

Министерство образования и науки Российской Федерации
Федеральное агентство по образованию
Владимирский государственный университет
Кафедра информационных систем и информационного менеджмента

ПРЕДСТАВЛЕНИЕ ЗНАНИЙ В ИНФОРМАЦИОННЫХ СИСТЕМАХ

Методические указания к лабораторным работам

Составители
Д.В. АЛЕКСАНДРОВ
А.Ю. ЛАРЮШКИН

Владимир 2005

Владимирский государственный университет

ПРЕДСТАВЛЕНИЕ ЗНАНИЙ В ИНФОРМАЦИОННЫХ СИСТЕМАХ

**Методические указания
к лабораторным работам**

Владимир 2005

УДК 681.518 (076)

ББК 32.988-5 я7

А46

Рецензент

Доктор технических наук, профессор
Владимирского государственного университета

А.В. Костров

Печатается по решению редакционно-издательского совета
Владимирского государственного университета

Представление знаний в информационных системах : метод.
А46 указания к лаб. работам / сост.: Д. В. Александров, А. Ю. Ларюшкин ; Владим. гос. ун-т. – Владимир : Ред.-издат. комплекс ВлГУ, 2005. – 20 с.

Представлены основные понятия теории вероятностей и математической статистики, основные законы распределения вероятностей случайных величин, задачи факторного анализа и линейного регрессионного анализа, а также методики выделения тренда из временного ряда и прогнозирования временных рядов.

Предназначены для студентов специальности 071900 – информационные системы в технике и технологиях при выполнении лабораторных занятий в системе *MathCad* по курсу «Представление знаний в информационных системах». Рекомендуются преподавателям при проведении лабораторных занятий. Могут быть полезны студентам, обучающимся по специальностям – системотехника, информатика, экономика и менеджмент всех форм обучения.

Ил. 2. Библиогр.: 5 назв.

УДК 681.518 (076)

ББК 32.988-5 я7

Лабораторная работа № 1

**ВВЕДЕНИЕ В ТЕОРИЮ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКУЮ
СТАТИСТИКУ. ЗНАКОМСТВО С СИСТЕМОЙ *MATHCAD***

1. Цель работы – знакомство с основными понятиями теории вероятностей (ТВ) и статистики, со средой *MathCad*.

2. Общие сведения

Теория вероятности занимается изучением закономерностей, которые порождаются случайными событиями.

Методы математической статистики позволяют оценивать параметры имеющих закономерностей.

Опыт (испытание) – осуществление определенного комплекса условий, который может быть воспроизведен сколько угодно раз.

Явления, происходящие в результате события, обычно обозначают: *A, B, C*.

Случайные величины обозначают: *X, Y, Z*.

События можно разделить:

- на достоверные;
- невозможные;
- случайные.

Вероятность – числовая мера правдоподобия события, возможного при случайном испытании.

События можно разделить:

- на зависимые;
- независимые.

Частота события A в *N* повторениях опыта есть величина $N(A)/N \approx P(A)$, если *N* велико.

Случайная величина – величина, которая в результате испытания принимает одно и только одно значение, заранее нам неизвестное. Случайные величины могут быть дискретными и непрерывными. Каждая случайная величина задает распределение вероятностей на множестве своих значений.

Закон распределения случайной величины – задание возможных значений и соответствующих им вероятностей.

Мода случайной величины – значение, имеющее максимальную вероятность.

Пусть ξ обозначает случайную величину, принимающую вещественное значение.

Функцией распределения $F(x)$ случайной величины ξ называют $F(x) = P(\xi \leq x)$, где x – вещественное число.

Случайную величину, принимающую вещественные значения, называют непрерывной, если непрерывна ее функция распределения (непрерывно и соответствующее распределение вероятностей).

Функция $P(t)$ называется *плотностью вероятности* в точке t , если для любых чисел a и b (пусть $a < b$) $P(a < \xi < b) = \int_b^a P(x)dx$.

Первый момент случайной величины ξ со значениями x_1, x_2, \dots , имеющими вероятности P_1, P_2, \dots , – это математическое ожидание $M(\xi) = \sum_k x_k P_k$.

Для непрерывной случайной величины ξ с плотностью $P(x)$ $M(\xi) = \int_{-\infty}^{\infty} xP(x)dx$, причем интеграл должен сходиться абсолютно.

Дисперсией D случайной величины ξ называется величина $D(\xi) = M(\xi^2) - (M(\xi))^2$. Существует не для всех случайных величин (не для всех распределений вероятностей), как и математическое ожидание.

Для дискретной случайной величины ξ со значениями x_1, x_2, \dots , имеющими вероятности P_1, P_2, \dots k -м моментом $M\xi^k$ называется величина $M\xi^k = \sum_i x_i P_i$, а k -м центральным моментом называется величина

$$\sum_i (x_i - M\xi)^k P_i.$$

Для непрерывной случайной величины с плотностью вероятности $P(x)$, k -м моментом называется величина $\int_{-\infty}^{\infty} x^k P(x) dx$, а k -м центральным моментом называется величина $M(\xi - M\xi)^k = \int_{-\infty}^{\infty} (x - M\xi)^k P(x) dx$.

Суммы и интегралы должны сходиться абсолютно. Моменты также существуют не для всех случайных величин.

Квантили используются как характеристики случайных величин, принимающих вещественные значения.

Квантилью x_p случайной величины, имеющей функцию распределения $F(x)$, называется решение x_p уравнения $F(x) = P$. Величину x_p называют P -квантилью или квантилью P распределения $F(x)$.

Уравнение $F(x) = p$, определяющее его P -квантили может не иметь решений или иметь несколько, следовательно, некоторые P -квантили не существуют или определены неоднозначно.

Случайные величины ξ и η независимы, если независимы любые два события, которые выражаются по отдельности через ξ и η .

Для зависимых величин важно знать степень зависимости: ковариацию и корреляцию.

Ковариацией $\text{cov}(\xi, \eta)$ случайных величин ξ и η называют $\text{cov}(\xi, \eta) = M(\xi - M\xi)(\eta - M\eta)$, если указанные математические ожидания существуют.

Коэффициент корреляции случайных величин ξ и η (обозначают ρ или $\text{corr}(\xi, \eta)$) называют $\rho = \frac{\text{cov}(\xi, \eta)}{\sqrt{D\xi}\sqrt{D\eta}}$.

Выбор одного объекта называют чисто случайным, если все объекты имеют равные вероятности оказаться выбранными.

Выбор n объектов из N называют чисто случайным, если все наборы из n объектов имеют одинаковые вероятности быть выбранными.

Выборка – последовательность независимых, одинаково распределенных случайных величин.

3. Задание к лабораторной работе

Познакомиться с основными понятиями теории вероятности и статистики. Изучить основы работы в системе *MathCad*.

4. Порядок выполнения работы

4.1. Основные операции в среде *MathCad*:

- работа с встроенной помощью системы *MathCad*;
- задание переменных, операции с переменными;
- задание массивов, матриц;
- операции с массивами, матрицами;
- вычисление значений функций;
- построение графика функции;
- импортирование данных из файла.

4.2. Оформить отчет по выполненной работе.

5. Содержание отчета

5.1. Назначение и цель работы.

5.2. Результаты вычислений в системе *MathCad*.

5.3. Выводы по результатам выполненной работы.

Лабораторная работа № 2

ЗАКОНЫ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ

1. Цель работы – знакомство с основными законами и понятиями распределения вероятностей случайных величин. Закрепление навыков работы с системой *MathCad*.

2. Общие сведения

Зная закон распределения вероятностей случайных величин (СВ), можно делать выводы о событиях, в которых участвуют эти СВ.

Биномиальное распределение (дискретное) возникает в тех случаях, когда нас интересует, сколько раз происходит некоторое событие в серии из определенного числа независимых наблюдений (опытов), выполняемых в одинаковых условиях.

Схема испытаний Бернулли – последовательность независимых испытаний, в которых результатом каждого из испытаний может быть один

из двух исходов (например, "успех" или "неудача"), и вероятность "успеха" (или "неудачи") в каждом из испытаний одна и та же.

Случайная величина X имеет биномиальное распределение с параметрами n и p , если она принимает значения $0, 1, \dots, n$ с вероятностями:

$$P(X = K) = C_n^k p^k (1 - p)^{n-k}, \quad k = 0, 1, \dots, n,$$

где n – число испытаний, p – вероятность успеха в испытании.

Событие $(X = K)$ – количество успешных испытаний.

$$C_n^k = \frac{n!}{k!(n-k)!}; \quad MX = np; \quad DX = np(1-p).$$

Выборочной (эмпирической) функцией распределения СВ ξ , построенной по выборке x_1, x_2, \dots, x_n , называется функция $F_n(x)$, равная доле таких значений x_i , что $x_i \leq x, i = 1, \dots, n$. Иначе говоря, $F_n(x)$ есть частота события $x_i \leq x$ в ряду x_1, x_2, \dots, x_n . Для построения выборочной функции распределения удобно от выборки x_1, \dots, x_n перейти к вариационному ряду $x_{(1)}, \dots, x_{(n)}$.

Вариационным рядом называют выборку, перенумерованную в порядке возрастания.

Среднее значение выборки или выборочный аналог математического ожидания:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Дисперсия выборки или выборочный аналог дисперсии:

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad \text{или} \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \quad \text{т.к.}$$
$$MS^2 = D\xi.$$

Выборочная квантиль – решение уравнения $F_n(x) = 0,5$.

Гистограмма выборки – графическое изображение зависимости частоты попадания элементов выборки от соответствующего интервала группировки (см. рисунок).

3. Задание к лабораторной работе

Для выборки диаметра головок заклепок (файл *D200gzmm.prn*) построить гистограмму частот с шагом группировки $0,075$ мк на интервале от 13 до $13,75$. Построить график соответствующего точного распределения.

4. Порядок выполнения работы

4.1. Получить у преподавателя исходные данные (файл *D200gzmm.prn*).

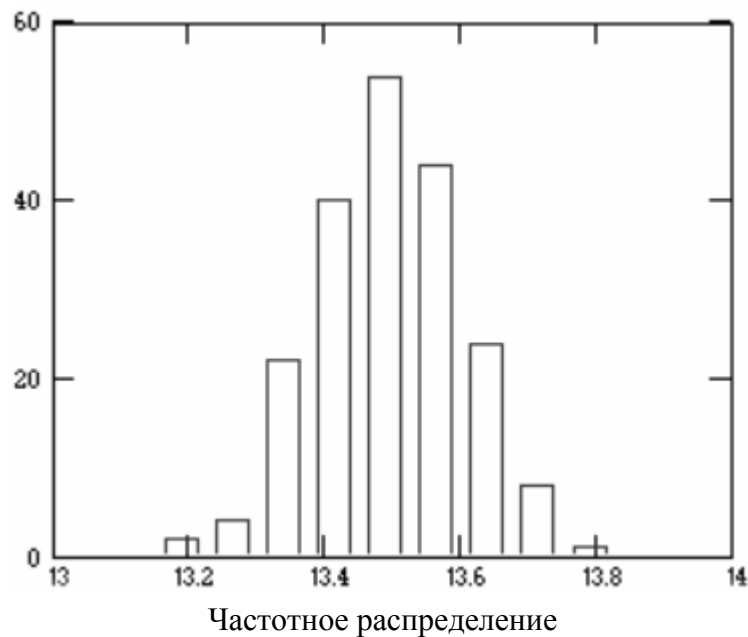
4.2. Определить число сегментов для построения гистограммы частотного распределения.

4.3. Построить график частотного распределения (используя функцию *hist*).

4.4. Определить значения коэффициентов, необходимых для построения графика точного распределения (математическое ожидание, дисперсия, шаг распределения, количество используемых данных).

4.5. Построить график точного распределения (используя функцию *dnorm*).

4.7. Оформить отчет по выполненной работе.



5. Содержание отчета

5.1. Назначение и цель работы.

5.2. Текст вычислений в системе *MathCad*.

5.3. График частотного распределения.

5.4. График точного распределения.

5.5. Выводы по результатам выполненной работы.

Файл *D200gzmm.prn*:

13.39 13.43 13.54 13.64 13.40 13.55 13.40 13.26 13.42 13.50 13.32 13.31 13.28
13.52 13.46 13.63 13.38 13.44 13.52 13.53 13.37 13.33 13.24 13.13 13.53 13.53
13.39 13.57 13.51 13.34 13.39 13.47 13.51 13.48 13.62 13.58 13.57 13.33 13.51
13.40 13.30 13.48 13.40 13.57 13.51 13.40 13.52 14.56 13.40 13.34 13.23 13.37
13.48 13.48 13.62 13.35 13.40 13.36 13.45 13.48 13.29 13.58 13.44 13.56 13.28
13.59 13.47 13.46 13.62 13.54 13.20 13.38 13.43 13.36 13.56 13.51 13.47 13.40
13.29 13.20 13.46 13.44 13.42 13.29 13.41 13.39 13.50 13.48 13.53 13.34 13.45
13.42 13.29 13.38 13.45 13.50 13.55 13.33 13.32 13.69 13.46 13.32 13.32 13.48
13.29 13.25 13.44 13.60 13.43 13.51 13.43 13.38 13.24 13.28 13.58 13.31 13.31
13.45 13.43 13.44 13.34 13.49 13.50 13.38 13.48 13.43 13.37 13.29 13.54 13.33
13.36 13.46 13.23 13.44 13.38 13.27 13.66 13.26 13.40 13.52 13.59 13.48 13.46
13.40 13.43 13.26 13.50 13.38 13.43 13.34 13.41 13.24 13.42 13.55 13.37 13.41
13.38 13.14 13.42 13.52 13.38 13.54 13.30 13.18 13.32 13.46 13.39 13.35 13.34
13.37 13.50 13.61 13.42 13.32 13.35 13.40 13.57 13.31 13.40 13.36 13.28 13.58
13.58 13.38 13.26 13.37 13.28 13.39 13.32 13.20 13.43 13.34 13.33 13.33 13.31
13.45 13.39 13.45 13.41 13.45

Лабораторная работа № 3

ОДНОФАКТОРНЫЙ И ДВУХФАКТОРНЫЙ АНАЛИЗЫ

1. Цель работы – знакомство с методами однофакторного и двухфакторного анализа. Закрепление навыков работы с системой *MathCad*.

2. Общие сведения

При исследовании зависимостей одной из наиболее простых является ситуация, когда можно указать только один фактор, влияющий на конечный результат, и этот фактор может принимать лишь конечное число значений (уровней). Такие задачи – задачи однофакторного анализа.

То, что должно оказывать влияние на конечный результат, называют фактором или факторами, если их несколько.

Конкретную реализацию фактора называют уровнем фактора или способом обработки.

Значение измеряемого признака (величину результата) называют откликом.

Критерий Краскела – Уоллиса (произвольные альтернативы). Критерий свободен от распределения для проверки гипотезы H_0 , если неизвестны альтернативы к H_0 . n_1, \dots, n_k – объемы выборок, $N = n_1 + \dots + n_k$ – общее число наблюдений.

$$\text{Средний ранг по столбцу } R_j = \frac{1}{n} \sum_{i=1}^{n_j} r_{ij}$$

$$\text{Статистика Краскела – Уоллиса: } H = \frac{12}{N(N+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3(N+1).$$

При больших объемах n_1, \dots, n_k случайная величина (при гипотезе H_0) приблизительно распределена как хи-квадрат с $k - 1$ степенями свободы. Отвергаем H_0 , если $H_{\text{набл}} > \chi^2_{1-\alpha}$, где $\chi^2_{1-\alpha}$ – квантиль уровня $(1 - \alpha)$ с распределением хи-квадрат с $k - 1$ степенями свободы.

Двухфакторный анализ. Пусть главный фактор A , а мешающий – B . Пусть A принимает k , а B – n различных значений. Фактор B разбивает все объекты наблюдений на n -блоков, каждый блок образуют наблюдения, проведенные при одном уровне фактора B . В блоке отклики могут значимо различаться только за счет применения к ним различных обработок, т.е. за счет различных уравнений фактора A .

Критерий Фридмана – непараметрический критерий проверки гипотез об отсутствии эффектов обработки.

$$S = \left[\frac{12}{nk(k+1)} \sum_{j=1}^k \left(\sum_{i=1}^n r_{ij} \right)^2 \right] - 3n(k+1)$$

3. Задание к лабораторной работе

Шести группам из пяти человек каждая предложили задачи одинаковой трудности. Каждому независимо от других. Группы отличаются величиной денежного вознаграждения. Проверить гипотезу об отсутствии влияния вознаграждения на число решенных задач.

4. Порядок выполнения работы

- 4.1. Получить у преподавателя исходные данные (файл *Tasks.prn*).
- 4.2. Вычислить средний ранг по столбцам.
- 4.3. Вычислить значение критерия Краскела – Уоллиса.

4.4. Выполнить проверку гипотезы об отсутствии влияния вознаграждения на число решенных задач.

4.5. Оформить отчет по выполненной работе.

5. Содержание отчета

5.1. Назначение и цель работы.

5.2. Текст вычислений в системе *MathCad*.

5.3. Выводы по результатам выполненной работы.

Файл *Tasks.prn*

10	8	12	12	24	19
11	10	17	15	16	18
9	16	14	16	22	27
13	13	9	16	18	25
7	12	16	19	20	24

Лабораторная работа № 4

ЛИНЕЙНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ

1. Цель работы – знакомство с методом линейного регрессионного анализа. Закрепление навыков работы с системой *MathCad*.

2. Общие сведения

Линейный регрессионный анализ объединяет широкий круг задач, связанных с построением функциональных зависимостей между двумя группами числовых переменных x_1, \dots, x_p и y_1, \dots, y_p , или между двумя многомерными переменными x и y , где x – независимая переменная, влияющая на значение y . Тогда y – отклик, а $x = (x_1, \dots, x_p)$ – факторы, влияющие на отклик. Исходные данные (x_i, y_i) , где y_i – значение отклика при заданном значении фактора x_i , $i = 1, \dots, n$. Пара (x_i, y_i) – результат одного измерения, n – число измерений.

Регрессионная модель: пусть значение отклика можно разделить на две части, одна зависит от x , другая – случайна по отношению к x .
 $y = f(x) + \varepsilon$, ε – ошибка эксперимента. Имеем $y_i = f(x_i) + \varepsilon_i$, $i = 1, \dots, n$.

$y = A + bx + \varepsilon \Rightarrow y_{ij} = A + bx_i + \varepsilon_{ij}$, где ε – внешний фактор,

$$y_{ij} = a + b(x_i - \bar{x}) + \varepsilon_{ij}.$$

Оценка параметров a и b :

$$\hat{a} = \bar{y} = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m y_{ij}; \quad y_i = \frac{1}{m} \sum_{j=1}^m y_{ij};$$

$$\hat{b} = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Для построения доверительных интервалов для истинных значений коэффициентов a и b и проверки качества модели надо найти оценку дисперсии:

$$S^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{a} - \hat{b}(x_i - \bar{x}))^2.$$

Таким образом, интервалы можно определить как:

$$a - \frac{S}{\sqrt{n}} t < a < a + \frac{S}{\sqrt{n}} t; \quad \hat{b} - \frac{S}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} t < b < \hat{b} + \frac{S}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} t,$$

где t – квантиль распределения Стьюдента со степенью свободы 4 при коэффициенте доверия $1 - \alpha$; $\alpha = 0,05 \Rightarrow t = 2,79$.

3. Задание к лабораторной работе

Измерительно-регистрационный тракт тензовесов используется для измерения сил и моментов сил, действующих на тело при продувке его в аэродинамической трубе.

На тензовесы оказывается воздействие эталонной силой (файл *Et_sila.prn*) и фиксируется значение отклика на выходе системы (файл *Kalibr.prn*). Эталонная сила – x_i , значение отклика – y_{ij} .

Определить зависимости между входом и выходом измерительно-регистрационной системы.

4. Порядок выполнения работы

- 4.1. Получить у преподавателя исходные данные (файлы *Et_sila.prn*, *Kalibr.prn*).
- 4.2. Вычислить значения оценки параметров a и b .
- 4.3. Вычислить значение оценки дисперсии.
- 4.4. Вычислить значения доверительных интервалов для параметров a и b .
- 4.5. Оформить отчет по выполненной работе.

5. Содержание отчета

- 5.1. Назначение и цель работы.
- 5.2. Текст вычислений в системе *MathCad*.
- 5.3. Выводы по результатам выполненной работы.

Файл *Kalibr.prn*

31.0	110.0	186.5	266.7	345.5	425.6
29.8	111.0	191.0	269.7	349.3	425.9
29.1	109.6	187.1	270.1	349.7	426.5
29.0	111.0	190.3	270.2	349.9	426.5
29.15	109.6	186.7	266.55	347.05	427.0
28.2	110.35	190.95	270.25	349.8	427.0

Файл *Et_sila.prn*

0.0 0.2 0.4 0.6 0.8 1.0

Лабораторная работа № 5

СТАТИСТИЧЕСКАЯ ОБРАБОТКА ДАННЫХ. ВЫДЕЛЕНИЕ ТРЕНДА

1. Цель работы – ознакомление с методикой выделения тренда из временного ряда. Закрепление навыков работы с системой *MathCad*.

2. Общие сведения

Временной ряд – последовательность чисел. Его элементы – значения некоторого, протекающего во времени, процесса. Значения измерены в надлежащие моменты времени, обычно через равные промежутки.

Трендом временного ряда tr_t при $t = 1, \dots, n$ называют плавно изменяющуюся, не циклическую компоненту, описывающую чистое влияние долговременных факторов, эффективность которых сказывается постоянно.

Адаптивная модель временного ряда:

$$x_t = D_t + E_t, \text{ при } t = 1, \dots, n \text{ или } X = D + E.$$

$$D_t = tr_t + S_t + C_t \text{ при } t = 1, \dots, n,$$

где tr_t – тренд, S_t – сезонная компонента, C_t – циклическая компонента.

Значение временного ряда X_t рассматривают как отклик (зависимую переменную), а время t – как фактор, влияющий на отклик (независимую переменную):

$$X_{it} = f(t_i, \Theta) + E_i, i = 1, \dots, n,$$

где f – функция тренда, Θ – неизвестные параметры модели временного ряда, а E_i – независимые нормально распределенные случайные величины.

Необходимо функцию тренда выбрать так, чтобы

$$\sum_{i=1}^n [X_{t_i} - f(t_i, \Theta)]^2 \rightarrow \min_{\Theta}.$$

Здесь тренд представляется в виде прямой: $tr_t = b_0 + b_1 t$.

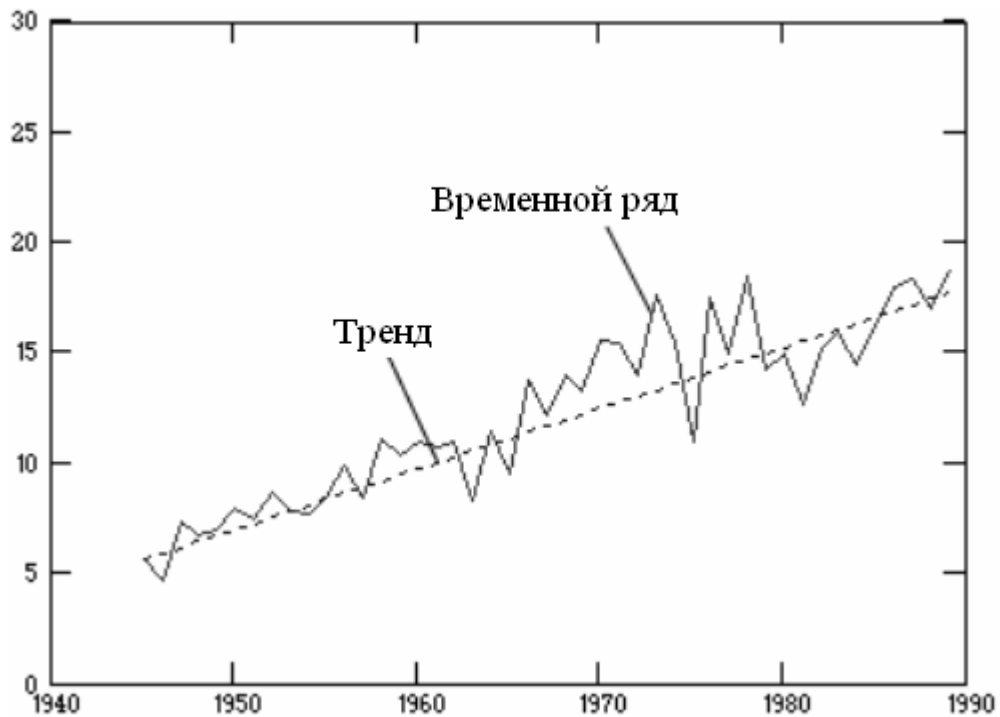
$$\hat{b}_0 = \bar{x}, \text{ где } \bar{x} = \frac{1}{n} \sum_{t=1}^n X_t, \quad \hat{b}_1 = \frac{\sum_{t=1}^n (x_t - \bar{x})([t+1] - \bar{t})}{\sum_{t=1}^n ([t+1] - \bar{t})^2},$$

\hat{b}_0 – средняя урожайность в начальный момент времени рассматриваемого ряда; \hat{b}_1 – оценка среднегодового прироста урожайности.

Кроме метода наименьших квадратов можно использовать метод перехода от исходного ряда к ряду разностей соседних значений ряда.

3. Задание к лабораторной работе

Удалить тренд из временного ряда (файл *Zerno.prn*) урожайности зерновых в СССР с 1945 по 1989 гг. (в центнерах с гектара в первоначально оприходованном весе), используя метод наименьших квадратов (используется для решения задач линейного регрессионного анализа (см. рисунок)).



4. Порядок выполнения работы

- 4.1. Получить у преподавателя исходные данные (файл *Zerno.prn*).
- 4.2. Построить график исходного временного ряда.
- 4.3. Вычислить значения тренда.
- 4.4. Построить график тренда.
- 4.5. Удалить из значений исходного временного ряда значения тренда.
- 4.6. Построить график полученного временного ряда.
- 4.7. Оформить отчет по выполненной работе.

5. Содержание отчета

- 5.1. Назначение и цель работы.
- 5.2. Текст вычислений в системе *MathCad*.

- 5.3. График исходного ряда.
- 5.4. График тренда исходного ряда.
- 5.5. График ряда после удаления тренда.
- 5.6. Выводы по результатам выполненной работы.

Файл *zerno.prn*

1945	5.6	1956	9.9	1967	12.1	1978	18.5
1946	4.6	1957	8.4	1968	14.0	1979	14.2
1947	7.3	1958	11.1	1969	13.2	1980	14.9
1948	6.7	1959	10.4	1970	15.6	1981	12.6
1949	6.9	1960	10.9	1971	15.4	1982	15.2
1950	7.9	1961	10.7	1972	14.0	1983	15.9
1951	7.4	1962	10.9	1973	17.6	1984	14.4
1952	8.6	1963	8.3	1974	15.4	1985	16.2
1953	7.8	1964	11.4	1975	10.9	1986	18.0
1954	7.7	1965	9.5	1976	17.5	1987	18.3
1955	8.4	1966	13.7	1977	15.0	1988	17.0
						1989	18.8

Лабораторная работа № 6

СТАТИСТИЧЕСКАЯ ОБРАБОТКА ДАННЫХ. ПРОГНОЗИРОВАНИЕ

1. Цель работы – ознакомление с методикой прогнозирования временных рядов. Закрепление навыков работы с системой *MathCad*.

2. Задание к лабораторной работе

По величинам месячного производства молока (в тысячах тонн) с января 1992 г. по октябрь 1996 г. спрогнозировать значение производства молока на ноябрь и декабрь 1996 г.

3. Порядок выполнения работы

Исходные данные – величины месячного производства молока (в тысячах тонн) в России с января 1992 г. по октябрь 1996 г.

Вычисляем значения тренда для исходного временного ряда. Линейная модель тренда: $tr_t = a + bt$, $t = 1, 2, 3, \dots, 58$, $a = 2841,1$, $b = -23,63$, где $\hat{tr} = \hat{a} + \hat{b}t$ – в каждой точке оценка тренда.

Строим таблицу значений отношений x_t/\hat{tr} . Необходимо упорядочить по строкам $x_t = tr_t S_t + \varepsilon_t$, $m = \frac{n}{p} - 1$, $n = 60$, $p = 12$.

$$\frac{x_{i+k_p}}{\hat{tr}_{i+k_p}} 100\%, i = 0, 1, 2, \dots, m.$$

Вычисляем сезонный индекс:

$$\hat{S}_i = \frac{1}{m+1} \sum_{k=0}^m \left(\frac{x_{i+k_p}}{\hat{tr}_{i+k_p}} 100\% \right), \text{ где } 1 \leq i \leq p.$$

Вычисляем окончательный прогноз $(prog_t \cdot S_i) / 100$, где $prog_t$ – предварительный прогноз по тренду tr_t .

4. Содержание отчета

- 4.1. Назначение и цель работы.
- 4.2. Текст вычислений в системе *MathCad*.
- 4.3. График исходного ряда.
- 4.4. График тренда исходного ряда.
- 4.5. График прогнозируемых значений.
- 4.6. Выводы по результатам выполненной работы.

Файл *moloko.prn*

	1992	1993	1994	1995	1996
1	2015	1759	1510	1172	1038
2	2123	1773	1484	1226	1104
3	2624	2361	1988	1651	1439
4	2891	2649	2211	1859	1521
5	3335	3203	2559	2392	1827
6	4071	3936	3209	2864	2446
7	4040	3861	3204	2714	2369
8	3392	3321	2687	2420	2081
9	2467	2438	2031	1925	1577
10	2092	1760	1506	1338	1081
11	1494	1299	1050	984	0
12	1562	1345	1054	1020	0

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Перечислите основные понятия теории вероятностей.
2. Что такое функция распределения СВ, дисперсия СВ?
3. Сформулируйте определение закона распределения вероятностей СВ.
4. Что такое вариационный ряд?
5. Каким образом строят гистограмму выборки в системе *MathCad*?
6. Приведите примеры задач на факторный анализ.
7. В каких случаях применяют линейный регрессионный анализ?
8. Что такое тренд временного ряда и как его можно построить?
9. Как выглядит адаптивная модель временного ряда?
10. Опишите порядок прогнозирования значений по временному ряду.

РЕКОМЕНДАТЕЛЬНЫЙ БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Бешелев, С. Д. Математико-статистические методы экспертных оценок / С. Д. Бешелев, Ф. Г. Гурвич. – М. : Статистика, 1980. – 263 с.
2. Дьяконов, В. П. *Mathcad 8 PRO* в математике, физике и Internet / В. П. Дьяконов, И. В. Абраменкова. – М. : «Нолидж», 1999. – 512 с. – ISBN
3. Замков, О. О. Математические методы в экономике : учебник / О. О. Замков, А. В. Толстопятенко, Ю. Н. Черемных. – М. : МГУ им. М.В. Ломоносова ; ДИС, 1998. – 368 с.
4. Статистика : учебник / И. И. Елисеева, И. И. Егорова [и др.] ; под ред. проф. И. И. Елисеевой. – М. : ТК Велби ; изд-во Проспект, 2004. – 448 с.
5. Тюрин, Ю. Н. Статистический анализ данных на компьютере / Ю. Н. Тюрин, А. А. Макаров ; под ред. В. Э. Фигурнова. – М. : ИНФРА-М, 1998. – 528 с.

ОГЛАВЛЕНИЕ

Лабораторная работа № 1. Введение в теорию вероятностей и математическую статистику. Знакомство с системой <i>MathCad</i>	3
Лабораторная работа № 2. Законы распределения вероятностей	6
Лабораторная работа № 3. Однофакторный и двухфакторный анализ ..	9
Лабораторная работа № 4. Линейный регрессионный анализы	11
Лабораторная работа № 5. Статистическая обработка данных. Выделение тренда	13
Лабораторная работа № 6. Статистическая обработка данных. Прогнозирование	16
Контрольные вопросы	18
Рекомендательный библиографический список	18

ПРЕДСТАВЛЕНИЕ ЗНАНИЙ
В ИНФОРМАЦИОННЫХ СИСТЕМАХ

Методические указания к лабораторным работам

Составители

АЛЕКСАНДРОВ Дмитрий Владимирович

ЛАРЮШКИН Андрей Юрьевич

Ответственный за выпуск – зав. кафедрой профессор А.В. Костров

Редактор Е.А. Амирсейидова

Корректор В. В. Гурова

Компьютерный набор А.Ю. Ларюшкина

Компьютерная верстка Д.В. Александрова

ЛР № 020275. Подписано в печать 16.03.05.

Формат 60x84/16. Бумага для множит. техники. Гарнитура Таймс.

Печать на ризографе. Усл. печ. л. 1,16. Уч.-изд. л. 1,24. Тираж 100 экз.

Заказ

Редакционно-издательский комплекс

Владимирского государственного университета.

600000, Владимир, ул. Горького, 87.

Министерство образования и науки Российской Федерации
Федеральное агентство по образованию
Владимирский государственный университет
Кафедра информационных систем и информационного менеджмента

ПРЕДСТАВЛЕНИЕ ЗНАНИЙ В ИНФОРМАЦИОННЫХ СИСТЕМАХ

Методические указания к лабораторным работам

Составители
Д.В. Александров
А.Ю. Ларюшкин

«В печать»

Составители:	Д.В. Александров А.Ю. Ларюшкин
Зав. кафедрой	А.В. Костров
Редактор	Е.А. Амирсейидова
Корректор	
Начальник РИО	Е.П. Викулова
Директор РИК	Ю.К. Жулев
Проректор ВлГУ по ИТ	В.А. Немонтов

Владимир 2005