

$$\left(\bar{x}-t_{q,n-1}\frac{s}{\sqrt{n}},\ \ \bar{x}+t_{q,n-1}\frac{s}{\sqrt{n}}\right), \tag{11}$$

где $t_{q,n-1}$ - квантиль уровня $q=1-\alpha/2$ распределения Стьюдента с $n-1$ степенями свободы, определяемый как корень уравнения

$$F_{n-1}\left(t_{q,n-1}\right)=q\equiv 1-\alpha/2\,, \tag{12}$$

$F_{n-1}(x)$ - функция распределения Стьюдента с $n-1$ степенями свободы. Если для оценки среднеквадратичной ошибки используется смещенная оценка дисперсии

$$S=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}\,, \tag{13}$$

то доверительный интервал для параметра a записывается в виде

$$\left(\bar{x}-t_{q,n-1}\frac{S}{\sqrt{n-1}},\ \ \bar{x}+t_{q,n-1}\frac{S}{\sqrt{n-1}}\right). \tag{14}$$

Доверительные интервалы при известном и неизвестном a:

$$\frac{n\hat{\sigma}^2}{c_{R,p}^2}<\sigma^2<\frac{n\hat{\sigma}^2}{c_{L,p}^2} \qquad \text{и} \qquad \frac{(n-1)s^2}{d_{R,p}^2}<\sigma^2<\frac{(n-1)s^2}{d_{L,p}^2}\,,$$

где

$$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-a\right)^2\,, \qquad s^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2\,.$$

Квантили $c_{L,p}^2, c_{R,p}^2$ и $d_{L,p}^2, d_{R,p}^2$ распределения «хи-квадрат» находятся как корни уравнений:

$$\chi_n^2\left(c_{L,p}^2\right)=\frac{\alpha}{2},\ \ \chi_n^2\left(c_{R,p}^2\right)=1-\frac{\alpha}{2}, \tag{15}$$

$$\chi_{n-1}^2\left(d_{L,p}^2\right)=\frac{\alpha}{2}, \qquad \chi_{n-1}^2\left(d_{R,p}^2\right)=1-\frac{\alpha}{2} \tag{16}$$

при n и $n-1$ степенях свободы.

<u>Задание:</u> Определить доверительные границы математического ожидания и дисперсии при известном и неизвестном параметре распределения.

Пример выполнения лабораторной работы №2 в среде MathCAD:

Считывание исходных данных из файла на жестком диске

$X :=$

	0
0	800
1	942.5
2	972.5
3	$1.025\cdot10^3$
4	665
5	$1.077\cdot10^3$

Параметры нормально распределенной выборки:

$n := \mathrm{rows}(X) \qquad \mu := 10 \qquad \sigma := 2 \qquad \sigma^2 = 4$

Определение точечных оценок математического ожидания и дисперсии:

$$Mx := \frac{1}{n}\cdot\sum_{i=0}^{n-1} X_i \qquad S2 := \frac{1}{n}\cdot\sum_{i=0}^{n-1}\left(X_i - Mx\right)^2 \qquad s2 := \frac{1}{n-1}\cdot\sum_{i=0}^{n-1}\left(X_i - Mx\right)^2$$

$Mx = 1.034\times10^3 \quad S2 = 4.684\times10^4 \quad \sqrt{S2} = 216.436 \quad s2 = 4.841\times10^4 \quad \sqrt{s2} = 220.014$

Построение 95%-х доверительных интервалов:

Уровень значимости: $\alpha := 0.05$

Квантиль нормального распределения: $u := \mathrm{qnorm}\left(1 - \frac{\alpha}{2}, 0, 1\right) \quad u = 1.96$

Доверительные границы для математического ожидания и дисперсии:

$$XL := Mx - u\cdot\frac{\sigma}{\sqrt{n}} \qquad XR := Mx + u\cdot\frac{\sigma}{\sqrt{n}} \qquad Lx := XR - XL$$

$XL = 1.033\times10^3 \qquad XR = 1.035\times10^3 \qquad Lx = 1.408$

Доверительные границы для мат. ожидания при неизвестном σ:

$t := \mathrm{qt}\left(1 - \frac{\alpha}{2}, n-1\right) \qquad t = 2.042$

$$XL := Mx - t\cdot\sqrt{\frac{s2}{n}} \qquad XR := Mx + \sqrt{\frac{s2}{n}} \qquad Lx := XR - XL$$

$XL = 953.218 \qquad XR = 1.073\times10^3 \qquad Lx = 120.217$

В случае использования смещенной оценки дисперсии:

$$XL := Mx - t\cdot\sqrt{\frac{S2}{n-1}} \qquad XR := Mx + \sqrt{\frac{S2}{n-1}} \qquad Lx := XR - XL$$

$XL = 953.218 \qquad XR = 1.073\times10^3 \qquad Lx = 120.217$

Доверительные границы для дисперсии при математическом ожидании μ

$cL := \mathrm{qchisq}\left(\frac{\alpha}{2}, n\right) \qquad cL = 17.539$

$cR := qchisq\left(1 - \frac{\alpha}{2}, n\right)$ $\quad cR = 48.232$

cR, cL - кванитили распределения Хи-квадрат с n - степенями свободы:

$dL := \frac{S2 \cdot n}{cR}$ $\quad dR := \frac{S2 \cdot n}{cL}$ $\quad Ld := dR - dL$

$dL = 3.011 \times 10^{4}$ $\quad dR = 8.28 \times 10^{4}$ $\quad Ld = 5.269 \times 10^{4}$

Доверительные интервалы при неизвестном математическом ожидании μ:

$cL := qchisq\left(\frac{\alpha}{2}, n - 1\right)$ $\quad cL = 16.791$

$cR := qchisq\left(1 - \frac{\alpha}{2}, n - 1\right)$ $\quad cR = 46.979$

$dL := \frac{s2 \cdot n - 1}{cR}$ $\quad dR := \frac{s2 \cdot n - 1}{cL}$ $\quad Ld := dR - dL$

Построение 99%-х доверительных интервалов:

$\alpha := 0.01$ уровень значимости $\quad u := qnorm\left(1 - \frac{\alpha}{2}, 0, 1\right)$ $\quad u = 2.576$

Квантиль нормального распределения:

Доверительные границы для математического ожидания и дисперсии:

$XL := Mx - u \cdot \frac{\sigma}{\sqrt{n}}$ $\quad XR := Mx + u \cdot \frac{\sigma}{\sqrt{n}}$ $\quad Lx := XR - XL$

$XL = 1.033 \times 10^{3}$ $\quad XR = 1.035 \times 10^{3}$ $\quad Lx = 1.851$

Доверительные границы для мат. ожидания при неизвестном σ:

$t = 2.75$ t - квантиль распределения Стьюдента

$t := qt\left(1 - \frac{\alpha}{2}, n - 1\right)$

$Lx := XR - XL$

$XL := Mx - t \cdot \sqrt{\frac{s2}{n}}$ $\quad XR := Mx + \sqrt{\frac{s2}{n}}$

$XL = 925.251$ $\quad XR = 1.073 \times 10^{3}$ $\quad Lx = 148.184$

Определение доверительных границ для дисперсии при математическом ожидании μ:

$cL := qchisq\left(\frac{\alpha}{2}, n\right)$ $\quad cL = 14.458$

$cR := qchisq\left(1 - \frac{\alpha}{2}, n\right)$ $\quad cR = 55.003$

$dL := \frac{S2 \cdot n}{cR}$ $\quad dR := \frac{S2 \cdot n}{cL}$ $\quad Ld := dR - dL$

$dL = 2.64 \times 10^{4}$ $\quad dR = 1.004 \times 10^{5}$ $\quad Ld = 7.404 \times 10^{4}$

Доверительные границы при неизвестном математическом ожидании μ:

$cL := qchisq\left(\frac{\alpha}{2}, n - 1\right)$ $\quad cL = 13.787$

$$cR := \text{qchisq}\left(1 - \frac{\alpha}{2}, n - 1\right) \qquad cR = 53.672$$

$$dL := \frac{s2\cdot(n-1)}{cR} \qquad dR := \frac{s2\cdot(n-1)}{cL} \qquad Ld := cR - cL$$

$$dL = 2.706\times 10^{4} \qquad dR = 1.053\times 10^{5} \qquad Ld = 39.885$$

2.3 Лабораторная работа №3

Тема работы: отсев грубых ошибок экспериментальных данных и оценка нормальности распределения.

Цель работы: ознакомиться с методикой проверки основных гипотез об однородности выборки и оценки соответствия нормальному закону распределения.

Теоретическая часть

Отсев грубых ошибок базируется на том, что критические значения максимального относительного отклонения

$$\tau = \frac{|x_* - \bar{x}|}{s} \tag{17}$$

выражаются через квантили распределения Стьюдента с $n-2$ степенями свободы:

$$\tau_{1-\alpha,n} = \frac{t_{1-\alpha,n-2}\sqrt{n-1}}{\sqrt{n-2+t^2_{1-\alpha,n-2}}} \tag{18}$$

при значениях $\alpha = 0.05$ и $\alpha = 0.001$.

Этими значениями вся область изменения τ разбивается на три интервала:

1) $-\infty < \tau \le \tau_1$;
2) $\tau_1 < \tau < \tau_2$;
3) $\tau_2 \le \tau < +\infty$.

Данные, попавшие в первый интервал, не подлежат отсеву. Наблюдения, попавшие во второй интервал, можно исключить, если имеются какие-либо дополнительные соображения в пользу их ошибочности. Наконец, наблюдения, попавшие в третий интервал, всегда отбрасываются как грубо ошибочные.

Проверка нормальности распределения осуществляется по результатам вычисления коэффициентов асимметрии, эксцесса и их дисперсий:

$$\widehat{A} = \frac{\widehat{\mu}_3}{s^3} \approx \frac{1}{s^3 n}\sum_{i=1}^{n}(x_i - \bar{x})^3, \qquad \widehat{E} = \frac{\widehat{\mu}_4}{s^4} \approx \frac{1}{s^4 n}\sum_{i=1}^{n}(x_i - \bar{x})^4 - 3,$$

$$D(A) = \frac{6(n-2)}{(n+1)(n+3)}, \qquad D(E) = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}.$$

Если вычисленные коэффициенты удовлетворяют условию:

$$|\widehat{A}| \le 3\sqrt{D(\widehat{A})} \qquad |\widehat{E}| \le 5\sqrt{D(\widehat{E})}, \tag{19}$$

то гипотеза о нормальности наблюдаемого распределения принимается, в противном случае гипотеза отклоняется.

Если выборка достаточно велика, применяются иные критерии согласия, наиболее надежным и универсальным из которых является критерий Пирсона χ^2. Применяя данный критерий, необходимо выполнить следующие действия.

Область возможных значений случайной величины $(-\infty,+\infty)$ разбивается на конечное число ($m \approx 8 \div 20$) непересекающихся интервалов:

$$(-\infty, x_2), (x_2, x_3), (x_3, x_4), \ldots, (x_m, +\infty)$$

Для каждого интервала (x_{i-1}, x_i) подсчитывается число n_i элементов выборки, попавших в данный интервал.

Вычисляется теоретическая вероятность p_i попадания в i-й интервал при нормальном законе распределения вероятностей

$$p_i \equiv P(x_{i-1} < X < x_i) = \Phi_0\left(\frac{x_i - \bar{x}}{s}\right) - \Phi_0\left(\frac{x_{i-1} - \bar{x}}{s}\right), \qquad (20)$$

где $\Phi_0(x)$ - функция Лапласа.

Проверяется выполнение условия $np_i \ge 5$ для всех интервалов; интервалы, для которых это условие не выполнено, объединяются с соседними интервалами.

Вычисляется сумма

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}, \qquad (21)$$

имеющая приближенно χ^2- распределение с $k-3$ степенями свободы.

При заданной доверительной вероятности $p = 1-\alpha$ (α - уровень значимости) и числе степеней свободы $k-3$ вычисляется (или находится по таблицам) критическое значение критерия $\chi^2_{p,k-3}$. Если $\chi^2 < \chi^2_{p,k-3}$, то эмпирическое распределение считается нормальным.

Задание: выявить наличие грубых ошибок в выборке и при необходимости произвести их отсев, проверить соответствие рассматриваемой выборки нормальному закону распределения.

Пример выполнения лабораторной работы №3 в среде MathCAD

Считывание исходных данных из файла на жестком диске

X :=

	0
0	800
1	942.5
2	972.5
3	$1.025\cdot10^3$
4	665
5	$1.077\cdot10^3$

$n := \mathrm{rows}(X) \qquad n = 31$

Исключение грубых погрешностей:

Выборочное среднее, дисперсия и стандартное отклонение:

$$Mx := \frac{1}{n} \cdot \sum_{i=0}^{n-1} X_i \qquad s2 := \frac{1}{n-1} \cdot \sum_{i=0}^{n-1} \left(X_i - Mx\right)^2 \qquad s := \sqrt{s2}$$

$$Mx = 1.034 \times 10^3 \qquad s2 = 4.841 \times 10^4 \qquad s = 220.014$$

Графическое изображение элементов выборки, среднего и пределов «три» сигма:

$$i := 0..\, n-1 \qquad msig3 := Mx - 3 \cdot s \qquad psig3 := Mx + 3 \cdot s$$

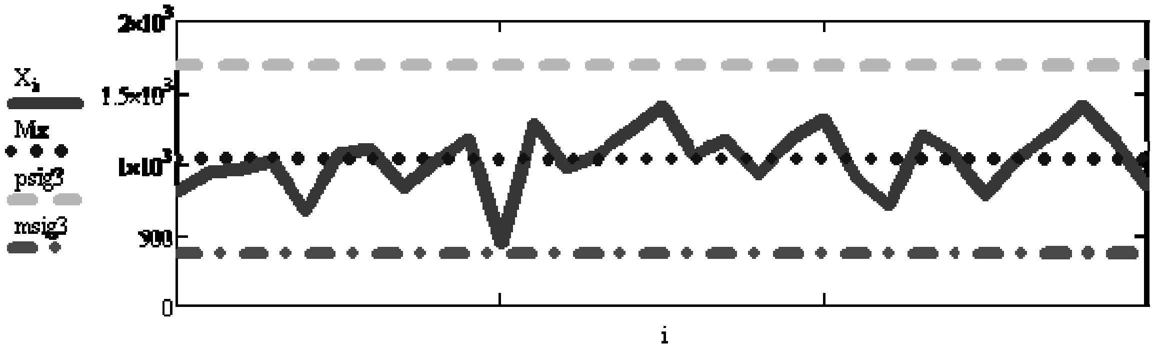


Рисунок 9 - Графичекое представление элементов выборки

Сортировка выборки и повторное графическое изображение:

$$X := \mathrm{sort}(X) \qquad Xmax := \max(X) \qquad Xmin := \min(X) \qquad |Xmax - Mx| = 381.081$$

$$Xmax = 1.415 \times 10^3 \qquad Xmin = 440 \qquad |Xmax - Mx| = 381.081$$

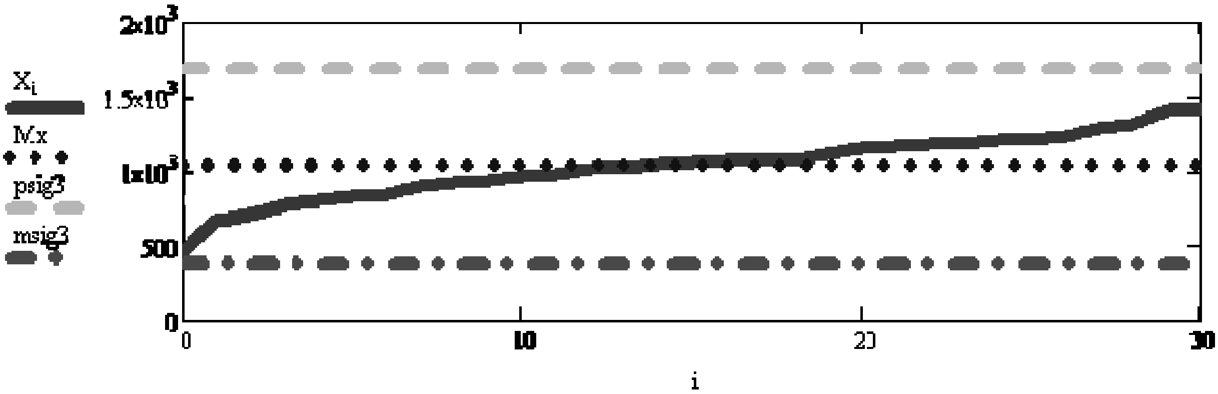


Рисунок 10 - Графичекое представление сортированной выборки

Проверка, является ли минимум выборки грубой ошибкой:

Максимальное относительное отклонение: $\tau := \frac{|Xmin - Mx|}{s} \qquad \tau = 2.699$

Уровень значимости: $\alpha := 0.05$

$$t := qt(1-\alpha, n-2) \quad t = 1.699 \qquad t1 := \frac{t \cdot \sqrt{n-1}}{\sqrt{n-2+t^2}} \qquad t1 = 1.648$$

Уровень значимости: $\alpha := 0.001$

$t := qt(1-\alpha, n-2)$ $\quad t = 3.396 \quad t2 := \dfrac{t\cdot\sqrt{n-1}}{\sqrt{n-2+t^2}} \quad t2 = 2.922$

Т.к. τ<τ2 минимальный элемент выборки не является грубой ошибкой, поэтому его удалять не нужно.

Проверка нормальности распределения выборки с использованием асимметрии и эксцесса:

$A := skew(X) \quad A = -0.54 \quad E := kurt(X) \quad E = 0.524$

$$DA := \frac{6\cdot(n-2)}{(n+1)\cdot(n+3)} \qquad DE := \frac{24\cdot n\cdot(n-2)\cdot(n-3)}{(n+1)^2\cdot(n+3)\cdot(n+5)}$$

$3\cdot\sqrt{DA} = 1.2 \qquad 5\cdot\sqrt{DE} = 3.471$

Поскольку выполняются неравенства |A|<3sqrt(DA) и |E|<5sqrt(DE) то гипотеза о нормальности распределения выборки не отклоняется.

Проверка нормальности распределения с использованием критерия согласия Пирсола: область значений случайной величины разбиваем на 10 интервалов и определяем количество элементов, попавших в каждый интервал, с вычислением вероятностей попадания значений в каждый интервал:

$m := 10 \qquad \Delta X := \dfrac{ceil(Xmax) - floor(Xmin)}{m} \qquad \Delta X = 97.5$

$i := 0..m+1$

$y_i := floor(Xmin) + \Delta X\cdot(i-1) \qquad X_0 := -\infty \qquad X_{m+1} := \infty$

$k := 0..m \qquad v := hist(y, X) \qquad p_k := pnorm(y_{k+1}, Mx, s) - pnorm(y_k, Mx, s)$

Для проверки определяются суммы: $\sum_{i=0}^{m} p_i = 0.958 \quad \sum_{i=0}^{m} v_i = 29 \quad pn := p\cdot n$

Вычисленные значения величин:

X =

	0
0	$-1\cdot10^{307}$
1	665
2	717.5
3	777.5
4	800
5	830
6	845
7	905
8	927.5
9	942.5
10	972.5

v =

	0
0	0
1	0
2	0
3	2
4	2
5	3
6	4
7	7
8	6
9	2
10	3

p =

	0
0	$2.635\cdot10^{-3}$
1	$8.553\cdot10^{-3}$
2	0.023
3	0.05
4	0.092
5	0.137
6	0.17
7	0.173
8	0.145
9	0.1
10	0.057

pn =

	0
0	0.082
1	0.265
2	0.709
3	1.564
4	2.841
5	4.256
6	5.255
7	5.349
8	4.488
9	3.104
10	1.77

Вычисляем сумму:

$$\chi 2 := \sum_{i=0}^{m} \frac{\left(v_i - pn_i\right)^2}{pn_i} \qquad \chi 2 = 4.365$$

Критическое значение критерия Хи-квадрат:

$\alpha := 0.05$ уровень значимости

$CR := qchisq(1 - \alpha, m - 3) \qquad CR = 14.067$

Т.к. вычисленное значение критерия Хи-квадрат меньше критического, то гипотеза о нормальности закона распределения принимается.

2.4 Лабораторная работа №4

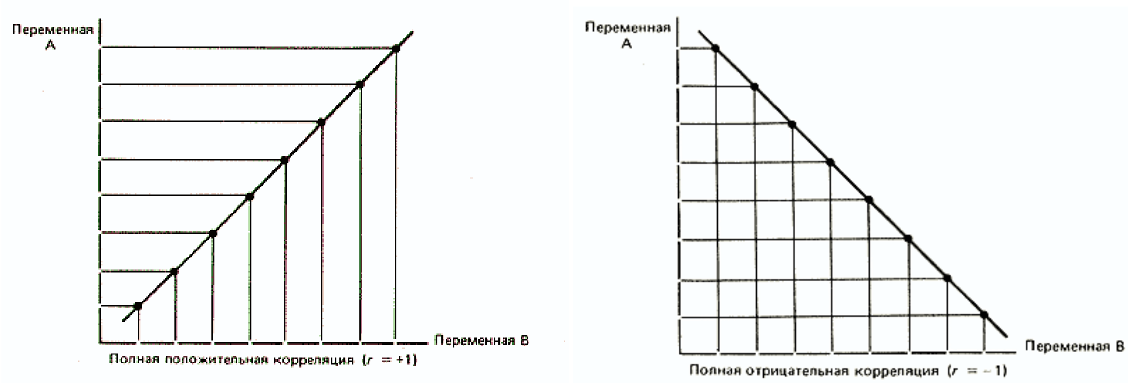
Тема работы: корреляционный анализ данных.

Цель работы: ознакомиться с методикой определения коэффициентов корреляции при помощи прямых вычислений и с использованием встроенных функций MathCAD.

Теоретическая часть

Корреляционная зависимость - статистическая взаимосвязь двух или нескольких случайных величин.

Коэффициент корреляции - это величина, которая может варьировать в пределах от +1 до -1. В случае полной положительной корреляции этот коэффициент равен плюс 1, а при полной отрицательной - минус 1.



а) б)

Рисунок 11 – Графическое представление положительной (а) и отрицательной (б) корреляции

Коэффициент корреляции определяется следующим образом:

$$\rho_{xy} = \frac{C_{xy}}{\sigma_x \times \sigma_y}, \qquad (22)$$

где $C_{xy} = \frac{1}{N-1} \times \sum_{i=0}^{N-1} \left((x_i - m_x) \times (y_{i+\tau} - m_y) \right)$ - функция взаимной корреляции; m_x, m_y - выборочные средние; σ_x^2, σ_y^2 - стандартные отклонения.

Задание: определить взаимную корреляцию между параметрами X1 и Y.

Пример выполнения лабораторной работы №4 в среде MathCAD

Считывание исходных данных из файлов на жестком диске

X1 :=

	0
0	800
1	942.5
2	972.5
3	$1.025\cdot10^3$
4	665
5	$1.077\cdot10^3$

Y :=

	0
0	772.5
1	847.5
2	945
3	$1.05\cdot10^3$
4	525
5	900

Определение взаимной корреляции между параметрами:

Максимальная величина задержки - $\tau max := 8$

Число данных в выборке- $n := rows(Y)$

Определим выборочные средние и стандартные отклонения:

$$\mu y := \frac{\sum_{i=0}^{n-1} Y_i}{n} \qquad \mu x := \frac{\sum_{i=0}^{n-1} X1_i}{n} \qquad oy := \sqrt{\frac{\sum_{i=0}^{n-1} \left(Y_i - \mu y\right)^2}{n}} \qquad ox := \sqrt{\frac{\sum_{i=0}^{n-1} \left(X1_i - \mu x\right)^2}{n}}$$

$$\rho(y, \mu y, x, \mu x, \tau max, n) := \left| \begin{array}{l} n \leftarrow n - 1 \\ \tau \leftarrow 0 \\ \text{for } i \in 0..\tau max \\ \quad \left| \begin{array}{l} Cyyr \leftarrow 0 \\ \text{for } j \in 0..n - \tau \\ \quad Cyyr \leftarrow Cyyr + \dfrac{\left[\left(x_j - \mu x\right)\cdot\left(y_{j+\tau} - \mu y\right)\right]}{n - \tau + 1} \\ \tau \leftarrow \tau + 1 \\ Cyy_i \leftarrow Cyyr \end{array} \right. \\ \rho \leftarrow Cyy \end{array} \right.$$

Корреляционная функция:

$$Cxy := \rho(Y, \mu y, X1, \mu x, \tau max, n) \qquad x := 0..\tau max - 1 \qquad Cxyn_x := \frac{Cxy_x}{ox\cdot oy}$$

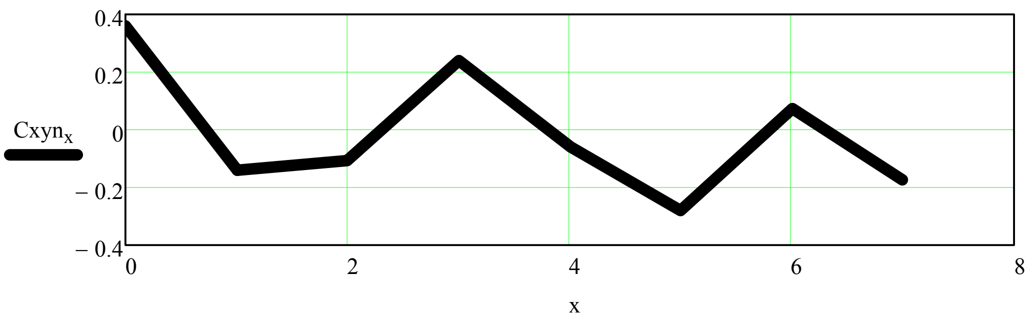


Рисунок 12 - График корреляционной функции

$$\text{Cxyn} = \begin{pmatrix} 0.362 \\ -0.14 \\ -0.106 \\ 0.243 \\ -0.061 \\ -0.277 \\ 0.076 \\ -0.176 \end{pmatrix}$$

Т.к. значение максимального коэффициента корреляции <0.7, то связь между факторами отсутствует.

Для проверки правильности разработанной программы определим максимальный коэффициент корреляции при помощи встроенной функции MathCAD.

$\text{Kmax} := \text{corr}(X1, Y)$

$\text{Kmax} = 0.362$

2.5 Лабораторная работа №5

Тема работы: парный регрессионный анализ.

Цель работы: ознакомиться с методикой проведения парного регрессионного анализа, способами определения коэффициентов регрессии.

Теоретическая часть

Парный регрессионный анализ имеет целью установить зависимость между двумя величинами в виде:

$$y(x) \equiv M(Y / X = x) = a_0 + a_1 x \qquad (23)$$

по результатам измерений

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n). \qquad (24)$$

Уравнение определяет прямую, которая является оценкой истинной линии регрессии:

$$\hat{y} = \hat{a}_0 + \hat{a}_1 x. \qquad (25)$$

Доверительная область для всей линии регрессии определяется с помощью уравнений

$$y'(x) = \hat{a}_0 + \hat{a}_1 x - s\sqrt{2 f_{\alpha,2,n-2}} \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}, \qquad (26)$$

$$y''(x) = \hat{a}_0 + \hat{a}_1 x + s\sqrt{2 f_{\alpha,2,n-2}} \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}. \qquad (27)$$

Описывает с вероятностью $p = 1 - \alpha$, что в данной области лежат точки выборки. В выражениях для определения доверительных границ $f_{\alpha,2,n-2}$ - квантиль распределения Фишера.

Проверка значимости уравнения регрессии выполняется оп критерию Фишера:

$$\frac{s_Y^2}{s^2} > f_{\alpha, n-1, n-2}\,. \qquad (28)$$

Задание: определить коэффициенты парной регрессионной модели, доверительную область уравнения регрессии и оценить значимость модели.

Пример выполнения лабораторной работы №5 в среде MathCAD

Считывание исходных данных из файлов на жестком диске

X2 :=

	0
0	800
1	942.5
2	972.5
3	$1.025 \cdot 10^3$
4	665
5	$1.077 \cdot 10^3$

Y :=

	0
0	257.5
1	282.5
2	315
3	350
4	175
5	300

Определение коэффициентов регрессионной модели с использованием встроенных функций MathCAD:

$a := \mathrm{intercept}(X2, Y) \quad b := \mathrm{slope}(X2, Y) \quad a = 274.59 \quad b = 0.052 \quad n := \mathrm{rows}(Y)$

$ab := \mathrm{line}(X2, Y)$

$$ab = \begin{pmatrix} 274.59 \\ 0.052 \end{pmatrix}$$

$a := ab_0 \quad b := ab_1 \quad a = 274.59 \quad b = 0.052$

Определение коэффициентов регрессионной модели при минимизации функции суммы квадратов отклонений точек от линии регрессии:

$$E(a, b) := \sum_{i=0}^{\mathrm{rows}(Y)-1} \left(a + b \cdot X2_i - Y_i\right)^2 \qquad \mathrm{Minimize}(E, a, b) = \begin{pmatrix} 274.59 \\ 0.052 \end{pmatrix}$$

Определение коэффициентов регрессии, посредством частных производных функции отклонения точек от линии регрессии:

$a := 0 \quad b := 0$ - начальное приближение

Given

$$\frac{d}{da} E(a, b) = 0 \qquad \frac{d}{db} E(a, b) = 0$$

$$\mathrm{Find}(a, b) = \begin{pmatrix} 274.59 \\ 0.052 \end{pmatrix}$$

Определим коэффициенты регрессии при помощи решения системы нормальных уравнений. Матрицы коэффициентов нормальных уравнений и вектор правых частей уравнений:

$$M := \begin{bmatrix} \mathrm{rows}(Y)-1 & \sum_{i=0}^{\mathrm{rows}(Y)-1} X2_i \\ \sum_{i=0}^{\mathrm{rows}(Y)-1} X2_i & \sum_{i=0}^{\mathrm{rows}(Y)-1} \left(X2_i\right)^2 \end{bmatrix} \qquad A := \begin{pmatrix} a \\ b \end{pmatrix} \qquad P := \begin{bmatrix} \sum_{i=0}^{\mathrm{rows}(Y)-1} Y_i \\ \sum_{i=0}^{\mathrm{rows}(Y)-1} \left(X2_i \cdot Y_i\right) \end{bmatrix}$$

$$M = \begin{pmatrix} 30 & 3.205\times 10^4 \\ 3.205\times 10^4 & 3.459\times 10^7 \end{pmatrix} \qquad P = \begin{pmatrix} 1.017\times 10^4 \\ 1.059\times 10^7 \end{pmatrix} \qquad P = \begin{pmatrix} 1.017\times 10^4 \\ 1.059\times 10^7 \end{pmatrix}$$

Обратная матрица и вектор коэффициентов:

$$M1 := M^{-1} \qquad A := M1 \cdot P \qquad A = \begin{pmatrix} 1.186\times 10^3 \\ -0.792 \end{pmatrix}$$

Определение коэффициентов регрессии с использованием явных формул:

$$a := \frac{\sum_{i=0}^{\mathrm{rows}(Y)-1} Y_i \cdot \sum_{i=0}^{\mathrm{rows}(Y)-1} \left(X2_i\right)^2 - \sum_{i=0}^{\mathrm{rows}(Y)-1} \left(Y_i \cdot X2_i\right) \cdot \sum_{i=0}^{\mathrm{rows}(Y)-1} X2_i}{n \cdot \sum_{i=0}^{\mathrm{rows}(Y)-1} \left(X2_i\right)^2 - \left(\sum_{i=0}^{\mathrm{rows}(Y)-1} X2_i\right)^2} \qquad a = 274.59$$

$$b := \frac{n \cdot \sum_{i=0}^{\mathrm{rows}(Y)-1} \left(X2_i \cdot Y_i\right) - \sum_{i=0}^{\mathrm{rows}(Y)-1} X2_i \cdot \sum_{i=0}^{\mathrm{rows}(Y)-1} Y_i}{n \cdot \sum_{i=0}^{\mathrm{rows}(Y)-1} \left(X2_i\right)^2 - \left(\sum_{i=0}^{\mathrm{rows}(Y)-1} X2_i\right)^2} \qquad b = 0.052$$

Определение доверительных интервалов для коэффициентов регрессии:

$$Mx := \frac{1}{n} \cdot \sum_{i=0}^{n-1} X2_i \qquad My := \frac{1}{n} \cdot \sum_{i=0}^{n-1} Y_i$$

$$sx := \sqrt{\frac{1}{n-1} \cdot \sum_{i=0}^{n-1} \left(X2_i - Mx\right)^2} \qquad sy := \sqrt{\frac{1}{n-1} \cdot \sum_{i=0}^{n-1} \left(Y_i - My\right)^2}$$

$$Mx = 1.034\times 10^3 \qquad My = 328.016 \qquad sx = 220.014 \qquad sy = 71.732$$

$$\alpha := 0.05 \qquad t := qt\left(1 - \frac{\alpha}{2}, n-2\right) \qquad t = 2.045$$

$$S2 := \frac{1}{n-2} \cdot \sum_{k=0}^{n-1} \left(Y_k - a - b \cdot X2_k\right)^2$$

$$S := \sqrt{S2} \qquad sx2 := (sx)^2 \qquad sy2 := (sy)^2 \qquad r := \mathrm{corr}(X2, Y)$$

$$aL := a - t \cdot S \cdot \sqrt{\frac{1}{n} + \frac{(Mx)^2}{\sum_{k=0}^{n-1} \left(X2_k - Mx\right)^2}} \qquad a - t \cdot \sqrt{\left[\frac{1}{n} + \frac{(Mx)^2}{(n-1) \cdot sx2}\right] \cdot \frac{n-1}{n-2} \cdot sy2 \cdot \left(1 - r^2\right)} = 145.444$$

$$aR := a + t\cdot S\cdot\sqrt{\frac{1}{n} + \frac{(Mx)^2}{\sum_{k=0}^{n-1}\left(X2_k - Mx\right)^2}} \qquad a + t\cdot\sqrt{\left[\frac{1}{n} + \frac{(Mx)^2}{(n-1)\cdot sx2}\right]\cdot\frac{n-1}{n-2}\cdot sy2\cdot\left(1-r^2\right)} = 403.735$$

Доверительный интервал для коэффициента а - $aL = 145.444 \quad aR = 403.735$

$\alpha := 0.05 \qquad t := qt\left(1 - \frac{\alpha}{2}, n-2\right) \qquad t = 2.045$

$$bL := b - \frac{t\cdot S}{\sqrt{\sum_{k=0}^{n-1}\left(X2_k - Mx\right)^2}} \qquad b - t\cdot\frac{sy\cdot\sqrt{\left(1-r^2\right)}}{sx\cdot\sqrt{(n-2)}} = -0.071$$

$$bR := b + \frac{t\cdot S}{\sqrt{\sum_{k=0}^{n-1}\left(X2_k - Mx\right)^2}} \qquad b + t\cdot\frac{sy\cdot\sqrt{\left(1-r^2\right)}}{sx\cdot\sqrt{(n-2)}} = 0.174$$

Доверительный интервал для коэффициента b - $bL = -0.071 \quad bR = 0.174$

Доверительная область для всей линии регрессии:

$f := qF(1-\alpha, 2, n-2) \qquad f = 3.328 \qquad i := 0..\,rows(Y) - 1$

$$yn_i := a + b\cdot X2_i - \sqrt{2\cdot f}\cdot S\cdot\sqrt{\frac{1}{n} + \frac{\left(X2_i - Mx\right)^2}{(n-1)\cdot sx2}} \qquad yv_i := a + b\cdot X2_i + \sqrt{2\cdot f}\cdot S\cdot\sqrt{\frac{1}{n} + \frac{\left(X2_i - Mx\right)^2}{(n-1)\cdot sx2}}$$

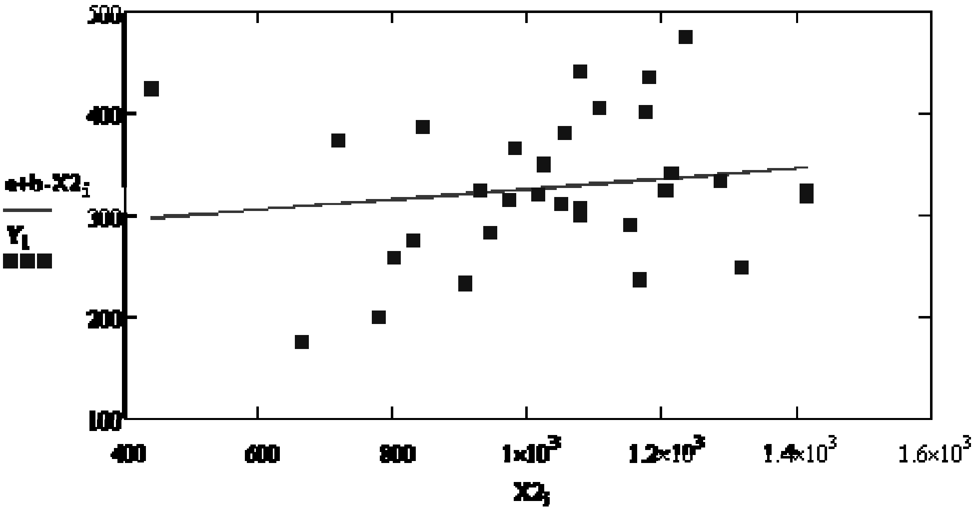


Рисунок 13 - Линия регрессии и реальные точки

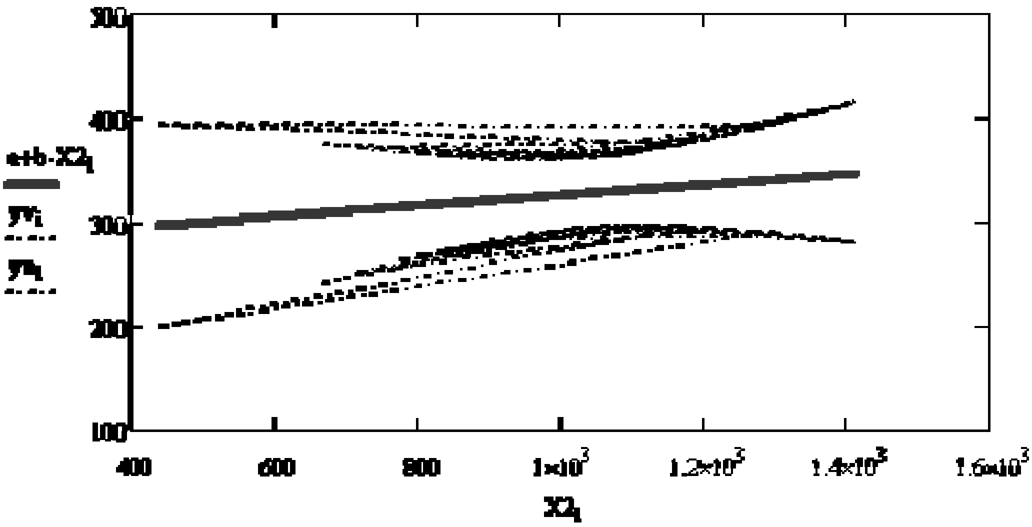


Рисунок 14 - Линия регрессии и доверительная область

Проверка адекватности уравнения регрессии:

$$sy2 := \frac{1}{n-1}\cdot\sum_{i=0}^{n-1}\left(Y_i - My\right)^2 \qquad s2 := \frac{1}{n-2}\cdot\sum_{i=0}^{n-1}\left(Y_i - a - b\cdot X2_i\right)^2 \qquad \frac{sy2}{s2} = 0.992$$

$$\alpha := 0.05 \qquad f := qF(1-\alpha, n-1, n-2) \qquad f = 1.854$$

Поскольку sy2/s2<f, то уравнение регрессии неадекватно отражает взаимосвязь между параметрами.

2.6 Лабораторная работа №6

Тема работы: множественный регрессионный анализ.

Цель работы: ознакомиться с методикой проведения множественного регрессионного анализа, способами определения коэффициентов регрессии .

Теоретическая часть

В случаях, когда функция отклика зависит от нескольких факторов, она ищется в виде:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k. \qquad (29)$$

При этом результаты наблюдений получаются из *n* опытов:

$$\begin{cases} y_1 = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k \\ y_2 = b_0 + b_1 x_{21} + b_2 x_{22} + \ldots + b_k x_{2k} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_n = b_0 + b_1 x_{n1} + b_2 x_{n2} + \ldots + b_k x_{nk}, \end{cases} \qquad (30)$$

Задачей множественного регрессионного анализа является построение такого уравнения прямой *k*-мерном пространстве, отклонение результатов наблюдений x_{ij} от которой были бы минимальными. Используя для этого метод наименьших квадратов, получаем систему нормальных уравнений:

$$\begin{cases} nb_0 + b_1\sum x_{i1} + b_2\sum x_{i2} + ... + b_k\sum x_{ik} = \sum y_i \\ b_0\sum x_{i1} + b\sum x_{i1}^2 + b_2\sum x_{i1}x_{i2} + ... + b_j\sum x_{i1}\sum x_{ij} + ... + b_k\sum x_{i1}x_{ik} = \sum y_i x_{i1} \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ b_0\sum x_{ij} + b_1\sum x_{i1}x_{ij} + ... + b_j\sum x_{ij}^2 + ... + b_k\sum x_{ik}x_{ij} = \sum y_i x_{ij} \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ b_0\sum x_{ik} + b_1\sum x_{i1}x_{ik} + ... + b_j\sum x_{ij}x_{ik} + ... + b_k\sum x_{ik}^2 = \sum y_i x_{ik}, \end{cases} \quad (31)$$

или в матричном виде:

$$X^T X = \begin{pmatrix} n & \sum x_{i1} & \sum x_{i2} & ... & \sum x_{ik} \\ \sum x_{i1} & \sum x_{i1}^2 & \sum x_{i2}x_{i1} & ... & \sum x_{i1}x_{ik} \\ \sum x_{i2} & \sum x_{i2}x_{i1} & \sum x_{i1} & ... & \sum x_{i2}x_{ik} \\ ... & ... & ... & ... & ... \\ \sum x_{ik} & \sum x_{ik}x_{i1} & \sum x_{ik}x_{i2} & ... & \sum x_{ik}^2 \end{pmatrix}. \quad (32)$$

Степень влияния коэффициентов оценивается посредством коэффициентов эластичности:

$$Э_i = \frac{bx_i}{a + bx_i} \quad (33)$$

Чем ближе значение коэффициента эластичности к 1 по абсолютной величине, тем сильнее влияние того или иного коэффициента регрессии на функцию отклика. Адекватность уравнения множественной регрессии оценивается аналогично парной регрессии (с вычислением дисперсии адекватности и критерия Фишера).

Задание: определить коэффициенты парной регрессионной модели, доверительную область уравнения регрессии и оценить значимость модели.

Пример выполнения лабораторной работы №6 в среде MathCAD

Считывание исходных данных из файлов на жестком диске

X1 :=

	0
0	800
1	942.5
2	972.5
3	$1.025\cdot10^3$
4	665
5	$1.077\cdot10^3$

X2 :=

	0
0	772.5
1	847.5
2	945
3	$1.05\cdot10^3$
4	525
5	900

Y :=

	0
0	257.5
1	282.5
2	315
3	350
4	175
5	300

X1 и X2 - объясняющие переменные

$i := 0..\mathrm{rows}(Y) - 1 \qquad X_{i,0} := 1 \qquad X_{i,2} := X2_i \qquad n := \mathrm{rows}(Y)$

$$D := X^T \cdot X \qquad D = \begin{pmatrix} 31 & 3.205\times 10^4 & 3.01\times 10^4 \\ 3.205\times 10^4 & 3.459\times 10^7 & 3.169\times 10^7 \\ 3.01\times 10^4 & 3.169\times 10^7 & 3.092\times 10^7 \end{pmatrix}$$

$$d := X^T \cdot Y \qquad d = \begin{pmatrix} 1.017\times 10^4 \\ 1.059\times 10^7 \\ 1.013\times 10^7 \end{pmatrix}$$

$$\text{rank}(D) = 3 \qquad \text{conde}(D) = 6.375\times 10^7$$

$$b := D^{-1} \cdot d = \begin{pmatrix} 187.888 \\ -7.972\times 10^{-3} \\ 0.153 \end{pmatrix}$$

$$b := \text{lsolve}(D, d) = \begin{pmatrix} 187.888 \\ -7.972\times 10^{-3} \\ 0.153 \end{pmatrix}$$

Определение коэффициентов эластичности модели:

$$i := 0..\, \text{rows}(Y) - 1 \qquad X3_{i,0} := X1_i \qquad X3_{i,1} := X2_i \qquad X := X3$$

$$b := \begin{pmatrix} 16.99 \\ 0.04 \\ 0.412 \end{pmatrix}$$

$$xcp1 := \text{mean}\left(X^{\langle 0\rangle}\right) \qquad xcp2 := \text{mean}\left(X^{\langle 1\rangle}\right) \qquad ycp := \text{mean}(Y)$$

$$s1 := \sqrt{\text{var}\left(X^{\langle 0\rangle}\right)} \qquad s2 := \sqrt{\text{var}\left(X^{\langle 1\rangle}\right)} \qquad sy := \sqrt{\text{var}(Y)}$$

$$bst1 := b_1 \cdot \frac{s1}{sy} \qquad bst2 := b_2 \cdot \frac{s2}{sy} \qquad bst1 = 0.123 \qquad bst2 = 1.362$$

$$E1 := b_1 \cdot \frac{xcp1}{ycp} \qquad E2 := b_2 \cdot \frac{xcp2}{ycp} \qquad E1 = 0.126 \qquad E2 = 1.22$$

Построение множественной регрессии средствами MathCad:

$$coef := \text{regress}(X, Y, 1)$$

$$yc(z) := \text{interp}(coef, X, Y, z)$$

$$i := 0..\, \text{rows}(Y) - 1$$

$$yy_i := yc\left[\left(X^T\right)^{\langle i\rangle}\right]$$

$$coef = \begin{pmatrix} 3 \\ 3 \\ 1 \\ -7.972\times 10^{-3} \\ 0.153 \\ 187.888 \end{pmatrix}$$

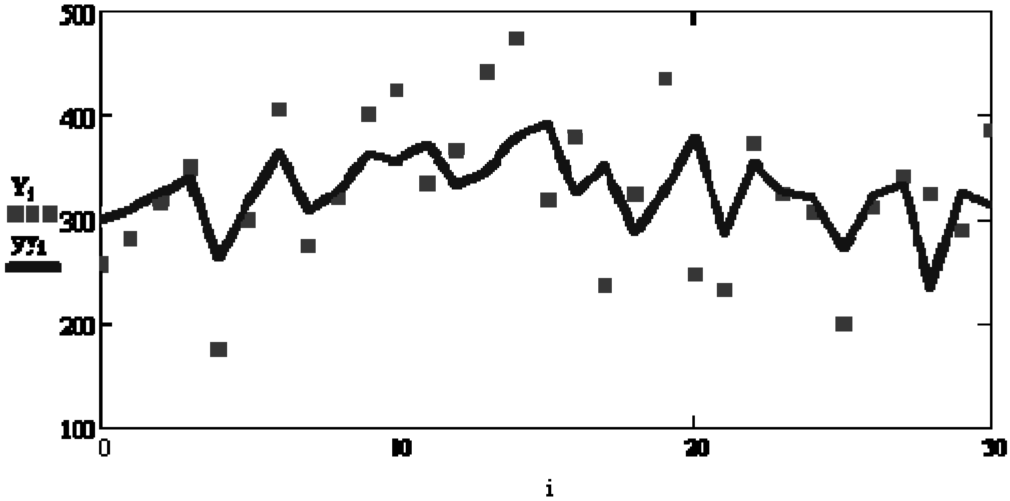
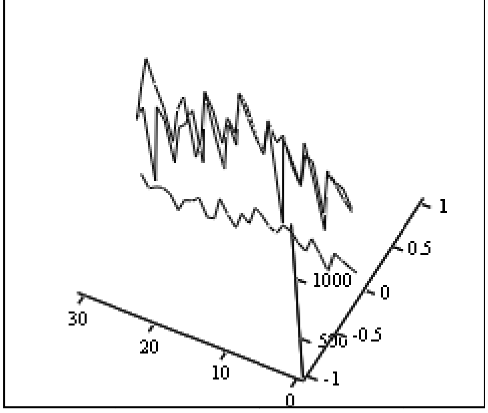


Рисунок 15 - График модели



X1, X2, Y

Рисунок 16 - Исходные данные

Определение интервальных оценок уравнения регрессии:

$$s2 := \frac{\sum_{i=0}^{rows(Y)-3} \left(Y_i - yy_i\right)^2}{n-3} \qquad s2 = 3.91\times 10^3$$

$$j := 0..2 \qquad s2b_j := s2\cdot\left(D^{-1}\right)_{j,j} \qquad s2b = \begin{pmatrix} 3.864\times 10^3 \\ 3.099\times 10^{-3} \\ 2.669\times 10^{-3} \end{pmatrix}$$

$$\delta b_j := qt\left(1 - \frac{0.05}{2}, n-3\right)\cdot\sqrt{s2b_j}$$

$b_j + \delta b_j =$

144.317
0.154
0.518

$$b = \begin{pmatrix} 16.99 \\ 0.04 \\ 0.412 \end{pmatrix}$$

$b_j - \delta b_j =$

-110.337
-0.074
0.306

$\chi0025 := \text{qchisq}(0.025, n-3) = 15.308$

$\chi0975 := \text{qchisq}(0.975, n-3) = 44.461$

$\frac{n \cdot s2}{\chi0975} = 2.726 \times 10^3 \qquad \frac{n \cdot s2}{\chi0025} = 7.918 \times 10^3$

Определение интервальной оценки математического ожидания:

$i := 0..n-1$

$z_i := \begin{pmatrix} 1 \\ X_{i,0} \\ 7 \end{pmatrix} \qquad yyc_i := yc\left(\begin{pmatrix} X_{i,0} \\ 7 \end{pmatrix}\right) \qquad s := \sqrt{s2}$

$\delta y_i := qt\left(1 - \frac{0.05}{2}, n-3\right) \cdot s \cdot \sqrt{z_i^T \cdot D^{-1} \cdot z_i}$

$yyH_i := yyc_i - \delta y_i \qquad yyB_i := yyc_i + \delta y_i$

$\delta 1y_i := qt\left(1 - \frac{0.05}{2}, n-3\right) \cdot \sqrt{s2} \cdot \sqrt{1 + z_i^T \cdot D^{-1} \cdot z_i}$

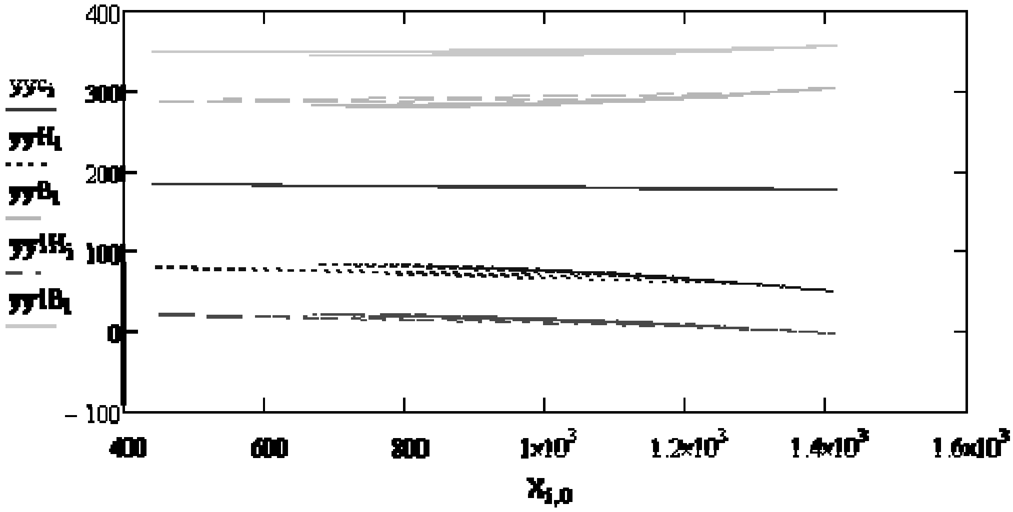


Рисунок 17 - Доверительные интервалы уравнения регрессии

Определение значимости коэффициентов модели:

$T_j := \frac{b_j}{s2b_j} \qquad T = \begin{pmatrix} 4.397 \times 10^{-3} \\ 12.907 \\ 154.376 \end{pmatrix}$

$t := qt(1 - 0.05, 7) \qquad t = 1.895$

Т.к. Tj<t, все коэффициенты модели не являются значимыми.

$i := 0..\text{rows}(Y) - 1 \qquad X_{i,0} := 1 \qquad X_{i,1} := X1_i \qquad X_{i,2} := X2_i$

$R2 := \frac{b^T \cdot X^T \cdot Y - n \cdot ycp^2}{Y \cdot Y - n \cdot ycp^2} \qquad R2 = 9.287$

$R2ck := 1 - \frac{n-1}{n-3} \cdot (1 - R2) \qquad R2ck = 9.879$

$F := \frac{R2 \cdot (n-3)}{(1 - R2) \cdot 2} \qquad F = -15.689$

$\text{Ftab} := \text{qF}(0.95, 2, 7) \qquad \text{Ftab} = 4.737$

Т.к. расчетное значение F-критерия меньше табличного, то можно сделать вывод о том, что полученная модель не является адекватной.

2.7 Лабораторная работа №7

Тема работы: нелинейный регрессионный анализ.

Цель работы: ознакомиться с методикой проведения нелинейного регрессионного анализа, способами определения коэффициентов регрессии и способом определения оптимальной степени обобщающего полинома.

Теоретическая часть

Если между факторами существуют нелинейные соотношения, то они выражаются с помощью соответствующих нелинейных функций.

Различают два класса нелинейных регрессий:

- регрессии, нелинейные относительно включенных в анализ объясняющих переменных (аргументов), но линейные по оцениваемым параметрам;

- регрессии, нелинейные по оцениваемым параметрам.

В данной работе будет рассматриваться первый класс моделей. Данный класс нелинейных регрессий включает уравнения, в которых зависимая переменная линейно связана с параметрами. Примером могут служить:

- полиномы разных степеней

$y_i = a_0 + a_1x_1 + a_2x_2^{\,2} + ... + a_kx_k^{\,k} + \varepsilon_i$. (полином k-й степени) (34)

- и равносторонняя гипербола

$y_i = a + \dfrac{b}{x_i} + \varepsilon_i$. (35)

При оценке параметров регрессий, нелинейных по объясняющим переменным, используется подход, именуемый «замена переменных». Суть его состоит в замене «нелинейных» объясняющих переменных новыми «линейными» переменными и сведение нелинейной регрессии к линейной регрессии. К новой «преобразованной» регрессии может быть применен обычный метод наименьших квадратов.

Среди нелинейной полиноминальной регрессии чаще всего используется парабола второй степени; в отдельных случаях - полином третьего порядка. Ограничение в использовании полиномов более высоких степеней связаны с требованием однородности исследуемой совокупности: чем выше порядок полинома, тем больше изгибов имеет кривая и, соответственно, менее однородна совокупность по результативному признаку. Обычно процедура подбора степени полинома заключается в последовательном повышении степени обобщающего полинома, начиная с 2, как только изменение погрешности модели становится несущественным, процесс прекращают.

Задание: определить коэффициенты полиномиальной модели и оценить оптимальную степень обобщающего полинома.

Пример выполнения лабораторной работы №7 в среде MathCAD

Считывание исходных данных из файлов на жестком диске

$X :=$

	0
0	800
1	942.5
2	972.5
3	$1.025 \cdot 10^3$
4	665
5	$1.077 \cdot 10^3$

$Y :=$

	0
0	257.5
1	282.5
2	315
3	350
4	175
5	300

Количество точек в файле $n := \mathrm{rows}(X)$ $n = 31$ $\sigma := 300$

Задание полиномов Чебышева нулевого Q0 и первого порядков Q1:

$Q0(X) := 1$

$Mx := \frac{1}{n} \cdot \sum_{i=0}^{n-1} X_i$ $Mx = 1.034 \times 10^3$

$Q1(X) := X - Mx$

Аппроксимация экспериментальных данных с помощью полиномов Чебышева:

$H0 := n$ $H1 := \sum_{i=0}^{n-1} Q1(X_i)^2$ $a0 := \frac{1}{H0} \cdot \sum_{i=0}^{n-1} \left(Y_i \cdot Q0(X_i)\right)$ $a1 := \frac{1}{H1} \cdot \sum_{i=0}^{n-1} \left(Y_i \cdot Q1(X_i)\right)$

$H0 = 31$ $H1 = 1.452 \times 10^6$ $a0 = 328.016$ $a1 = 0.052$

$f1(X) := a0 \cdot Q0(X) + a1 \cdot Q1(X)$ $\mathrm{intercept}(X, Y) = 274.59$ $\mathrm{slope}(X, Y) = 0.052$

$i := 0 .. n - 1$

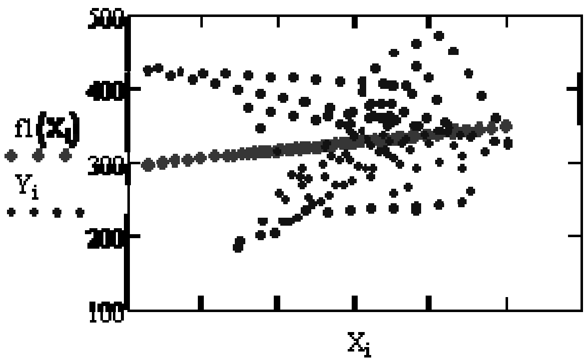


Рисунок 18 - Результаты построения модели

Проверка необходимости корректировки полученной формулы:

$\alpha := 0.05$

$qf := qF(1 - \alpha, n - 2, 1000)$ $qf = 1.479$

$$D1 := \frac{1}{n-1} \cdot \sum_{i=0}^{n-1} \left(Y_i - f1\left(X_i\right)\right)^2 \qquad d := \frac{D1}{\sigma^2} \qquad d = 0.056$$

Т.к. d<qf, полученная формула не нуждается в уточнении.

2.8 Лабораторная работа №8

Тема работы: моделирование искусственных нейронных сетей.

Цель работы: ознакомиться с методикой моделирования искусственных нейронных сетей.

Теоретическая часть

Искусственной нейронной сетью (ИНС) называется вычислительная структура, моделирующая способ работы человеческого мозга, состоящая из набора нейронов, соединенных между собой синаптическими связями.

Нейрон представляет собой единицу обработки информации в нейронной сети.

Синапсы или связи, каждая из которых характеризуется своим весом. В частности, сигнал x_j на входе синапса j, связанного с нейроном k, умножается на вес ω_{kj}. Синаптические веса искусственных нейронов могут иметь как положительные, так и отрицательные значения.

Функция активации ограничивает амплитуду выходного сигнала нейрона. Обычно нормализованный диапазон амплитуд выхода нейрона лежит в пределе [0,1] или [-1,1]. Математически функционирование искусственного нейрона можно представить следующим образом:

$$u_k = \sum_{j=1}^{m} \omega_{kj} \times x_j, \qquad (36)$$

$$y_k = \varphi(u_k + b_k), \qquad (37)$$

где x_m, y_k - входные и выходные сигналы;

ω_{km} - синаптические веса нейрона k;

u_k - линейная комбинация входных воздействий;

φ - функция активации.

Одним из самых важных свойств ИНС является их способность к обучению. Обучение – процесс, в котором свободные параметры ИНС настраиваются посредством моделирования среды, в которую эта ИНС встроена. Наиболее широко сегодня применяется метод обучения ИНС, основанный на коррекции ошибок.

Процесс обучения ИНС посредством коррекции ошибок можно представить следующей схемой:

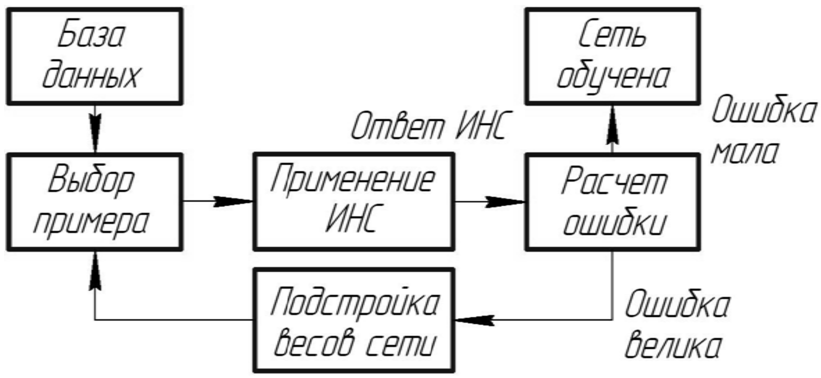


Рисунок 19 - Обучение ИНС, основанное на коррекции ошибок

Обучение производится по следующей схеме: на ИНС поступает входной сигнал $x(n)$, далее формируется ответ ИНС в виде выходного сигнала $y(n)$, который в свою очередь сравнивается с желаемым выходным сигналом $d(n)$. В результате получаем сигнал ошибки:

$$e(n) = d(n) - y(n), \tag{38}$$

где n - дискретное время или номер шага итеративного процесса настройки синапсов ИНС.

После формирования сигнала ошибки инициализируется механизм управления, цель которого заключается в применении последовательности корректировок синаптических весов нейронов ИНС. Вычисления повторяются до тех пор, пока ошибка не станет меньше требуемой.

Задание: произвести моделирование обучения искусственной нейронной сети с m нейронами (m – номер варианта).

Пример выполнения лабораторной работы №8 в среде MathCAD

Формируем исходные данные:

$ORIGIN := 1$

$m := 10$

$i := 1 .. m$

$j := 1 .. m$

$p_i := rnd(1)$

$t_i := rnd(1)$

$w_{i,j} := 0$

$p =$

	1
1	$1.268\cdot10^{-3}$
2	0.193
3	0.585
4	0.35
5	0.823
6	0.174

$t =$

	1
1	0.989
2	0.119
3	$8.923\cdot10^{-3}$
4	0.532
5	0.602
6	0.166

Настройка весов нейронной сети:

Giver

$$\sum_{i=1}^{m}\left(t_i - \frac{1}{1+\exp\left(-w^{\langle i\rangle T}\cdot p\right)}\right)^2 = 0$$

$w := \mathrm{Minerr}(w)$

$w =$

	1	2	3	4	5
1	13.956	-11.817	-1.496	9.999	3.015
2	0.092	-0.078	$-9.815\cdot10^{-3}$	0.066	0.02
3	0.03	-0.026	$-3.244\cdot10^{-3}$	0.022	$6.539\cdot10^{-3}$
4	0.051	-0.043	$-5.417\cdot10^{-3}$	0.036	0.011
5	3.931	-2.265	-5.085	0.015	0.46
6	0.102	-0.086	-0.011	0.073	0.022
7	0.025	-0.021	$-2.671\cdot10^{-3}$	0.018	$5.384\cdot10^{-3}$
8	0.058	-0.049	$-6.242\cdot10^{-3}$	0.042	0.013
9	0.194	-0.164	-0.021	0.139	0.042
10	0.12	-0.102	-0.013	0.086	...

$$T := \left|\begin{array}{l} \text{for } i \in 1..m \\ \quad \left|\begin{array}{l} \Sigma_i \leftarrow w^{\langle i\rangle T}\cdot p \\ a_i \leftarrow \dfrac{1}{1+\exp\left(-\Sigma_i\right)} \\ i \end{array}\right. \\ T \leftarrow a \\ T \end{array}\right.$$

$T =$

	1
1	0.968
2	0.119
3	0.015
4	0.532
5	0.602
6	0.166
7	0.451
8	0.064
9	0.783
10	0.52

Определяем величину ошибки обучения:

$\Delta_i := t_i - T_i$

$\max(\Delta) = 0.021$

$\delta_i := \frac{\Delta_i}{\max(t)}$

$\delta =$

	1
1	0.021
2	$-2.601\cdot10^{-4}$
3	$-5.9\cdot10^{-3}$
4	$-1.729\cdot10^{-10}$
5	$7.268\cdot10^{-11}$
6	$-4.428\cdot10^{-6}$
7	$2.117\cdot10^{-10}$
8	$-6.911\cdot10^{-3}$
9	$2.257\cdot10^{-10}$
10	$2.578\cdot10^{-4}$

СПИСОК ЛИТЕРАТУРЫ

1 Воскобойников, Ю. Е. Построение регрессионных моделей в пакете MathCad [текст] : учебное пособие / Ю. Е. Воскобойников. – Новосибирск : НГАСУ, 2009. - 220 с.

2 Корн, Г. Справочник по математике для научных работников и инженеров [текст] / Г. Корн, Т.Корн. - М. : Наука, 1973. - 831 с.

3 Симонов, А. М. Основы обеспечения качества поверхности деталей машин с использованием динамического мониторинга [Текст] / А. М. Симонов, А.К. Остапчук, В.Е.Овсянников. - Курган : Изд-во Курганского гос. ун-та, 2010. - 118 с.

4 Яньков, В.Ю. Лабораторный практикум по Маткаду. Модуль 3. Моделирование в Маткаде. Для преподавателей, аспирантов и студентов технических, технологических и экономических специальностей всех форм обучения [текст] / В.Ю. Яньков.- М. : МГУТУ, 2009. – 68 с.

5. http://portal.tpu.ru/SHARED/v/VNDEMIDOV/public.

Овсянников Виктор Евгеньевич

Инженерное и управленческое моделирование
в компьютерной системе

Лабораторный практикум для студентов, обучающихся по направлениям 221700.62 «Стандартизация и метрология» и 222000.62 «Инноватика»

Редактор А.С. Мокина

Подписано в печать	Формат 60×84 1/16	Бумага тип. №1
Печать трафаретная	Усл. печ. л. 2,5	Уч.-изд. л. 2,5
Заказ	Тираж 22	Цена свободная

Редакционно-издательский центр КГУ.
640669, г. Курган, ул. Гоголя, 25.
Курганский государственный университет.