

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение  
высшего профессионального образования

**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**  
ЮРГИНСКИЙ ТЕХНОЛОГИЧЕСКИЙ ИНСТИТУТ

---

УТВЕРЖДАЮ

Зам. директора ЮТИ ТПУ по УР

\_\_\_\_\_ В.Л. Бибик

« \_\_ » \_\_\_\_\_ 2016 г.

**А.А. Мицель**

## **СИСТЕМНЫЙ АНАЛИЗ, УПРАВЛЕНИЕ И ОБРАБОТКА ИНФОРМАЦИИ В АНАЛИТИЧЕСКОЙ ИНФОРМАЦИИ**

Методические указания к выполнению лабораторных работ  
по дисциплине «Системный анализ, управление и обработка информации в аналитиче-  
ской экономике»

для магистрантов, обучающихся по направлению 09.04.03  
«Прикладная информатика»

Издательство

Юргинского технологического института (филиала)  
Томского политехнического университета

2016

УДК 681  
ББК 32.81  
З–38

Составитель **Мицель А.А.**

З–38 Системный анализ, управление и обработка информации в аналитической экономике: методические указания к выполнению лабораторных работ по дисциплине «Системный анализ, управление и обработка информации в аналитической экономике» для магистрантов, обучающихся по направлению 9.04.03 «Прикладная информатика» / Мицель А.А.; Юргинский технологический институт. – Юрга: Изд–во Юргинского технологического института (филиала) Томского политехнического университета, 2016. – 62 с.

УДК 681  
ББК 32.81

Методические указания рассмотрены и рекомендованы  
к изданию методическим семинаром кафедры  
информационных систем ЮТИ ТПУ  
« » \_\_\_\_\_ 2016 г.

Зав. кафедрой ИС  
кандидат технических наук

\_\_\_\_\_ *А.А. Захарова*

Председатель учебно–методической  
комиссии

\_\_\_\_\_ *Е.В. Молнина*

*Рецензент*

Кандидат технических наук, доцент кафедры ИС ЮТИ ТПУ  
*А.В. Маслов*

© ФГБОУ ВПО НИ ТПУ Юргинский  
технологический институт (филиал), 2016  
© Мицель А.А., 2016

## Оглавление

ЛАБОРАТОРНАЯ РАБОТА 1 .....	4
<i>Тема 1. Генерация случайных чисел с заданным законом распределения</i>	4
ЛАБОРАТОРНАЯ РАБОТА 2 .....	15
<i>Тема 2. Свойства выборочных характеристик</i> .....	15
ЛАБОРАТОРНАЯ РАБОТА 3 .....	19
<i>Тема 3. Оценка закона распределения на основе выборочных данных</i> ..	19
ЛАБОРАТОРНАЯ РАБОТА 4 .....	35
<i>Тема 4. Линейная регрессия</i> .....	35
ЛАБОРАТОРНАЯ РАБОТА 5 .....	50
<i>Тема 5. Комплексная оценка финансового</i> .....	50
<i>состояния предприятия</i> .....	50
СОДЕРЖАНИЕ ОТЧЕТА ПО ЛАБОРАТОРНЫМ РАБОТАМ .....	61
СПИСОК ЛИТЕРАТУРЫ .....	62
Приложение А .....	63

# ЛАБОРАТОРНАЯ РАБОТА 1

## Тема 1. Генерация случайных чисел с заданным законом распределения

### *Цель работы:*

1. Научиться использовать генератор случайных чисел пакета EXCEL для генерации случайных чисел с заданным законом распределения.
2. Познакомиться со способами представления выборочных данных.

### **1. Практическое задание**

1. Задан закон распределения  $F$  *дискретной* случайной величины (приложение 1). Требуется:
  - a) Сгенерировать средствами пакета EXCEL выборку из 100 значений случайной величины с законом  $F$  (для генерации случайных чисел распределенных по законам, которые отсутствуют в генераторе случайных чисел пакета EXCEL см. приложение 3).
  - b) Представить выборку в виде вариационного ряда.
  - c) Построить статистический ряд абсолютных частот, относительных частот и накопленных частот.
  - d) Построить полигон частот и сравнить его с многоугольником теоретического распределения  $F$ .
  - e) Найти основные выборочные характеристики –  $\bar{X}$  и  $s^2$  и сравнить их с математическим ожиданием и дисперсией теоретического распределения  $F$ .
2. Задан закон распределения  $F$  *непрерывной* случайной величины (приложение 1). Требуется:
  - a) Сгенерировать средствами пакета EXCEL выборку из 100 значений случайной величины с законом  $F$  (для генерации случайных чисел распределенных по законам, которые отсутствуют в генераторе случайных чисел пакета EXCEL см. приложение 3).
  - b) Представить выборку в виде вариационного ряда.
  - c) Построить сгруппированный статистический ряд абсолютных частот, относительных частот и плотностей частот.
  - d) Построить гистограмму и сравнить ее с графиком плотности теоретического распределения  $F$ . Для корректного сопоставления гистограммы с графиком плотности теоретического распределения, следует помнить, что EXCEL при одновременном отображении графика и гистограммы, помещает точки графика в середину столбца гистограммы.

Следовательно, значения плотности должны быть подсчитаны для середин столбцов гистограммы.

е) Построить график эмпирической функции распределения и сравнить с графиком теоретического распределения  $F$  (для построения графиков использовать не менее 40 точек).

ф) Найти основные выборочные характеристики –  $\bar{X}$  и  $s^2$  и сравнить их с математическим ожиданием и дисперсией теоретического распределения  $F$ .

## Приложение 1 Варианты заданий

Вариант 1.

1).  $F$  – биномиальное распределение с параметрами  $n = 20$  и  $p = 0,7$ .

2).  $F$  – распределение  $\chi^2$  с одной степенью свободы.

Вариант 2.

1).  $F$  – биномиальное распределение с параметрами  $n = 100$  и  $p = 0,15$ .

2).  $F$  – закон равномерной плотности на  $(-2; 5)$ .

Вариант 3.

1)  $F$  – биномиальное распределение с  $n = 50$  и  $p = 0,42$ .

2)  $F$  – показательный закон с параметром  $\lambda = 0,04$ .

Вариант 4.

1).  $F$  – закон Пуассона с параметром  $\lambda = 8$ .

2).  $F$  – распределение  $\chi^2$  с 2 степенями свободы.

Вариант 5..

1).  $F$  – биномиальное распределение с параметрами  $n = 80$  и  $p = 0,2$ .

2).  $F$  – распределение Стьюдента с 3 степенями свободы.

Вариант 6.

1).  $F$  – закон Пуассона с параметром  $\lambda = 12$ .

2).  $F$  – распределение Коши с плотностью  $f(x) = 1/(\pi(1+x^2))$ ,  $x \in R$ .

Вариант 7.

- 1).  $F$  – биномиальное распределение с параметрами  $n = 30$  и  $p = 0,6$ .
- 2).  $F$  – нормальный закон с параметрами  $a = 0$  и  $\sigma = 3$ .

Вариант 8.

- 1).  $F$  – геометрическое распределение с параметром  $p = 0,2$ .
- 2).  $F$  – нормальный закон с параметрами  $a = -2$  и  $\sigma = 3$ .

Вариант 9.

- 1).  $F$  – закон Пуассона с параметром  $\lambda = 10$ .
- 2).  $F$  – показательный закон с параметром  $\lambda = 0,1$ .

Вариант 10.

- 1).  $F$  – биномиальное распределение с параметрами  $n = 50$  и  $p = 0,3$ .
- 2).  $F$  – распределение  $\chi^2$  с одной степенью свободы.

## Приложение 2

### Генерация случайных чисел

#### *Генерация случайных чисел при помощи пакета EXCEL.*

Для генерации случайных чисел в пакете EXCEL 2003 выберите меню “Сервис”, “Пакет анализа”, ”Генерация случайных чисел”. Если в меню “Сервис” отсутствует подменю “Пакет анализа”, следует зайти в меню “Сервис”, “Надстройки” и подключить “Пакет анализа”. В EXCEL 2007 выберите меню “Данные”, “Анализ данных”. Если в меню “Данные” отсутствует подменю “Анализ данных”, то следует нажать кнопку “Office”, перейти в “Параметры Excel”, выбрать “Надстройки”, нажать “Перейти” и подключить “Пакет анализа”. Параметры генератора случайных чисел “число переменных” и ”число случайных чисел”. определяют соответственно число столбцов и строк для вывода случайных чисел. Можно использовать, например, “число переменных” для того, чтобы получить сразу несколько независимых выборок – столбцов объема, определяемого параметром “число случайных чисел”.

#### **Статистические функции пакета EXCEL**

**1. Функции связанные с основными законами распределения случайных величин**

**БИНОМРАСП (число успехов; число испытаний; вероятность успеха; интегральная)**

Возвращает вероятности связанные с биномиальным распределением. Функция БИНОМРАСП используется для подсчета вероятностей числа успехов в испытаниях по схеме Бернулли.

**Число успехов** – количество успешных испытаний ( $m$ ).

**Число испытаний** – общее число независимых испытаний ( $n$ ).

**Вероятность успеха** – вероятность успеха в каждом испытании ( $p$ ).

**Интегральная.** – это логическое значение, определяющее форму функции. Если аргумент интегральная имеет значение ИСТИНА (1), то функция БИНОМРАСП возвращает вероятность того, что число успешных испытаний не более значения число успехов; если этот аргумент имеет значение ЛОЖЬ (0), то возвращается вероятность того, что число успешных испытаний в точности равно значению аргумента число успехов.

Таким образом:

$$\text{БИНОМРАСП}(m; n; p; 0) = P_n(m) = C_n^m p^m (1-p)^{n-m};$$

$$\text{БИНОМРАСП}(m; n; p; 1) = \sum_{k=0}^m P_n(k) = \sum_{k=0}^m C_n^k p^k (1-p)^{n-k}.$$

**ПУАССОН ( $x$ ; среднее; интегральная)**

Возвращает вероятности, связанные с распределением Пуассона (например, вероятности числа событий в простейшем потоке за некоторый промежуток времени, при известном среднем числе событий)

$x$  – количество событий.

**Среднее** – среднее число событий ( $\lambda$ ).

**Интегральная** – логическое значение, определяющее форму возвращаемого распределения вероятностей. Если аргумент «интегральная» имеет значение ИСТИНА (1), то функция ПУАССОН возвращает вероятность того, что число случайных событий будет от 0 до  $x$  включительно. Если этот аргумент имеет значение ЛОЖЬ (0), то возвращается вероятность того, что событий будет в точности  $x$ .

Таким образом:  $\text{ПУАССОН}(x; \lambda; 0) = \lambda^x e^{-\lambda} / x!$ ;  $\text{ПУАССОН}(x; \lambda; 1)$

$$= \sum_{k=0}^x \lambda^k e^{-\lambda} / k!.$$

**ГИПЕРГЕОМЕТ (число успехов в выборке; размер выборки; число успехов в совокупности; размер совокупности).**

Возвращает вероятности для гипергеометрического распределения. ГИПЕРГЕОМЕТ возвращает вероятность заданного количества успехов в выборке, если заданы размер выборки, количество успехов в генеральной совокупности и размер генеральной совокупности.

**Число успехов в выборке** – это количество успешных испытаний в выборке ( $m$ ).

**Размер выборки** – размер выборки ( $n$ ).

**Число успехов в совокупности** – количество успешных испытаний в генеральной совокупности ( $M$ ).

**Размер совокупности** – размер генеральной совокупности ( $N$ ).

Например, из генеральной совокупности, содержащей  $N$  шаров, среди которых  $M$  красных, выбирается наудачу  $n$  шаров. Тогда, вероятность того, что среди них ровно  $m$  красных равна:  $P = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n} =$  ГИ-

ПЕРГЕОМЕТ( $m; n; M; N$ ).

**НОРМРАСП( $x$ ; среднее; стандартное откл; интегральная)**

Возвращает вероятности, связанные с нормальным распределением.

$x$ – значение, для которого определяется вероятность.

**Среднее** – математическое ожидание распределения ( $a$ ).

**Стандартное откл** – среднеквадратическое отклонение распределения ( $\sigma$ ).

**Интегральная** – логическое значение, определяющее форму функции. Если интегральная имеет значение ИСТИНА (1), то функция НОРМРАСП возвращает функцию распределения от аргумента  $x$ ; если это аргумент имеет значение ЛОЖЬ (0), то возвращается плотности распределения от аргумента  $x$ .

Таким образом:

$$\text{НОРМРАСП}(x; a; \sigma; 0) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}};$$

$$\text{НОРМРАСП}(x; a; \sigma; 1) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(t-a)^2}{2\sigma^2}} dt.$$

**НОРМСТРАСП( $z$ )**

Возвращает функцию распределения стандартной нормальной величины, т.е.  $\text{НОРМСТРАСП}(z) = P(\xi < z) = F_\xi(z)$ , где  $\xi \in N_{0,1}$ ,

$$F_\xi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{t^2}{2}} dt.$$

**НОРМОБР(вероятность; среднее; стандартное откл)**

Возвращает квантиль нормального распределения для указанной вероятности, то есть  $\text{НОРМСТОБР}(\alpha)$  возвращает значение  $\tau_\alpha$ , для которого  $P(\xi < \tau_\alpha) = \alpha$ ,  $\xi \in N_{a,\sigma^2}$ .

**Вероятность** – вероятность, соответствующая квантили.



**Среднее** – математическое ожидание распределения.

**Стандартное откл** – среднеквадратическое отклонение распределения.

**НОРМСТОБР(вероятность)**

Возвращает квантиль стандартного нормального распределения для указанной вероятности, то есть  $\text{НОРМСТОБР}(\alpha)$  возвращает значение  $\tau_\alpha$ , для которого  $P(\xi < \tau_\alpha) = \alpha$ ,  $\xi \in N_{0,1}$

**Вероятность** – вероятность, соответствующая квантили.

**СТЬЮДРАСП(x; степени свободы; хвосты)**

Возвращает вероятности, связанные с распределением Стьюдента.

$X$  – численное значение, для которого требуется вычислить вероятности.

**Степени свободы** – число степеней свободы распределения.

**Хвосты** – число учитываемых хвостов распределения. Если хвосты = 1, то функция СТЬЮДРАСП возвращает вероятность того, что случайная величина, распределенная по закону Стьюдента, примет значение большее чем  $x$ . Т.е.  $\text{СТЬЮДРАСП}(x; n; 1) = P(\xi > x) = 1 - F_\xi(x)$ , где  $\xi \in T_n$ . Если хвосты = 2, то функция СТЬЮДРАСП возвращает вероятность того, что случайная величина, распределенная по закону Стьюдента, примет значение, большее, чем  $|x|$ . Т.е.  $\text{СТЬЮДРАСП}(x; n; 2) = P(|\xi| > x)$ , где  $\xi \in T_n$ .

**СТЬЮДРАСПОБР(вероятность; степени свободы)**

возвращает коэффициент Стьюдента  $t_\alpha$ , соответствующий заданной вероятности  $\alpha$ : т.е. значение  $t_\alpha$  для которого  $P(|\xi| > t_\alpha) = \alpha$ , что тоже самое, что квантиль распределения Стьюдента уровня  $1 - \alpha/2$ , то есть значение  $\tau_{1-\alpha/2}$ , для которого  $P(\xi < \tau_{1-\alpha/2}) = 1 - \alpha/2$ ,  $\xi \in T_n$ .

**Вероятность** – вероятность, для которой находится значение коэффициента.

**Степени свободы** – число степеней свободы, характеризующее распределение.

**ХИ2РАСП(x; степени свободы)**

Возвращает вероятность того, что случайная величина, распределенная по закону хи-квадрат примет значение, большее, чем  $x$ , т.е.  $\text{ХИ2РАСП}(x; n) = P(\xi > x) = 1 - F_\xi(x)$ , где  $\xi \in \chi_n^2$ .

$X$  – это значение, для которого требуется вычислить вероятность.

**Степени свободы** – это число степеней свободы распределения хи-квадрат.

**ХИ2ОБР(вероятность; степени свободы)**

возвращает критическую точку распределения хи-квадрат для заданной вероятности, то есть  $\text{ХИ2ОБР}(\alpha, n) = t_\alpha$ , где  $t_\alpha$  значение, для которого  $P(\xi > t_\alpha) = \alpha$ , что тоже самое, что квантиль распределения хи-квадрат уровня  $1 - \alpha$ .

**Вероятность** – вероятность, для которой находится критическая точка.

**Степени свободы** – число степеней свободы, характеризующее распределение.

**ФРАСП(x; степени свободы1; степени свободы2)**

Возвращает вероятность того, что случайная величина, распределенная по закону Фишера примет значение, большее, чем  $x$ , т.е.  $\text{ФРАСП}(x; n1; n2) = P(\xi > x) = 1 - F_\xi(x)$ , где  $\xi \in F_{n1, n2}$ .

$X$  – это значение, для которого требуется вычислить вероятность.

**Степени свободы 1, степени свободы2** – число степеней свободы, характеризующих распределение.

**ФРАСПОБР(вероятность; степени свободы1; степени свободы2)**

возвращает критическую точку распределения Фишера для заданной вероятности, то есть  $\text{ФРАСПОБР}(\alpha, n1, n2) = t_\alpha$ , где  $t_\alpha$  значение, для которого  $P(\xi > t_\alpha) = \alpha$ , что тоже самое, что квантиль распределения Фишера уровня  $1 - \alpha$ .

**Вероятность** – вероятность, для которой находится критическая точка.

**Степени свободы1, степени свободы2** – число степеней свободы, характеризующих распределение.

## 2. Примеры использования функций

**Пример 1.** Игральная кость подбрасывается 24 раза. Найти вероятность того, что 6 очков выпадут ровно 3 раза. Найти точное значение вероятности и приближенные, используя локальную формулу Муавра–Лапласа и формулу Пуассона.

**Решение.** Требуется найти вероятность того, что в  $n=24$  испытаниях по схеме Бернулли с вероятностью успеха  $1/6$ , число успехов будет равно 3. Для точного вычисления вероятности используем функцию **БИНОМРАСП**. Если параметр **интегральная** имеет значение **ЛОЖЬ (0)**, то функция **БИНОМРАСП** возвращает вероятность того, что число успешных испытаний в точности равно значению аргумента **число успехов**. Таким образом, для  $n = 24$ ,  $m = 3$ ,  $p = 1/6$  находим:

$$P_{24}(3) = \text{БИНОМРАСП}(3; 24; 1/6; 0) = 0,203681.$$

Найдем приближенное значение вероятности, используя локальную формулу Муавра–Лапласа. Согласно этой формуле, вероятность

$$P_n(m) \approx \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{npq}} e^{-\frac{(m-np)^2}{2npq}}, \text{ т.е. приближенно равна плотности нормального}$$

распределения со средним  $np$  и среднеквадратичным отклонением  $\sqrt{npq}$  в точке  $m$ . Значения плотности распределения нормальной величины возвращает функция **НОРМРАСП**, при значении параметра **интегральная** равном **ЛОЖЬ (0)**.

Таким образом:  $P_{24}(3) \approx \text{НОРМРАСП}(3; 24 \cdot 1/6; \sqrt{24 \cdot 1/6 \cdot 5/6}; 0) = 0,188073$ .

Найдем приближенное значение той же вероятности, используя формулу Пуассона. Согласно этой формуле, при малых  $p$  вероятность

$$P_n(m) \approx \frac{\lambda^m e^{-\lambda}}{m!}, \lambda = np, \text{ т.е. приближенно равна вероятности пуассоновско-$$

го распределения с параметром (средним значением)  $\lambda = np$  в точке  $m$ .

Вероятности отдельных значений для распределения Пуассона возвращает функция ПУАССОН при значении параметра **интегральная** равном **ЛОЖЬ (0)**. Таким образом:  $P_{24}(3) \approx \text{ПУАССОН}(3; 24 \cdot 1/6; 0) = 0,195367$ .

Заметим, что погрешность при использовании формулы Муавра–Лапласа составила  $\approx 7,8\%$ , а при использовании формулы Пуассона  $\approx 4,1\%$ .

**Пример 2.** Вероятность искажения одного символа при передаче сообщения равна 0,01. Какова вероятность, что сообщение, содержащее 200 символов, содержит не более 2–х искажений. Найти точное значение вероятности и приближенные, используя локальную формулу Муавра–Лапласа и формулу Пуассона.

**Решение.** Требуется найти вероятность того, что в  $n=200$  испытаниях по схеме Бернулли с вероятностью успеха 0,01, число успехов будет не более 2. Для точного вычисления вероятности используем функцию **БИНОМРАСП**. Если параметр **интегральная** имеет значение **ИСТИНА (1)**, то функция **БИНОМРАСП** возвращает вероятность того, что число успешных испытаний лежит в пределах от 0 до значения, определяемого аргументом **число успехов**. Таким образом, для  $n = 200$ ,  $m = 2$ ,  $p = 0,01$  находим:  $P_{200}(0 \leq m \leq 2) = \text{БИНОМРАСП}(2; 200; 0,01; 1) = 0,676679$ .

Найдем приближенное значение вероятности, используя локальную формулу Муавра–Лапласа. Согласно этой формуле, вероятность  $P_n(m)$ , приближенно равна плотности нормального распределения в точке  $m$  со средним  $np$  и среднеквадратичным отклонением  $\sqrt{npq}$ . Значения плотности распределения нормальной величины возвращает функция

**НОРМРАСП**, при значении параметра **интегральная** равно **ЛОЖЬ (0)**. Таким образом:  $P_{200}(0 \leq m \leq 2) = P_{200}(0) + P_{200}(1) + P_{200}(2) \approx$   
 $\approx$ НОРМ-  
РАСП(0; 2;  $\sqrt{1,98}$ ; 0)+НОРМРАСП(1; 2;  $\sqrt{1,98}$ ; 0)+НОРМРАСП(2; 2;  $\sqrt{1,98}$ ; 0)  
 $=0,607013$ .

Найдем приближенное значение той же вероятности, используя формулу Пуассона. Согласно этой формуле, вероятность  $P_n(m)$  при малых  $p$  приближенно равна вероятности пуассоновского распределения в точке  $m$  со средним значением  $\lambda = np$ . Значения вероятностей для распределения Пуассона возвращает функция **ПУАССОН**. Причем, если значение параметра **интегральная** равно **ИСТИНА (1)**, то функция **ПУАССОН** возвращает вероятность того, что случайная величина, имеющая распределение Пуассона примет значения в пределах от 0 до значения, определяемого аргументом  $x$ . Таким образом:  
 $P_{200}(0 \leq m \leq 2) \approx$ ПУАССОН(2;  $200 \cdot 0,01$ ; 1) $=0,676676$ .

Заметим, что погрешность, полученная при использовании формулы Пуассона, в данном случае на порядок ниже, чем при использовании локальной формулы Муавра–Лапласа. Смысла использовать интегральную формулу Муавра–Лапласа в данном случае нет, поскольку интервал значений  $m$  мал ( $0 \leq m \leq 2$ ).

Найдем приближенное значение вероятности, используя теперь интегральную формулу Муавра–Лапласа. Согласно этой формуле, вероятность  $P_n(m_1 \leq m \leq m_2) \approx \Phi\left(\frac{m_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{m_1 - np}{\sqrt{npq}}\right)$ , т.е. приближенно равна вероятности попадания в интервал  $(m_1; m_2)$  нормальной случайной величины со средним  $np$  и среднеквадратичным отклонением  $\sqrt{npq}$ . Следовательно, если  $F_\xi(x)$  – функция распределения нормальной случайной величины с параметрами  $a = np$  и  $\sigma^2 = npq$ , то  $P_n(m_1 \leq m \leq m_2) \approx F_\xi(m_2) - F_\xi(m_1)$ . Значения функции распределения нормальной величины возвращает функция **НОРМРАСП**, при значении параметра **интегральная** равно **ИСТИНА (1)**. Таким образом:  
 $P_n(0 \leq m \leq 2) \approx$ НОРМРАСП(2; 2;  $\sqrt{1,98}$ ; 1)–НОРМРАСП(0; 2;  $\sqrt{1,98}$ ; 1) $=$ .

**Пример 3.** Монета подбрасывается 10000 раз. Найти вероятность того, что орел выпадет более 5100 раз. Найти точное значение вероятности и приближенное, используя интегральную формулу Муавра–Лапласа.

**Решение.** Требуется найти вероятность того, что в  $n=10000$  испытаниях по схеме Бернулли с вероятностью успеха  $1/2$ , число успехов будет более 5100. Для точного вычисления вероятности используем функцию **БИНОМРАСП**. Если параметр **интегральная** имеет значение **ИСТИНА (1)**, то функция **БИНОМРАСП** возвращает вероятность того, что число успешных испытаний не менее значения аргумента **число успехов**. Таким образом, находим:  $P_{10000}(m > 5100) = 1 - P_{10000}(m \leq 5100) = 1 - \text{БИНОМРАСП}(5100; 10000; 1/2; 1) = 0,022213$ .

Найдем приближенное значение вероятности, используя интегральную формулу Муавра–Лапласа. Согласно этой формуле, вероятность  $P_n(m_1 \leq m \leq m_2) \approx \Phi\left(\frac{m_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{m_1 - np}{\sqrt{npq}}\right)$ , т.е. приближенно равна вероятности попадания в интервал  $(m_1; m_2)$  нормальной случайной величины со средним  $np$  и среднеквадратичным отклонением  $\sqrt{npq}$ . Следовательно, если  $F_\xi(x)$  – функция распределения нормальной случайной величины с параметрами  $a = np$  и  $\sigma^2 = npq$ , то  $P_n(m_1 \leq m \leq m_2) \approx F_\xi(m_2) - F_\xi(m_1)$ .

Значения функции распределения нормальной величины возвращает функция **НОРМРАСП**, при значении параметра **интегральная** равном **ИСТИНА (1)**. Таким образом:

$$P_{10000}(m > 5100) = 1 - P_{10000}(m \leq 5100) \approx 1 - F_\xi(5100) = \text{НОРМРАСП}(5100; 10000 \cdot 1/2; \sqrt{10000 \cdot 1/2 \cdot 1/2}; 1) = 0,02275.$$

Заметим, что более точной является приближенная формула  $P_n(m_1 \leq m \leq m_2) \approx F_\xi(m_2 + 0,5) - F_\xi(m_1 - 0,5)$ . Если использовать ее, то получим:

$$P_{10000}(m > 5100) = 1 - P_{10000}(m \leq 5100) \approx 1 - F_\xi(5100 + 1/2) = \text{НОРМРАСП}(5100; 10000 \cdot 1/2; \sqrt{10000 \cdot 1/2 \cdot 1/2}; 1) = 0,022216.$$

**Пример 4.** Случайная величина распределена по нормальному закону с параметрами  $a = 3$  и  $\sigma^2 = 7$ . Найти:

- а) вероятность того, что  $\xi$  примет значение в интервале  $(1; 4)$ ;
- б) квантиль распределения уровня  $0,85$ ;
- в) критическую точку распределения уровня  $0,07$ ;
- г) интервал, симметричный относительно математического ожидания, в котором с вероятностью  $0,95$  содержатся значения  $\xi$ .

**Решение.**

а) Вероятности, связанные с нормальным распределением, можно вычислять, используя функцию **НОРМРАСП**. Данная функция, при значении параметра **интегральная** равном **ИСТИНА(1)**, возвращает значения функции распределения  $F_{\xi}(x)$  нормальной случайной величины.

Поскольку  $P(a < \xi < b) = F_{\xi}(b) - F_{\xi}(a)$ , то

$$P(1 < \xi < 4) = F_{\xi}(4) - F_{\xi}(1) = \\ = \text{НОРМРАСП}(4; 3; \sqrt{7}; 1) - \text{НОРМРАСП}(1; 3; \sqrt{7}; 1) = 0,422426.$$

б) Квантили, критические точки и, вообще, различные значения, связанные с вероятностями для нормальной случайной величины, можно вычислять, используя функцию **НОРМОБР** или **НОРМСТОБР**. Функция **НОРМОБР** возвращает квантиль нормального распределения для указанной вероятности, то есть  $\text{НОРМОБР}(\beta; a; \sigma) = \tau_{\beta}$ , для которого  $P(\xi < \tau_{\beta}) = \beta$ ,  $\xi \in N_{a, \sigma^2}$ . Таким образом, квантиль уровня 0,85 равна  $\tau_{0,85} = \text{НОРМОБР}(0,85; 3; \sqrt{7}) = 5,742$ .

в) Критическая точка уровня  $\beta$  по определению есть значение  $t_{\beta}$ , для которого  $P(\xi > t_{\beta}) = \beta$ . Критическая точка уровня  $\beta$  совпадает с квантилью уровня  $1 - \beta$ . Поэтому, критическая точка уровня 0,07 есть  $t_{0,07} = \tau_{0,93} = \text{НОРМОБР}(0,07; 3; \sqrt{7}) = 6,905$ .

г) Требуется найти такое значение  $\delta$ , для которого  $P(|\xi - M(\xi)| < \delta) = 0,95$ . Удобнее в данном случае воспользоваться функцией **НОРМСТОБР**, которая возвращает квантиль стандартного нормального распределения для указанной вероятности. Имеем:  $P(|\xi - M(\xi)| < \delta) = 2\Phi(\delta/\sigma) = 2F_{0,1}(\delta/\sigma) - 1 = 0,95$ , где  $F_{0,1}(x)$  – функция распределения стандартной нормальной величины. Или  $F_{0,1}(\delta/\sigma) = 1,95/2 = 0,975$ . Используя **НОРМСТОБР**, находим квантиль для стандартной нормальной величины уровня 0,975:  $\tau_{0,975} = \text{НОРМСТОБР}(0,975) = 1,96$ . Тогда  $\delta/\sigma = 1,96$ , откуда  $\delta = \sigma \cdot 1,96 = 5,186$ . Таким образом, искомый интервал имеет вид:  $(a - \delta; a + \delta) = (-2,186; 8,186)$ .

## ЛАБОРАТОРНАЯ РАБОТА 2

### Тема 2. Свойства выборочных характеристик

**Цель работы:**

Исследование свойств основных числовых характеристик выборки.

**Задание**

Задан закон распределения  $F$  случайной величины  $\xi$  (приложение 1). Требуется:

а) Для каждого из  $n = \{15, 60, 240, 960\}$ , сгенерировать, используя генератор случайных чисел пакета EXCEL, по 10 выборок объемом  $n$  из генеральной совокупности  $F$ . Для каждой выборки определить выборочные характеристики:  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  и  $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2$ . Данные по  $\bar{X}$  и  $s^2$  представить в виде таблиц:

Таблица 2.1.

*Значения выборочных средних  $\bar{X}$  для выборок различного объема*

Выборочное среднее $\bar{X}$				
№ п/п выборки	n=15	n=60	n=240	n=960
1				
2				
...				
10				
$\bar{X}_{\min}$				
$\bar{X}_{\max}$				
$W = \bar{X}_{\max} - \bar{X}_{\min}$				

Таблица 2.2.

*Значения  $s^2$  для выборок различного объема*

Выборочная дисперсия $s^2$				
№ п/п выборки	n=15	n=60	n=240	n=960
1				
2				

...				
10				
$s_{\min}^2$				
$s_{\max}^2$				
$W = s_{\max}^2 - s_{\min}^2$				

Сделать выводы о сходимости выборочных характеристик. Оценить изменение величины разброса  $W = \bar{X}_{\max} - \bar{X}_{\min}$  с ростом объема выборки.

б) Если наблюдается сходимость выборочного среднего, используя центральную предельную теорему определить для заданной в задании (приложение 1) вероятности  $\beta$  и величины отклонения  $\varepsilon$  необходимый объем выборки  $N$ , так чтобы  $P(|\bar{X} - M(\xi)| < \varepsilon) \approx \beta$ . Проверить, сгенерировав 10 выборок найденного объема  $N$  и подсчитав для каждой величину  $|\bar{X} - M(\xi)|$ .

*Справка.* Вероятность, что случайная величина  $\bar{X} - M(\xi)$  находится в интервале  $-\varepsilon < \bar{X} - M(\xi) < \varepsilon$ , равна

$$\begin{aligned}
 P(|\bar{X} - M(\xi)| < \varepsilon) &= \int_{-\sqrt{n}\varepsilon}^{\sqrt{n}\varepsilon} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{t^2}{2\sigma^2}} dt = \\
 &= 2 \int_0^{\sqrt{n}\varepsilon} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{t^2}{2\sigma^2}} dt = 2 \int_0^{\sqrt{n}\varepsilon/\sigma} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = 0,95
 \end{aligned}$$

Из таблицы интегралов вероятностей находим:  $\sqrt{n}\varepsilon/\sigma = 1,96$ .

$$\text{Отсюда } n = \left( \frac{1,96\sigma}{\varepsilon} \right)^2.$$

## Приложение 1 Варианты заданий

Вариант 1.

1)  $F$  – закон равномерной плотности на  $(1; 3)$ .  $\varepsilon = 0,01$ ,  $\beta = 0,95$ .

2)  $F$  – закон с плотностью распределения  $f(x) = \frac{1}{\pi} \frac{1}{(1+x^2)}$ ,

$$x \in (-\infty, \infty). \varepsilon = 0,001, \quad \beta = 0,9. \quad (F(x) = \frac{1}{2} + \frac{\text{arctg}(x)}{\pi}),$$

$$x = \text{tg}(\pi(z - 0,5)), \quad z \in U_{0,1}.$$



Вариант 2.

1)  $F$  – нормальный закон с параметрами  $a = -1$  и  $\sigma = 2$ .  $\varepsilon = 0,05$ ,  $\beta = 0,9$ .

2)  $F$  – закон с плотностью распределения  $f(x) = \frac{1}{(1+x)^2}$ ,  $x \in (0, \infty)$ .  $\varepsilon = 0,01$ ,  $\beta = 0,99$ .

Вариант 3.

3)  $F$  – биномиальное распределение с  $n = 20$  и  $p = 0,2$ .  $\varepsilon = 0,05$ ,  $\beta = 0,99$ .

4)  $F$  – закон с плотностью распределения  $f(x) = \frac{1}{\pi} \frac{1}{1+(x-2)^2}$ ,  $x \in (-\infty, \infty)$ .  $\varepsilon = 0,02$ ,  $\beta = 0,97$ .

Вариант 4.

1).  $F$  – закон Пуассона с параметром  $\lambda = 4$ .  $\varepsilon = 0,03$ ,  $\beta = 0,8$ .

2).  $F$  – закон с плотностью распределения  $f(x) = \frac{1/3}{(1+x/3)^2}$ ,  $x \in (0, \infty)$ .  $\varepsilon = 0,01$ ,  $\beta = 0,9$ .

Вариант 5.

1.  $F$  – показательный закон с параметром  $\alpha = 0,8$ .  $\varepsilon = 0,02$ ,  $\beta = 0,85$ .

2.  $F$  – закон с плотностью распределения  $f(x) = \frac{1}{\pi} \frac{1}{1+(x+1)^2}$ ,  $x \in (-\infty, \infty)$ .  $\varepsilon = 0,03$ ,  $\beta = 0,85$

Вариант 6.

1.  $F$  – показательный закон с параметром.  $\alpha = 2$ .  $\varepsilon = 0,01$ ,  $\beta = 0,9$ .

2.  $F$  – закон с плотностью распределения  $f(x) = \frac{1}{(2+x)^2}$ ,  $x \in (-1, \infty)$ .  $\varepsilon = 0,02$ ,  $\beta = 0,999$ .

Вариант 7.

1.  $F$  – закон Бернулли с параметром  $p = 0,16$ .  $\varepsilon = 0,01$ ,  $\beta = 0,95$ .

2.  $F$  – закон с плотностью распределения  $f(x) = \frac{1}{5\pi} \cdot \frac{1}{1+(x/5)^2}$ ,  
 $x \in (-\infty, \infty)$ .  $\varepsilon = 0,03$ ,  $\beta = 0,9$ .

Вариант 8.

1.  $F$  – нормальный закон с параметрами  $a = 7$  и  $\sigma = 2$ .  $\varepsilon = 0,03$ ,  
 $\beta = 0,8$ .

2.  $F$  – закон с плотностью распределения  $f(x) = \frac{1}{(x-1)^2}$ ,  
 $x \in (2, \infty)$ .  $\varepsilon = 0,05$ ,  $\beta = 0,8$ .

Вариант 9.

1.  $F$  – биномиальный закон с параметрами  $n = 50$  и  $p = 0,35$ .  
 $\varepsilon = 0,1$ ,  $\beta = 0,95$ .

2.  $F$  – закон с плотностью распределения  $f(x) = \frac{1}{1+\pi^2 x^2}$ ,  
 $x \in (-\infty, \infty)$ .  $\varepsilon = 0,01$ ,  $\beta = 0,92$ .

Вариант 10.

1.  $F$  – закон равномерной плотности на  $(-3; 3)$ .  $\varepsilon = 0,04$ ,  
 $\beta = 0,92$ .

2.  $F$  – закон с плотностью распределения  $f(x) = \frac{1}{(x-2)^2}$ ,  
 $x \in (3, \infty)$ .  $\varepsilon = 0,002$ ,  $\beta = 0,9$ .

## ЛАБОРАТОРНАЯ РАБОТА 3

### Тема 3. Оценка закона распределения на основе выборочных данных

#### **Цель работы:**

Оценка закона распределения генеральной совокупности на основе выборочных данных.

#### **1. Необходимые теоретические сведения**

##### **1.1. Критерий $\chi^2$ (Пирсона) для простой гипотезы**

Пусть  $\{X_1, X_2, \dots, X_n\}$  выборка из генеральной совокупности  $F$ . Проверяется гипотеза  $H_0 : F = F_1$  против альтернативы  $H_1 : F \neq F_1$ .

Представим выборку в виде группированного ряда, разбив предполагаемую область значений случайной величины на  $m$  интервалов. Пусть  $n_i$  – число элементов выборки попавших в  $i$ -ый интервал, а  $p_i$  – теоретическая вероятность попадания в этот интервал

при условии истинности  $H_0$ . Составим статистику  $\rho(X) = \sum_{i=1}^m \frac{(n_i - np_i)^2}{np_i}$ ,

которая характеризует сумму квадратов отклонения наблюдаемых значений  $n_i$  от ожидаемых  $np_i$  по всем интервалам группирования.

**Теорема Пирсона.** Если  $H_0$  верна, то при фиксированном  $m$  и  $n \rightarrow \infty$

$$\rho(X) = \sum_{i=1}^m \frac{(n_i - np_i)^2}{np_i} \Rightarrow \chi_{m-1}^2. \quad (1)$$

Таким образом, статистику  $\rho(\vec{X})$  можно использовать в качестве статистики критерия согласия для проверки гипотезы о виде закона распределения, который будет иметь вид:

$$F(X) = \begin{cases} H_0, & \rho(X) < \tau_{1-\alpha} \\ H_1, & \rho(X) \geq \tau_{1-\alpha} \end{cases}, \text{ где } \rho(X) = \sum_{i=1}^m \frac{(n_i - np_i)^2}{np_i}, \quad (2)$$

где  $\tau_{1-\alpha}$  – квантиль распределения  $\chi_{m-1}^2$ .

Данный критерий называется критерием  $\chi^2$  или критерием согласия Пирсона.

**Замечание.** Критерий не состоятелен для альтернатив, для которых  $\tilde{p}_i = p_i$  для всех  $i \in \{1, 2, \dots, m\}$ . Поэтому, следует стремиться к как можно

большому числу интервалов группирования. Однако, с другой стороны, сходимость к  $\chi^2$  величины  $\frac{(n_i - np_i)^2}{np_i}$  обеспечивается ЦПТ, то есть ожидаемое значение  $np_i$  для каждой ячейки не должно быть слишком мало. Поэтому обычно число интервалов выбирают таким образом, чтобы  $np_i \geq 5$ .

## 1.2. Критерий $\chi^2$ (Пирсона) для сложной гипотезы

Пусть  $\{X_1, X_2, \dots, X_n\}$  выборка из генеральной совокупности  $F$ . Проверяется сложная гипотеза  $H_0 : F = F_\theta$ , где  $\theta$  – неизвестный параметр распределения  $F$  (или вектор параметров), против альтернативы  $H_1 : F \neq F_\theta$ .

Пусть выборка по прежнему представлена в виде группированного ряда и  $n_i$  – число элементов выборки попавших в  $i$ -ый интервал,  $i \in \{1, 2, \dots, m\}$ . Статистику (1) мы не можем в этом случае использовать для построения критерия Пирсона, так как не можем вычислить теоретические значения вероятностей  $p_i$ , которые зависят от неизвестного параметра  $\theta$ . Пусть  $\theta^*$  – оценка параметра  $\theta$ , а  $p_i^*(\theta^*)$  – соответствующие ей оценки вероятностей  $p_i$ . Составим статистику

$$\rho(X) = \sum_{i=1}^m \frac{(n_i - np_i^*)^2}{np_i^*}.$$

**Теорема Пирсона.** Если  $H_0$  верна, и  $l$  – число компонент вектора  $\theta$  (число неизвестных параметров распределения), то при фиксированном  $m$  и  $n \rightarrow \infty$

$$\rho(X) = \sum_{i=1}^m \frac{(n_i - np_i^*)^2}{np_i^*} \Rightarrow \chi_{m-l-1}^2. \quad (3)$$

Таким образом, критерий Пирсона для параметрической гипотезы будет иметь вид:

$$F(X) = \begin{cases} H_0, & \rho(X) < \tau_{1-\alpha} \\ H_1, & \rho(X) \geq \tau_{1-\alpha} \end{cases}, \quad \rho(X) = \sum_{i=1}^m \frac{(n_i - np_i^*)^2}{np_i^*}, \quad (4)$$

где  $\tau_{1-\alpha}$  – квантиль распределения  $\chi_{m-l-1}^2$ .

**Замечание.** Вообще говоря, оценки, используемые для построения статистики критерия хи-квадрат, должны быть определены из условия минимума статистики  $\rho(X)$ . Поэтому желательно уточнить оценки,

найденные другим способом (методом максимального правдоподобия или методом моментов) путем минимизации  $\rho(X)$ .

### 1.3. Метод моментов оценки параметров распределения

Идея этого метода заключается в приравнивании теоретических и эмпирических моментов.

Пусть  $X = (x_1, x_2, \dots, x_n)$  – независимая выборка из распределения  $P_\theta$ , зависящего от неизвестного параметра  $\theta = (\theta_1, \theta_2, \dots, \theta_k) \in \Theta \subset R^k$ . Моментом  $i$ -го порядка называется функция

$$\mu_i(\theta_1, \dots, \theta_k) = E[x^i] = \begin{cases} \int x^i f(x, \theta_1, \dots, \theta_k) dx, & \text{if } x \text{ continuous value} \\ \sum_j x_j^i p(x_j, \theta_1, \dots, \theta_k), & \text{if } x \text{ discrete value} \end{cases}$$

где  $f(x, \theta)$  – плотность распределения непрерывной случайной величины  $x$ ,  $p(x_j, \theta)$  – вероятность дискретной случайной величины. Теоретический момент является функцией неизвестных параметров  $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ .

Выборочным (эмпирическим) моментом  $i$ -го порядка называется величина.

$$m_i(x_1, x_2, \dots, x_n) = \frac{1}{n} \sum_{j=1}^n x_j^i.$$

Отметим, что по своему определению эмпирические моменты являются функциями от выборки.

Для нахождения неизвестных параметров (будем обозначать их  $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_k)$ ) составим систему уравнений

$$\begin{aligned} \mu_1(\hat{\theta}_1, \dots, \hat{\theta}_k) &= m_1, \\ \mu_2(\hat{\theta}_1, \dots, \hat{\theta}_k) &= m_2, \\ &\dots\dots\dots \\ \mu_k(\hat{\theta}_1, \dots, \hat{\theta}_k) &= m_k. \end{aligned}$$

Далее решаем систему относительно параметров  $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_k)$ . В результате получим.

$$\begin{aligned}\hat{\theta}_1 &= \hat{\theta}_1(x_1, \dots, x_n), \\ \hat{\theta}_2 &= \hat{\theta}_2(x_1, \dots, x_n), \\ &\dots\dots\dots, \\ \hat{\theta}_k &= \hat{\theta}_k(x_1, \dots, x_n).\end{aligned}$$

Найденные параметры зависят от выборки  $X = (x_1, x_2, \dots, x_n)$ .

**Пример 1.** Пусть  $X \in \Pi_\alpha$ , где  $\Pi_\alpha$  – показательный закон распределения с параметром  $\alpha$ . Найти оценку параметра  $\hat{\alpha}$ .

*Решение.*

$$\mu_1(\alpha) = E[x] = \int_0^\infty x \cdot \alpha e^{-\alpha x} dx = \frac{1}{\alpha}.$$

Приравниваем к  $m_1 = \frac{1}{n} \sum_{j=1}^n x_j$ . Отсюда получим: 
$$\hat{\alpha} = \frac{1}{m_1} = \frac{1}{\frac{1}{n} \sum_{j=1}^n x_j}.$$

**Пример 2.** Пусть  $X \in U_{\alpha, \beta}$ , где  $U_{\alpha, \beta}$  – равномерный закон распределения с параметрами  $\alpha, \beta$ . Найти оценки параметров  $\hat{\alpha}$  и  $\hat{\beta}$ .

*Решение.*

$$\mu_1(\alpha, \beta) = E[x] = \frac{1}{\beta - \alpha} \int_\alpha^\beta x dx = \frac{\alpha + \beta}{2},$$

$$\mu_2(\alpha, \beta) = E[x^2] = \frac{1}{\beta - \alpha} \int_\alpha^\beta x^2 dx = \frac{\beta^2 + \alpha\beta + \alpha^2}{3}.$$

Получим систему

$$\begin{aligned}\frac{\alpha + \beta}{2} &= m_1, \\ \frac{\beta^2 + \alpha\beta + \alpha^2}{3} &= m_2.\end{aligned}$$

Решение системы:  $\hat{\beta} = m_1 + \sqrt{3} \sqrt{m_2 - m_1^2} = m_1 + \sqrt{3} \sigma$ ,  $\hat{\alpha} = m_1 - \sqrt{3} \sigma$ .  
Здесь  $\sigma^2 = m_2 - m_1^2$  – дисперсия выборочного распределения.

#### 1.4. Метод максимального правдоподобия

Пусть  $X = (x_1, x_2, \dots, x_n)$  – независимая выборка из распределения  $P_\theta$ , зависящего от неизвестного параметра  $\theta = (\theta_1, \theta_2, \dots, \theta_k) \in \Theta \subset R^k$ .

Функцией правдоподобия  $L(\theta, x) = L(\theta_1, \dots, \theta_k; x_1, \dots, x_n)$  называют функцию.

$$L(\theta, x) = \begin{cases} \prod_{j=1}^n f(x_j, \theta_1, \dots, \theta_k), & \text{if } x \text{ continuous value} \\ \prod_{j=1}^n p(x_j, \theta_1, \dots, \theta_k), & \text{if } x \text{ discrete value} \end{cases}$$

В качестве оценки параметров  $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_k)$  примем значения этих параметров, при которых функция правдоподобия принимает максимальное значение, т.е.  $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_k) = \arg\left(\max_{\theta} L(\theta, x)\right)$ . Если функция  $L(\theta, x) = L(\theta_1, \dots, \theta_k; x_1, \dots, x_n)$  является дифференцируемой по переменным  $\theta_1, \theta_2, \dots, \theta_k$ , оценки параметров удовлетворяют системе уравнений:

$$\frac{dL(\theta_1, \dots, \theta_k; x)}{d\theta_i} = 0, \quad i = 1, 2, \dots, k.$$

**Пример 3.** Пусть  $X \in \Pi_{\alpha}$ , где  $\Pi_{\alpha}$  – показательный закон распределения с параметром  $\alpha$ . Найти оценку параметра  $\hat{\alpha}$  методом максимального правдоподобия.

*Решение.* Запишем функцию правдоподобия

$$L(\alpha, x) = \prod_{j=1}^n \alpha e^{-\alpha x_j} = \alpha^n e^{-\alpha \sum_{j=1}^n x_j}. \quad \text{Из условия максимума } \frac{dL(\alpha; x)}{d\alpha} = 0$$

получим следующее уравнение:  $n\alpha^{n-1} e^{-\alpha \sum_{j=1}^n x_j} - \alpha^n \sum_{j=1}^n x_j e^{-\alpha \sum_{j=1}^n x_j} = 0$ . Отсюда

следует  $\hat{\alpha} = \frac{1}{\frac{1}{n} \sum_{j=1}^n x_j}$ . Таким образом, эта оценка совпала с оценкой, по-

лученной методом моментов (см. пример 1).

**Пример 4.** Пусть  $X \in U_{\alpha, \beta}$ , где  $U_{\alpha, \beta}$  – равномерный закон распределения с параметрами  $\alpha, \beta$ . Найти оценки параметров  $\hat{\alpha}$  и  $\hat{\beta}$  методом максимального правдоподобия.

*Решение.*

Функция правдоподобия равна

$$L(\alpha, \beta) = \begin{cases} \prod_{j=1}^n \frac{1}{(\beta - \alpha)}, & \alpha \leq \tilde{x}_1 \leq \tilde{x}_2 \leq \dots \leq \tilde{x}_n \leq \beta \\ 0, & \tilde{X} \in [\alpha, \beta] \end{cases} \quad \text{или}$$

$$L(\alpha, \beta) = \begin{cases} \frac{1}{(\beta - \alpha)^n}, & \alpha \leq \tilde{x}_1 \leq \tilde{x}_2 \leq \dots \leq \tilde{x}_n \leq \beta \\ 0, & \tilde{X} \in [\alpha, \beta] \end{cases}, \quad \text{где } \tilde{X} = (\tilde{x}_1, \dots, \tilde{x}_n) -$$

упорядоченная выборка (вариационный ряд).

Проанализируем неравенства  $\alpha \leq \tilde{x}_1 \leq \tilde{x}_2 \leq \dots \leq \tilde{x}_n \leq \theta$ . Если взять  $\beta < \tilde{x}_n$ , т.е. меньше максимального значения выборки, то  $L(\alpha, \beta)$  обратится в ноль, и только при  $\beta = \tilde{x}_n$  функция  $L(\alpha, \beta)$  будет отлична от нуля и равна  $\frac{1}{(\beta - \alpha)^n}$ . Этот результат будет тем более верен, если  $\beta < \tilde{x}_i$ ,  $i < n$ . Поэтому получим  $\hat{\beta} = \tilde{x}_n$ . Проводя аналогичные рассуждения относительно левой границы интервала, получим  $\hat{\alpha} = \tilde{x}_1$ .

Таким образом, оценки параметров  $\hat{\alpha}, \hat{\beta}$ , полученные методом максимального правдоподобия, отличаются от оценок этих параметров, полученных методом моментов (см. пример 2).

### 1.5. Метод наименьших квадратов

Пусть дана табличная функция  $y(x)$

$x$	$x_1$	$x_2$		$x_n$
$y$	$y_1$	$y_2$		$y_n$

Необходимо аппроксимировать эти данные некоторой параметрической функцией  $f(\theta_1, \dots, \theta_k; x)$ , т.е. заменить функцию  $y(x)$  функцией  $f(\theta_1, \dots, \theta_k; x)$ :

$$y(x) \approx f(\theta_1, \dots, \theta_k; x).$$

Параметры  $\theta_1, \dots, \theta_k$  будем подбирать таким образом, расхождение табличной функции с функцией  $f(\theta_1, \dots, \theta_k; x)$  было минимальным. Для этого построим функционал  $F(\theta_1, \dots, \theta_k) = \sum_{j=1}^n (y_j - f(\theta_1, \dots, \theta_k; x_j))^2$  и найдем его минимум. Необходимое условие минимума имеет вид



$$\frac{dF(\theta_1, \dots, \theta_k)}{d\theta_i} = 0, \quad i = 1, \dots, k.$$

Решаем эту систему уравнений и получаем значения параметров  $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_k)$ .

**Пример 5.** Пусть дана независимая выборка из  $X = (x_1, x_2, \dots, x_n)$  из распределения  $P_\theta$ . Разобьем весь диапазон данных  $[x_{\min}, x_{\max}]$  на  $m$  интервалов и построим гистограмму. Обозначим середины интервалов  $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_m$ . Тогда относительные частоты (высоты столбиков гистограммы) будут значениями табличной функции  $y$ . Таким образом, мы получили табличную функцию

$\tilde{x}$	$\tilde{x}_1$	$\tilde{x}_2$		$\tilde{x}_n$
$y$	$y_1$	$y_2$		$y_n$

Поставим следующую задачу. Подобрать параметры известного закона непрерывного распределения  $f(\theta_1, \dots, \theta_k; x)$  так, чтобы расхождение между гистограммой и функцией  $f(\theta, x)$  было минимально. В результате мы приходим к методу наименьших квадратов.

## 2. Практическое задание

Имеется выборка объемом  $n = 100$  из неизвестного распределения  $F$  (см. приложение 2). Предполагается, что  $F$  может быть одним из следующих распределений:

1)  $F$  – нормальное распределение с плотностью

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad x \in R, \text{ где параметры } a \text{ и } \sigma > 0 \text{ – неизвестны;}$$

2)  $F$  – распределение Лапласа с плотностью  $f(x) = \frac{1}{\sigma\sqrt{2}} e^{-\frac{|x-a|\sqrt{2}}{\sigma}}$ ,

$x \in R$ , где параметры  $a$  и  $\sigma > 0$  – неизвестны.

Справка.

Моменты:  $m_1 = \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2}} e^{-\frac{|x-a|\sqrt{2}}{\sigma}} dx = a,$

$$m_2 = \int_{-\infty}^{\infty} x^2 \frac{1}{\sigma\sqrt{2}} e^{-\frac{|x-a|\sqrt{2}}{\sigma}} dx = a^2 + \sigma^2,$$

$$D = m_2 - m_1^2 = \sigma^2.$$

$$\text{Коэффициент асимметрии } A = \frac{1}{\sigma^3} M(x-a)^3 = 0,$$

$$\text{коэффициент эксцесса } E = \frac{1}{\sigma^4} M(x-a)^4 - 3 = 3.$$

3)  $F$  – распределение Рэлея с плотностью  $f(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}$ ,  $x > 0$ , где параметр  $\sigma > 0$  – неизвестен;

4)  $F$  – показательное распределение с плотностью  $f(x) = \alpha e^{-\alpha x}$ ,  $x > 0$ , где параметр  $\alpha > 0$  – неизвестен;

5)  $F$  – равномерное распределение на отрезке  $[a, b]$ , где параметры  $a$  и  $b$  – неизвестны.

$$6) F \text{ – распределение с плотностью } f(x) = \frac{2a}{(1+ax)^3}, \quad x \in [0, \infty),$$

где параметр  $a > 0$  неизвестен.

7)  $F$  – распределение Коши с плотностью  $f(x) = \frac{1}{\pi \cdot \theta} \cdot \frac{1}{1 + ((x-a)/\theta)^2}$ ,  $x \in R$ , где параметры  $a$  и  $\theta > 0$  – неизвестны;

### Требуется:

1) Представить выборку в виде группированного статистического ряда. При разбивке на интервалы следует следить за тем, чтобы частоты  $n_i$  для всех интервалов были одного порядка, причем количество выборочных значений  $n_i$  попавших в каждый интервал должно быть не меньше 5 ( $n_i \geq 5 \quad \forall i = \overline{1, m}$ ). В противном случае следует объединять соседние интервалы, добиваясь относительно равномерного распределения частот по интервалам.

2) Найти числовые характеристики выборки: выборочное среднее  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ , выборочную дисперсию  $s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ , выборочный коэффициент асимметрии  $\bar{A} = \frac{1}{(n-1)s^3} \sum_{i=1}^n (X_i - \bar{X})^3$  и выборочный коэффициент эксцесса  $\bar{E} = \frac{1}{(n-1)s^4} \sum_{i=1}^n (X_i - \bar{X})^4 - 3$ .

3) Построить гистограмму и сравнить ее с кривыми плотности возможных теоретических распределений. При построении кривых

плотностей неизвестные параметры теоретических распределений можно заменить оценками, найденными по методу максимального правдоподобия или по методу моментов.

4) Выдвинуть гипотезу  $H_0$  о виде закона распределения (на основе сравнения гистограммы с графиком плотности теоретического распределения). Дополнительно для выбора можно использовать сравнение выборочных значений  $\bar{A}$  и  $\bar{E}$  с теоретическими:  $A = \frac{\mu_3}{\sigma^3}$ ,

$E = \frac{\mu_4}{\sigma^4} - 3$ , где  $\mu_k$  – центральный момент  $k$ -го порядка,  $\sigma$  – среднеквадратичное отклонение (теоретические значения  $A$  и  $E$  для каждого из возможных распределений предварительно подсчитать).

5) Используя критерий Пирсона, проверить гипотезу  $H_0$  для уровня значимости  $\alpha = 0,01 * [(N + 1) / 2]$ , где  $N$  – номер варианта задания. Если гипотеза отвергается, следует выдвинуть другую и аналогично подвергнуть ее проверке.

6) Для принятой гипотезы уточнить значение оценок параметров распределения, используя метод наименьших квадратов (определяем оценки, исходя из минимума статистики критерия Пирсона

$$\rho(\bar{X}) = \sum_{i=1}^m \frac{(n_i - np_i^*)^2}{np_i^*}$$

7) Найти реально достигнутый уровень значимости  $\alpha_0$ , то есть вероятность того, что при истинности гипотезы  $H_0$  значение статистики  $\rho(\bar{X})$  будет больше наблюдаемого значения статистики  $\rho_{набл}$ :  $P(\rho > \rho_{набл}) = \alpha_0$ .

## Приложение 1

### Пример выполнения задания

*Дана выборка объемом  $n = 100$  из неизвестного распределения  $F$ :*

12,35	3,72	2,96	4,57	1,01	7,08	10,24	1,66	3,87	20,32
0,89	3,81	0,95	7,62	23,47	2,56	19,15	0,00	8,38	0,67
15,56	0,30	4,83	10,32	13,91	12,25	11,34	13,68	9,42	3,12
11,49	4,11	41,99	15,87	18,78	6,99	17,25	4,27	16,06	5,78
15,69	10,35	11,22	4,24	36,67	22,37	0,22	8,61	16,05	18,06
0,21	20,82	12,36	10,04	2,36	0,44	7,36	10,99	6,36	22,54
28,67	18,77	5,95	17,91	2,59	7,76	1,92	21,88	5,54	11,59
0,71	8,41	2,18	2,50	0,65	4,67	11,17	4,76	0,60	3,39
3,88	39,28	5,70	1,46	3,25	3,57	17,85	1,18	1,34	9,14

25,64    4,07    8,95    25,71    4,94    15,65    1,25    3,01    18,10    6,52

Построим статистический ряд, осуществив группировку данных. Находим  $x_{\min} = 0,04$ ,  $x_{\max} = 26,52$ . Число интервалов группирования  $m$  определяем по формуле Стерджесса:  $m = 1 + [\log_2 n] = 7$ . Для удобства возьмем в качестве нижней границы первого интервала значение  $\tilde{x}_0 = 0$ , а в качестве верхней границы последнего интервала значение  $\tilde{x}_7 = 42$ , тогда длина каждого интервала группирования будет равна  $\Delta x = 42/7 = 6$ . Подсчитывая частоты, получаем следующий ряд:

Интервал $\tilde{x}_{i-1} - \tilde{x}_i$	0 – 6	6 – 12	12 – 18	18 – 24	24 – 30	30 – 36	36 – 42
Частота $n_i$	46	23	14	11	3	0	3

Видим, что частоты распределены по интервалам крайне неравномерно, поэтому делаем перегруппировку данных, добиваясь более равномерного распределения частот по интервалам. В результате получаем следующий ряд:

Интервал $\tilde{x}_{i-1} - \tilde{x}_i$	0 – 4	4 – 8	8 – 12	12 – 16	16 – 20	20 – 24	24 – 42
Середина $\bar{x}_i$	2	6	10	12	18	22	33
Частота $n_i$	33	20	16	9	10	6	6
Относительная частота $\omega_i$	0,338	0,20	0,16	0,09	0,10	0,06	0,06
Плотность частоты $\rho_i$	0,0825	0,0500	0,0400	0,0225	0,025	0,0150	0,00333

*Соответствующая гистограмма приведена на рисунке 3.1.*

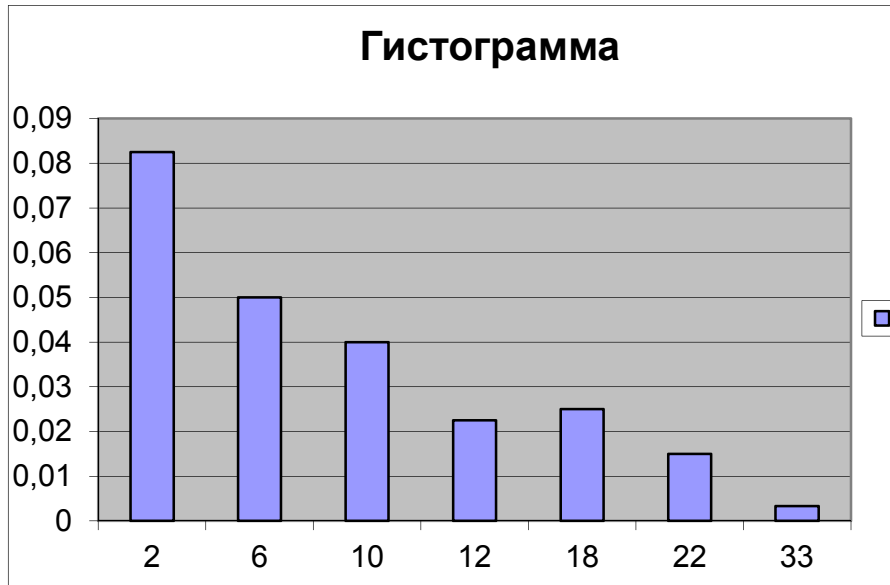


Рис. 3.1 Гистограмма ряда

Анализируя гистограмму, видим, что распределение экспериментальных данных похоже на показательное распределение. Таким образом, выдвигаем гипотезу о том, что выборочные данные имеют показательное распределение. В качестве оценки неизвестного параметра  $\alpha$  этого распределения возьмем оценку, полученную по методу моментов (через первый момент). Так как  $m_1 = 1/\lambda$ , то  $\lambda^* = 1/\bar{X} = 1/9,72 \approx 0,103$ . Вычислим значения плотности показательного распределения  $f(x) = \lambda^* e^{-\lambda^* x}$  в точках, соответствующих серединам интервалов группирования, и сравним гистограмму с графиком плотности:

Интервал $\tilde{x}_{i-1} - \tilde{x}_i$	0 – 4	4 – 8	8 – 12	12 – 16	16 – 20	20 – 24	24 – 42
Середина $\bar{x}_i$	2	6	10	12	18	22	33
Плотность частоты $\rho_i$	0,0825	0,0500	0,0400	0,0225	0,025	0,0150	0,00333
Теоретическая плотность	0,08377	0,0555	0,03677	0,02993	0,01614	0,01069	0,0034

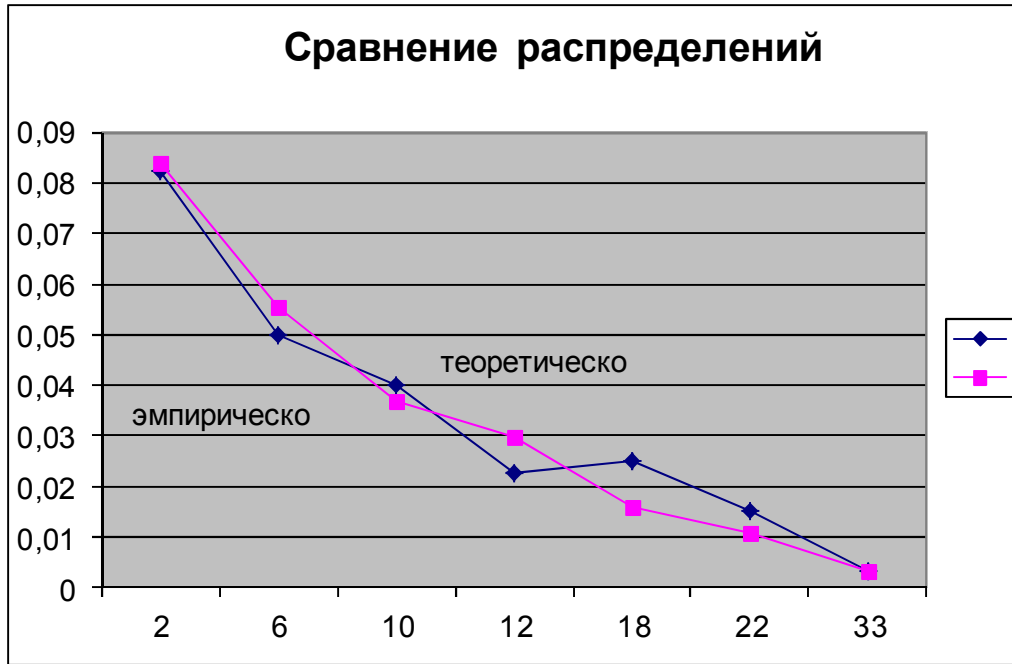


Рис. 3.2. График сравнения определений

Найдем также оценки коэффициента асимметрии и эксцесса распределения и сравним их с коэффициентом асимметрии и эксцессом показательного распределения.

Выборочный коэффициент асимметрии

$$\bar{A} = \frac{1}{(n-1)s^3} \sum_{i=1}^n (X_i - \bar{X})^3 \approx 1,4$$

Выборочный эксцесс  $\bar{E} = \frac{1}{(n-1)s^4} \sum_{i=1}^n (X_i - \bar{X})^4 - 3 \approx 2,15$ .

Коэффициент асимметрии для показательного распределения  $A = 2$ , эксцесс  $E = 6$ . Хотя различие есть, тем не менее можно утверждать, что данные выборочные характеристики имеют смещение (относительно нуля) в сторону характеристик показательного закона.

Применим критерий Пирсона для проверки нашей гипотезы о законе распределения выборочных данных. Подсчитаем вероятности  $p_i^*$  попадания в каждый интервал при условии, что генеральная совокупность имеет показательное распределение с параметром  $\lambda^* = 0,103$ :  $p_i^* = F(\tilde{x}_i) - F(\tilde{x}_{i-1})$ ,  $i = 1, 7$ ,  $F(x) = 1 - e^{-\lambda^* x}$  — функция распределения показательного закона. Причем для последнего интервала, полагаем  $\tilde{x}_7 = \infty$  и, соответственно,  $F(\tilde{x}_7) = 1$ , поскольку теоретически для показательного распределения плотность отлична от нуля на интервале  $(0, \infty)$ . Далее находим ожидаемые значения —  $np_i^*$  и

нормированные квадраты отклонений  $(np_i^* - n_i)^2 / np_i^*$  по всем интервалам. Результаты оформляем в виде таблицы:

Интервал	0 – 4	4 – 8	8 – 12	12 – 16	16 – 20	20 – 24	24 – 42
Частота $n_i$	33	20	16	9	10	6	6
Вероятность $p_i^*$	0,3374	0,2236	0,1481	0,0981	0,0650	0,0431	0,0133
Ожидаемые значения $np_i^*$	33,74	22,36	14,81	9,81	6,502	4,31	8,46
$(n_i - np_i^*)^2 / np_i^*$	0,0164	0,248	0,0951	0,06758	1,8811	0,6643	0,7147

Находим наблюдаемое значение статистики критерия Пирсона:

$$\rho_{\text{набл}} = \sum_{i=1}^7 \frac{(n_i - np_i^*)^2}{np_i^*} \approx 3,15.$$

Зададим уровень значимости  $\alpha = 0,05$ . Для заданного уровня значимости и числа степеней свободы  $k = m - l - 1 = 7 - 1 - 1 = 5$  ( $l = 1$  – так как один параметр распределения  $\lambda$  мы оценивали по выборке) найдем критическое значение статистики, как критическую точку распределения  $\chi_5^2$  уровня  $\alpha = 0,05$  (или что тоже самое – квантиль уровня 0,95):  $\rho_{\text{кр}} = 11,07$ . (Критическую точку заданного уровня можно получить, например, используя функцию пакета EXCEL – ХИ2ОБР).

Так как  $\rho_{\text{набл}} < \rho_{\text{кр}}$ , то гипотеза о распределении данных по показательному закону принимается.

Уточним значение  $\lambda^*$ , минимизируя наблюдаемое значение статистики  $\rho_{\text{набл}}$ . Используя последовательные итерации, можно получить оценку  $\lambda_1^* = 0,101$ , при которой наблюдаемое значение  $\rho_{\text{набл}} \approx 3$ . Видим, что в данном случае эта оценка практически не отличается от оценки метода моментов.

## Приложение 2 Варианты заданий

Вариант 1

-12,97	8,18	5,19	2,54	8,40	3,26	1,31	6,33	-0,25	2,74
-2,78	7,41	0,55	-8,43	12,40	2,55	8,70	9,26	-2,15	7,97
-0,10	8,96	2,98	-1,88	1,43	4,85	-2,54	-5,23	8,60	-1,71
-1,25	-0,51	7,77	-3,47	3,24	2,85	7,10	-1,86	4,06	2,98
7,20	-1,48	-6,20	-0,54	-4,02	4,01	2,65	2,27	1,91	8,36
-0,25	6,03	3,51	3,10	1,73	1,11	-3,05	0,41	3,39	1,49
-5,13	5,43	2,41	-5,07	0,17	3,02	6,37	4,56	-10,89	3,61

-1,45	-6,24	6,71	8,44	-8,09	1,38	6,81	-5,66	3,37	4,16
4,47	4,26	-4,15	2,01	-3,46	3,20	-1,90	2,85	5,50	-0,76
0,83	-0,24	-8,96	-0,09	2,57	8,09	11,05	8,75	2,73	3,50

Вариант 2

0,41	0,33	0,01	0,34	0,55	0,89	0,96	0,91	0,56	0,62
-18,85	7,41	4,29	2,32	7,68	2,79	1,59	5,36	0,50	2,44
-1,82	6,51	1,09	-9,66	13,61	2,33	8,06	8,79	-1,18	7,16
0,60	8,40	2,59	-0,92	1,66	3,99	-1,57	-4,76	7,94	-0,76
-0,34	0,29	6,93	-2,58	2,77	2,51	6,17	-0,90	3,36	2,59
6,27	-0,55	-6,11	0,26	-3,21	3,33	2,38	2,15	1,95	7,63
0,49	5,07	2,96	2,68	1,85	1,46	-2,11	0,99	2,87	1,70
-4,62	4,51	2,24	-4,53	0,81	2,63	5,40	3,76	-14,30	3,03
-0,52	-6,16	5,75	7,73	-9,08	1,63	5,86	-5,33	2,86	3,44
3,68	3,52	-3,38	2,00	-2,57	2,74	-0,94	2,51	4,57	0,08
1,28	0,50	-10,59	0,61	2,33	7,31	11,38	8,12	2,44	2,95

Вариант 3

0,10	4,22	3,28	2,50	4,29	2,71	2,17	3,63	1,78	2,56
1,22	3,97	1,98	0,39	5,64	2,51	4,39	4,57	1,35	4,15
1,81	4,47	2,62	1,41	2,20	3,17	1,27	0,78	4,36	1,44
1,54	1,72	4,09	1,09	2,70	2,59	3,87	1,41	2,94	2,63
3,90	1,49	0,64	1,71	0,99	2,92	2,53	2,43	2,33	4,28
1,78	3,53	2,78	2,66	2,28	2,12	1,17	1,94	2,74	2,22
0,80	3,35	2,47	0,81	1,88	2,64	3,64	3,09	0,20	2,81
1,50	0,64	3,75	4,30	0,42	2,19	3,78	0,72	2,74	2,97
3,06	3,00	0,96	2,36	1,09	2,69	1,40	2,59	3,37	1,66
2,05	1,78	0,34	1,82	2,51	4,19	5,17	4,40	2,56	2,77

Вариант 4

0,10	0,02	0,18	0,46	0,43	0,64	0,00	0,10	0,40	0,03
0,06	0,01	0,01	0,04	0,05	0,00	0,07	0,08	0,16	0,09
0,09	0,09	0,48	0,13	0,11	0,07	0,74	0,33	0,95	0,06
0,61	0,01	0,24	0,34	0,72	0,13	0,07	0,28	0,09	0,30
0,02	0,04	0,01	0,09	0,13	0,14	0,09	0,85	0,01	0,05
0,00	0,52	0,02	0,06	0,30	0,23	0,33	0,26	0,02	0,03
0,28	0,20	0,04	0,10	0,16	0,25	0,16	0,04	0,70	0,23
0,15	0,32	0,33	0,06	0,04	0,40	0,02	0,01	0,29	0,27
0,86	0,52	0,47	0,16	0,14	0,22	0,13	0,03	0,01	0,32
0,22	0,26	0,18	0,03	0,45	0,09	0,04	0,05	0,08	0,08



Вариант 5

2,00	4,67	4,22	3,63	4,70	3,80	3,34	4,42	2,98	3,68
2,51	4,58	3,16	2,06	4,94	3,63	4,73	4,78	2,61	4,65
3,01	4,75	3,73	2,66	3,36	4,15	2,55	2,22	4,72	2,69
2,77	2,92	4,63	2,41	3,79	3,70	4,54	2,66	3,98	3,73
4,55	2,73	2,15	2,92	2,34	3,97	3,65	3,56	3,48	4,70
2,98	4,37	3,86	3,76	3,44	3,29	2,47	3,13	3,83	3,38
2,23	4,26	3,60	2,24	3,07	3,74	4,43	4,09	2,01	3,88
2,73	2,15	4,48	4,70	2,07	3,35	4,50	2,19	3,82	4,00
4,07	4,02	2,33	3,50	2,41	3,78	2,65	3,70	4,27	2,87
3,22	2,98	2,04	3,01	3,64	4,67	4,89	4,73	3,67	3,85

Вариант 6

0,00	-0,41	-0,19	-0,10	-0,43	-0,12	-0,07	-0,25	-0,04	-0,10
-0,02	-0,34	-0,06	0,00	-1,26	-0,10	-0,47	-0,54	-0,02	-0,39
-0,05	-0,50	-0,11	-0,03	-0,07	-0,17	-0,02	-0,01	-0,45	-0,03
-0,03	-0,04	-0,37	-0,02	-0,12	-0,10	-0,31	-0,03	-0,14	-0,11
-0,32	-0,03	-0,01	-0,04	-0,01	-0,14	-0,10	-0,09	-0,08	-0,43
-0,04	-0,24	-0,12	-0,11	-0,08	-0,06	-0,02	-0,05	-0,12	-0,07
-0,01	-0,20	-0,09	-0,01	-0,05	-0,11	-0,26	-0,16	0,00	-0,13
-0,03	-0,01	-0,28	-0,44	0,00	-0,07	-0,29	-0,01	-0,12	-0,15
-0,16	-0,15	-0,01	-0,08	-0,02	-0,11	-0,03	-0,10	-0,21	-0,04
-0,06	-0,04	0,00	-0,05	-0,10	-0,40	-0,87	-0,47	-0,10	-0,12

Вариант 7

-1156,93	16,11	6,65	2,68	17,36	3,62	1,13	9,19	-1,03	2,94
-6,48	12,65	0,13	-83,96	86,84	2,70	19,18	23,34	-4,74	15,09
-0,81	21,01	3,24	-4,09	1,28	6,01	-5,77	-19,13	18,57	-3,70
-2,76	-1,44	14,17	-8,89	3,59	3,07	11,52	-4,04	4,74	3,24
11,86	-3,21	-29,27	-1,50	-11,31	4,68	2,82	2,33	1,89	17,13
-1,03	8,44	3,96	3,40	1,67	0,87	-7,36	-0,06	3,78	1,36
-18,32	7,12	2,51	-17,79	-0,41	3,30	9,31	5,53	-317,92	4,10
-3,16	-29,75	10,27	17,55	-71,03	1,21	10,58	-23,01	3,76	4,90
5,38	5,05	-11,99	2,01	-8,85	3,54	-4,14	3,07	7,26	-1,87
0,51	-1,02	-110,08	-0,79	2,71	15,69	47,01	19,52	2,92	3,94

Вариант 8

14,78	5,88	2,03	8,86	13,16	7,44	7,81	11,34	10,38	13,17
7	12,44	10,86	12,15	6,44	13,43	7,94	4,8	9,55	4,5

5,01	11,31	14,26	6,34	10,13	7,65	12,01	5,19	5,69	17,97
7,3	7,01	16,28	12	14,82	2,9	7,18	9,31	14,23	12,54
2,38	9,04	1,73	17,93	3,03	4,48	14,27	14,62	2,16	13,29
15,87	5,12	11,18	12,15	13,23	10,37	13,4	16,85	5,65	17,32
16,93	17,25	6,81	13,77	3,16	10,81	10,26	9,36	10,68	4,88
6,24	18,48	3,15	6,71	13,4	22,71	4,75	4,97	9,08	11,09
12,92	5,39	6,64	9,39	10,7	5,93	11,35	3,92	12,22	4,89
4,18	4,75	10,97	24,92	6,85	6,94	7,21	13,25	14,35	1,93

Вариант 9

3,36	0,88	0,62	0,65	3,95	2,19	8,10	2,54	-5,06	0,85
6,56	0,63	0,40	0,37	0,73	0,16	0,10	0,43	1,32	0,94
-1,47	2,59	-2,96	-2,23	2,51	0,21	-3,62	2,24	2,12	3,59
0,42	5,47	2,53	1,36	0,62	-1,20	-1,56	1,68	4,17	3,66
1,92	6,12	1,54	0,00	2,74	1,69	0,74	-0,42	-0,43	1,75
0,56	0,63	-0,92	-0,62	3,45	0,98	2,67	1,48	-0,29	-0,55
1,41	0,10	0,85	-0,64	-0,31	2,28	2,11	0,76	1,02	0,28
2,59	2,80	-3,72	3,51	2,83	-1,26	0,29	-0,47	0,65	-2,78
0,91	0,95	-2,47	-3,63	-1,62	-2,04	0,70	0,86	-0,60	2,99
1,18	1,22	0,31	-0,42	4,48	1,00	1,50	-1,42	1,34	0,53

Вариант 10

0,98	0,75	0,96	2,46	0,39	0,10	1,68	1,23	2,93	1,49
0,46	0,31	0,94	0,33	0,67	2,28	1,40	1,78	2,28	1,62
1,35	0,26	2,20	1,56	0,36	0,46	0,62	1,81	2,16	1,33
2,14	0,60	1,12	1,84	0,94	0,77	1,02	0,78	1,25	0,57
2,08	0,70	1,05	2,68	1,16	1,73	1,27	0,63	1,18	2,11
2,52	0,19	0,85	1,12	1,20	1,51	1,58	2,01	1,50	0,97
0,17	0,82	2,73	0,84	0,97	1,82	1,27	0,49	1,16	0,67
1,02	0,92	1,81	1,67	2,92	1,61	0,63	0,35	0,56	0,68
1,99	1,27	3,08	0,93	0,29	0,42	2,65	1,69	0,28	0,90
0,55	0,94	0,77	1,73	0,72	0,53	1,52	1,64	1,78	0,89

# ЛАБОРАТОРНАЯ РАБОТА 4

## Тема 4. Линейная регрессия

### **Цель работы:**

Оценка уравнения линейной регрессии на основе выборочных данных

### **1. Необходимые сведения из теории**

#### **1.1. Построение модели парной регрессии**

Рассмотрим линейную по коэффициентам модель парной регрессии:

$$y = f(x) + \varepsilon = \beta_0 + \beta_1 f_1(x) + \beta_2 f_2(x) + \dots + \beta_k f_k(x) + \varepsilon, \quad (1)$$

где  $\varepsilon$  – случайная величина с математическим ожиданием равным нулю и дисперсией  $\sigma^2$ .

Полагая,  $x_j = f_j(x)$ ,  $j = \overline{1, k}$  перейдем к модели множественной линейной регрессии:

$$y = f(x) + \varepsilon = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon. \quad (2)$$

Пусть для оценки неизвестных параметров  $\beta_j$ ,  $j = \overline{0, k}$  уравнения регрессии (2) взята выборка объемом  $n$  из значений величин  $(Y, X_1, X_2, \dots, X_k)$ . Тогда

$$Y = XB + \varepsilon,$$

где  $Y = (y_1, y_2, \dots, y_n)^T$  – вектор значений переменной  $y$ ;

$B = (\beta_0, \beta_1, \dots, \beta_k)^T$  – вектор параметров модели;

$\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$  – вектор ошибок, где  $\varepsilon_i \in N(0, \sigma^2)$  и независимы;

$X$  – матрица исходных данных переменных  $X_j$  размерами  $n \times (k+1)$ . Первый столбец матрицы  $X$  содержит единицы (значения фиктивной переменной  $x_0$ ), остальные столбцы значения переменных  $x_1, x_2, \dots, x_k$ :

$$X = \begin{pmatrix} 1 & x_1^1 & \dots & x_k^1 \\ 1 & x_1^2 & \dots & x_k^2 \\ & & \dots & \\ 1 & x_1^n & \dots & x_k^n \end{pmatrix}.$$

Для нахождения оценки  $B^*$  вектора параметров  $B = (\beta_0, \beta_1, \dots, \beta_k)^T$  используем метод наименьших квадратов, согласно которому в качестве оценок  $\beta_0^*, \beta_1^*, \dots, \beta_k^*$  берутся такие, которые минимизируют сумму квадратов  $Q$  отклонений значений  $y_i$  от  $f(\bar{x}_i)$ :

$$Q = \sum_{i=1}^n (y_i - f(\bar{x}_i))^2 = \sum_{i=1}^n \varepsilon_i^2 = \varepsilon^T \varepsilon = (Y - XB)^T (Y - XB). \quad (3)$$

Оценка  $B^*$  метода наименьших квадратов имеет вид:

$$B^* = (X^T X)^{-1} X^T Y. \quad (4)$$

## 1.2. Оценка погрешности регрессии

Качество регрессионной модели можно оценить, используя оценку  $s^2$  дисперсии предсказания  $\sigma^2$ :

$$s^2 = \frac{1}{n-k-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \frac{1}{n-k-1} \sum_{i=1}^n e^2 = \frac{1}{n-k-1} e^T e, \text{ где}$$

$\hat{y}_i = \beta_0^* + \beta_1^* x_i + \dots + \beta_k^* x_k$ . Качество модели также можно оценить с исполь-

зованием оценки коэффициента детерминации:  $R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$ .

Чем ближе значения  $R^2$  к 1, тем большую долю дисперсии величины  $Y$  объясняет модель регрессии.

Оценка дисперсии коэффициента  $\beta_j$  находится по формуле:  $s_j^2 = s^2 [(X^T X)^{-1}]_{jj}$ , где  $[(X^T X)^{-1}]_{jj}$  соответствующий диагональный элемент матрицы  $(X^T X)^{-1}$ .

Доверительные интервал для  $\sigma^2$  находится с использованием статистики  $\chi^2 = (n-k-1)s^2 / \sigma^2$ , которая при нормальном распределении  $\varepsilon_i$  имеет распределение хи-квадрат с  $n-k-1$  степенью свободы.

Для проверки значимости коэффициентов уравнения регрессии используем статистику  $t_j = \frac{\beta_j^*}{\sqrt{s^2 [(X^T X)^{-1}]_{jj}}}$ , которая при истинности ги-

потезы  $H_0: \beta_j = 0$ , имеет распределение Стьюдента с  $n-k-1$  степенью свободы. Если для заданного уровня значимости  $\alpha$  значение  $|t_j|$  больше

критического  $t_{крит} = t_{1-\alpha/2}$ , то нулевая гипотеза отвергается и коэффициент признается значимым. В противном случае коэффициент признается незначимым, и соответствующее слагаемое исключается из модели.

В пакете Excel рассчитывается также уровень значимости  $\alpha$  статистики  $|t_j|$ , т.е. вероятность  $P(x > |t_j|)$ . Степень значимости параметров распределения качественно определяется по уровню значимости: незначимые ( $\alpha \geq 0,100$ ), слабо значимые ( $0,100 > \alpha \geq 0,050$ ), статистически значимые ( $0,050 > \alpha \geq 0,010$ ), сильно значимые ( $0,010 > \alpha \geq 0,001$ ), высоко значимые ( $0,001 > \alpha$ ).

Для нахождения доверительных интервалов для коэффициентов  $\beta_j$  используют статистики  $\tilde{t}_j = \frac{\beta_j^* - \beta_j}{\sqrt{s^2 [(X^T X)^{-1}]_{jj}}}$ , имеющие распределение

Стьюдента с  $n-k-1$  степенью свободы. Для уровня значимости  $\alpha$  доверительный интервал рассчитывается по формуле  $\beta_j^* \pm t_\alpha \sqrt{s^2 [(X^T X)^{-1}]_{jj}}$ , где  $t_\alpha$  – квантиль распределение Стьюдента с  $n-k-1$  степенью свободы.

Доверительный интервал для значений  $y$  в точке  $X_0 = (1, x_1^0, \dots, x_k^0)^T$  будет иметь вид:  $\left[ X_0^T B^* \pm t_{1-\alpha/2} s \left( 1 + \sqrt{X_0^T [(X^T X)^{-1}] X_0} \right) \right]$ , так как погрешность  $y = f(x) + \varepsilon$  будет определяться двумя источниками: погрешностью  $(\Delta f)^2 = s^2 \left( X_0^T [(X^T X)^{-1}] X_0 \right)$ , связанной с погрешностями параметров модели, и погрешностью собственно модели  $\varepsilon^2 = s^2$ . Здесь  $t_\alpha$  – квантиль распределение Стьюдента с  $n-k-1$  степенью свободы.

## 2. Пример выполнения задания

Имеется выборка значений совместно наблюдаемых величин  $X$  и

$Y$ :

X	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5
Y	2,96	0,61	4,63	2,44	2,23	4,89	4,98	3,89	6,74	8,07
X	5,5	6	6,5	7	7,5	8	8,5	9	9,5	10
Y	8,34	9,56	9,30	12,35	11,46	11,09	7,91	8,16	6,54	7,88

Требуется подобрать подходящую модель регрессии, характеризующую зависимость  $Y$  от  $X$ , если известно, что ошибка  $\sigma^2 = 1,3$ .

Нанесем точки  $(X, Y)$  на координатную плоскость – построим корреляционное поле, соответствующее нашей выборке (рис. 4.1)

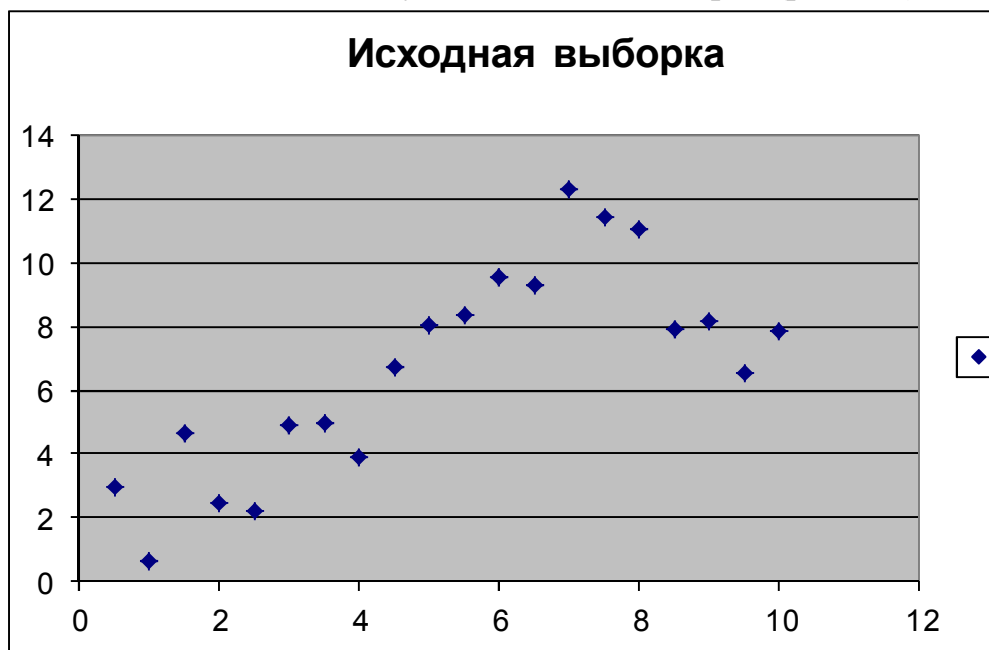


Рис. 4.1. Исходные данные

Видим, что существует зависимость, между значениями  $X$  и  $Y$ , причем зависимость явно нелинейная. Попробуем аппроксимировать эту зависимость для начала полиномами различных порядков. Возьмем в качестве уравнения регрессии квадратное уравнение:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2$$

Чтобы воспользоваться МНК для оценки коэффициентов, проведем линеаризацию модели, положив  $x_1 = x$ ,  $x_2 = x^2$ , получим

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Тогда оценку вектора параметров, согласно МНК, найдем как

$$B^* = (X^T X)^{-1} X^T Y$$

Здесь  $X$  – матрица, первый столбец которой содержит единицы, а второй и последующий значения  $x_1$  и  $x_2$ .

Для облегчения подбора модели можно воспользоваться встроенными функциями пакета EXCEL (для выбранной модели все равно потом потребуется провести все вычисления вручную, чтобы построить доверительные интервалы). В пакете анализа необходимо выбрать функцию “регрессия”, задать столбец значений  $Y$  и матрицу, соответ-

вующую  $X$  (единичный столбец в этом случае задавать не надо). Если выбрать вывод остатков, то помимо регрессионной статистики, будут выведены и предсказанные значения  $Y$ , т.е.

$$y^* = \beta_0^* + \beta_1^* x + \beta_2^* x^2$$

Для нашей модели регрессионная статистика, полученная пакетом Fxcel будет иметь следующий вид (приведена лишь часть):

ВЫВОД ИТО-  
ГОВ

<i>Регрессионная статистика</i>	
Множественный	
R	0,852395595
R-квадрат	0,72657825
Нормированный	
R-квадрат	0,694410986
Стандартная ошибка	1,820197731
Наблюдения	20

Дисперсионный анализ

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>
Регрессия	2	149,6702188	74,83510939	22,58750495	1,63336E-05
Остаток	17	56,32303623	3,313119778		
Итого	19	205,993255			

	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>
Y-пересечение	-0,963008772	1,354193804	-0,711130688	0,486646844	-3,82010793	1,89409
Переменная X 1	2,604789474	0,593991366	4,385231207	0,000403783	1,351577249	3,8580
Переменная X 2	-0,16754386	0,054949757	-3,049037334	0,007254125	-0,28347771	-0,05

Здесь в первой таблице:

- Множественный R – корень квадратный из коэффициента детерминации;

- R-квадрат – коэффициент детерминации,  $R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$ ;

- Нормированный R-квадрат – это скорректированная величина коэффициента детерминации, вычисляемая по формуле  $\hat{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-k-1}$  (здесь  $n$  – объем выборки,  $n = 20$ ;  $k$  – количество переменных модели,  $k = 2$ );
- Стандартная ошибка – значение  $s = \sqrt{s^2}$ , где  $s^2$  – оценка дисперсии предсказания  $\sigma^2$ ,  $s^2 = \frac{1}{n-k-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2$ ;
- Наблюдения – объем выборки,

Во второй таблице:

- $df$  – степени свободы;
- $SS$  – сумма квадратов. В первой строке для регрессии

$$SS = \sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2, \quad \text{во второй строке для остатка}$$

$$SS = \sum_{i=1}^n (y_i - \hat{y}_i)^2;$$

- $MS$  – средние квадраты. В первой строке – для регрессии

$$MS = \frac{1}{k} \sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2, \quad \text{во второй строке – для остатка}$$

$$MS = \frac{1}{n-k-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad \text{т.е. } MS = s^2;$$

- $F = \frac{\frac{1}{k} \sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2}{\frac{1}{n-k-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$  – статистика для проверки нулевой

гипотезы о значимости регрессии. Заметим, что статистика  $F$



может быть рассчитана также по формуле 
$$F = \frac{\frac{1}{k}R^2}{\frac{1}{n-k-1}(1-R^2)}.$$

В обоих случаях мы получим одинаковый результат.

- значимость  $F$  – достигнутый уровень значимости (вероятность ошибки первого рода).

В третьей таблице:

- Коэффициенты – значения оценок коэффициентов  $\beta_0^*, \beta_1^*, \beta_2^*$ ;
- Стандартная ошибка – значения оценок среднеквадратичных отклонений оценок коэффициентов  $\beta_0^*, \beta_1^*, \beta_2^*$ ;
- t-статистика – наблюдаемые значения статистик критерия проверки значимости коэффициентов для соответственно коэффициентов  $\beta_0, \beta_1, \beta_2$ ;
- P-значения – достигнутые значения уровня значимости  $P(x > |t_j|)$ .

Соответствующий график предсказанных значений в сравнении с исходными данными имеет вид (рис. 4.2):

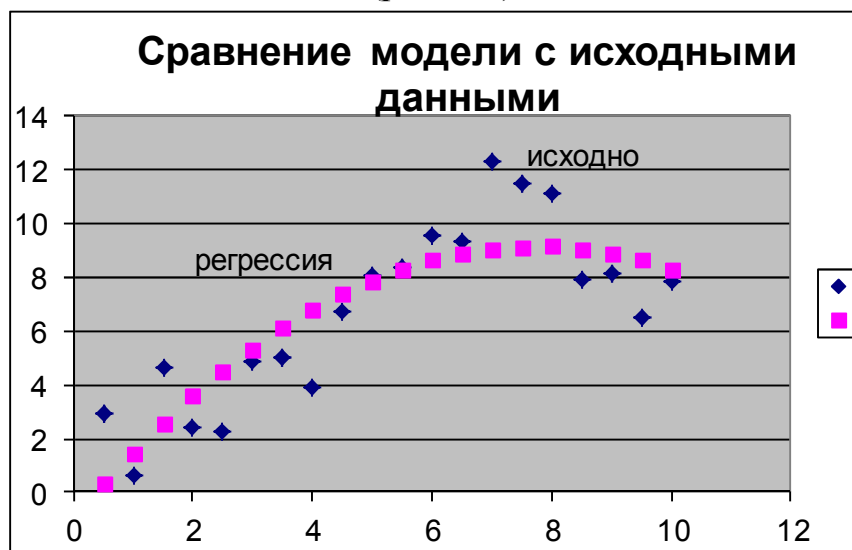


Рис.4.2. Сравнение модели  $y = \beta_0 + \beta_1x + \beta_2x^2$  с исходными данными

Отметим, что полученная оценка значения  $\sigma$  велика:  $s = 1,82$ . Что касается коэффициентов модели, то кроме  $\beta_0$  все они значительно отличаются от нуля (достигнутый уровень значимости достаточно мал, поэтому можно отвергнуть гипотезу о равенстве коэффициентов нулю). Попробуем улучшить модель, увеличим порядок полинома, пусть  $y = \beta_0 + \beta_1x + \beta_2x^2 + \beta_3x^3$

Проводим линеаризацию, полагая  $x_1 = x$ ,  $x_2 = x^2$ ,  $x_3 = x^3$ , и оцениваем коэффициенты новой модели.

### ВЫВОД ИТОГОВ

<i>Регрессионная статистика</i>	
Множественный R	0,925507816
R-квадрат	0,856564717
Нормированный R-квадрат	0,829670602
Стандартная ошибка	1,358958106
Наблюдения	20

	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>
Y-пересечение	3,176148907	1,484433212	2,13963746	0,04814219
Переменная X 1	-1,619155418	1,194561911	-1,355438679	0,19410420
Переменная X 2	0,814063749	0,261005991	3,118946605	0,00661209
Переменная X 3	-0,062325275	0,016365816	-3,808259635	0,00154556

График показан на рис. 4.3

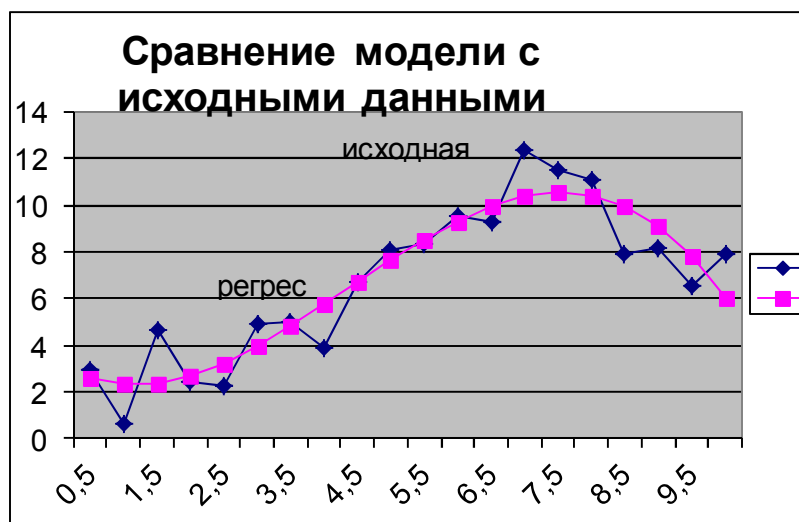


Рис. 4.3. Сравнение модели  $y = \beta_0 + \beta_1x + \beta_2x^2 + \beta_3x^3$  с исходными данными

Заметим, что коэффициент детерминации увеличился, а оценка  $\sigma$  (стандартная ошибка) уменьшилась, что говорит о лучшем качестве модели по сравнению с предыдущей. Причем значение этой оценки близко к значениям  $\sigma = 1,3$ , указанному в задании. Из коэффициентов можно считать, что  $\beta_1$  не значительно отличается от нуля (достигнутый уровень значимости  $\alpha = 0,194$ , говорит о том, что при истинности гипотезы  $H_0 : \beta_1 = 0$ , такое или большее значение  $t$ -статистики критерия могло наблюдаться с вероятностью  $0,194$ ). Поэтому, можно положить  $\beta_1 = 0$  и модель соответственно примет вид:

$$y = \beta_0 + \beta_2x^2 + \beta_3x^3$$

Результаты регрессионного анализа для этой модели:

Регрессионная статистика	
Множественный R	0,916566765
R-квадрат	0,840094635
Нормированный R-квадрат	0,82128224
Стандартная ошибка	1,39201886
Наблюдения	20

	Коэффициенты	Стандартная ошибка	$t$ -статистика	$P$ -Значение
Y-пересечение	1,356030259	0,648133062	2,092209668	0,05173357

Переменная X				6,05722E-
1	0,469599546	0,060941974	7,70568321	07
Переменная X				
2	-0,041727702	0,006223514	-6,70484584	3,6971E-06

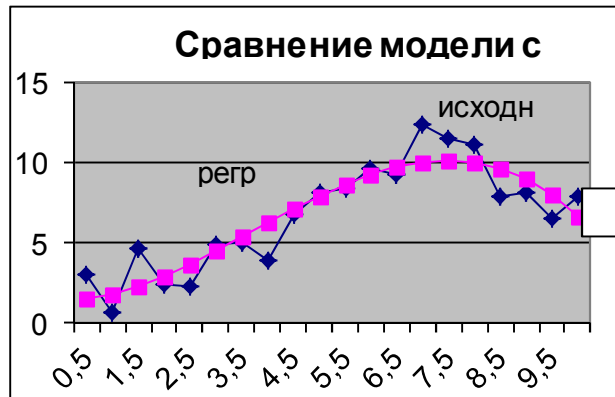


Рис. 4.4. Сравнение модели  $y = \beta_0 + \beta_2 x^2 + \beta_3 x^3$  с исходными данными

И хотя, параметры модели немного ухудшились (сравните R-квадрат и стандартную ошибку!), тем не менее все коэффициенты получились значимыми, поэтому данная модель, предпочтительнее предыдущей.

Можно ли повысить еще качество модели? В классе полиномов это сделать не удастся. Повышение порядка полинома (можно проверить!), уже больше не понижает стандартную ошибку. Следовательно, улучшение нужно искать, используя иные классы функций. Например, можно предположить, что существует периодическая зависимость значений  $Y$  от  $X$ , тогда надо добавить в модель гармонические составляющие вида  $\beta_1 \cos(\omega x) + \beta_2 \sin(\omega x)$ . Частоты этих составляющих  $\omega$  придется подбирать отдельно. Можно предположить, анализируя зависимость, что у нас присутствует периодическая составляющая с частотой порядка 1 (одно колебание за весь интервал изменений  $X$ ). Построим модель вида

$$y = \beta_0 + \beta_1 x + \beta_2 \cos(2\pi x / 10) + \beta_3 \sin(2\pi x / 10).$$

Получим оценки для этой модели:

#### ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,934315451
R-квадрат	0,872945362
Нормированный R-квадрат	0,849122618
Стандартная ошибка	1,279008211

	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение
Y-пересечение	5,080728355	0,88616922	5,733361352	3,08053E-05
Переменная X 1	0,308759559	0,159762044	1,932621493	0,0711836
Переменная X 2	-1,299656278	0,412270758	-3,152433814	0,00616364
Переменная X 3	-3,032825609	0,646493647	-4,691191664	0,00024523

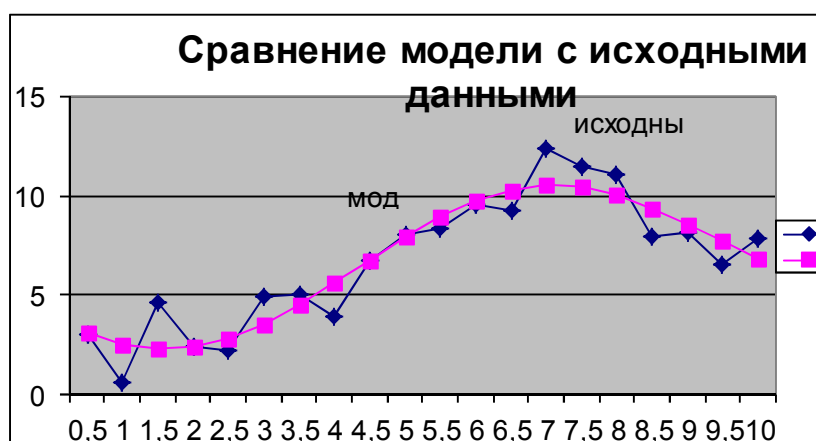


Рис. 4.5. Сравнение модели  $y = \beta_0 + \beta_1 x + \beta_2 \cos(2\pi x/10) + \beta_3 \sin(2\pi x/10)$

с исходными данными

Заметим, что мы получили значение стандартной ошибки меньше заданной величины  $s$ . Это говорит о том, что дальше улучшать модель бессмысленно. Если мы продолжим, то будем по сути аппроксимировать случайные ошибки, а не реальную существующую зависимость  $Y$  от  $X$ . Качество модели выше, чем у модели вида  $y = \beta_0 + \beta_2 x^2 + \beta_3 x^3$ . Однако, данная модель содержит на один коэффициент больше, и кроме того содержит параметр, значение которого по сути определяется вручную. Поскольку модели принципиально различные, то предпочесть следует ту, которая более соответствует физике явления (если она известна). Мы остановимся на последней модели  $y = \beta_0 + \beta_1 x + \beta_2 \cos(2\pi x/10) + \beta_3 \sin(2\pi x/10)$ .

Найдем доверительный интервал для  $\sigma^2$  соответствующий уровню 0,95. Находим квантили распределения хи-квадрат уровней 0,025 и

0,975 соответственно для числа степеней свободы  $\nu = n - k - 1 = 20 - 3 - 1 = 16$  ( $k$  – число коэффициентов модели, не считая  $\beta_0$ ):  $\tau_{0,025} = 6,91$ ,  $\tau_{0,975} = 28,85$ . Тогда  $\frac{(n - k - 1)s^2}{\tau_{0,975}} < \sigma^2 < \frac{(n - k - 1)s^2}{\tau_{0,025}} \Rightarrow 0,91 < \sigma^2 < 3,79 \Rightarrow 0,95 < \sigma < 1,95$ .

Доверительные интервалы для коэффициентов уравнения регрессии можно найти в Итоговой статистике.

Доверительный интервал для значений  $y$  для тех же значений  $X$ , что приведены в выборке, рассчитываем по формуле

$$y_j = \left[ (X^T B^*)_j \pm t_\alpha s \left( 1 + \sqrt{(X(X^T X)^{-1} X^T)_{jj}} \right) \right],$$

где  $t_\alpha$  квантиль распределения Стьюдента с  $n - k - 1$  степенью свободы (доверительный уровень возьмем 0,67, тогда  $\alpha = 0,33$  и  $t_\alpha = 1,0047$ ).

Доверительный интервал для значений  $y$ , полученный по этим формулам, отображен на рис 4.6.

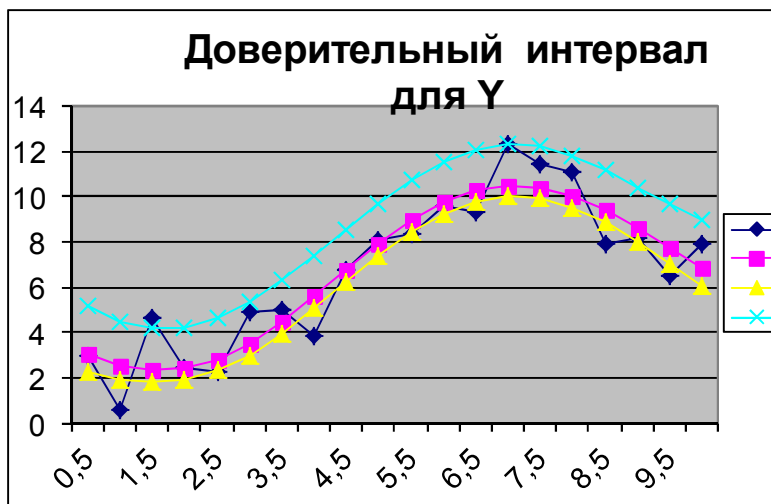


Рис. 4.6. Доверительный интервал для  $Y$

### 3. Задание

Имеется выборка значений совместно наблюдаемых величин  $X$  и  $Y$  (приложение).

#### Требуется:

8) Отобразить графически поле наблюдаемых значений величин  $X$  и  $Y$ .

9) Подобрать 2–3 аппроксимирующих зависимости для уравнения регрессии, провести линеаризацию моделей.

Методом наименьших квадратов найти оценки параметров каждой модели. Оценить для каждой модели остаточную дисперсию (диспер-

сию предсказания). Сравнить графически каждую кривую регрессии с наблюдаемыми значениями  $Y$  и выбрать модель регрессии. Желательно, чтобы для выбранной модели оценка дисперсии  $s^2$  была в пределах указанной в задании величины  $\sigma^2$  (Величина  $\sigma^2$  в расчетах нигде не используется и указана только лишь для сопоставления с  $s^2$  !!!).

10) Для выбранной модели найти оценки дисперсий коэффициентов уравнения регрессии и проверить значимость коэффициентов уравнения регрессии в предположении нормальности распределения ошибок (доверительный уровень принять равным 0,95).

11) Определить доверительные интервалы для остаточной дисперсии и коэффициентов уравнения регрессии.

12) Определить границы доверительного интервала для средних значений  $\hat{y} = M(Y|X = x)$  и значений  $y$  при каждом наблюдаемом значении  $x$  (отобразить графически).

## Приложение

### Варианты заданий

Вариант 1.

X	1	2	3	4	5	6	7	8	9	10
Y	0,53	0,64	0,21	0,38	1,85	0,69	5,11	4,12	6,09	4,01
X	11	12	13	14	15	16	17	18	19	20
Y	2,85	8,70	6,28	4,31	4,76	0,82	3,07	6,32	6,74	3,15

$\sigma = 2$

Вариант 2.

X	1	2	3	4	5	6	7	8	9	10
Y	7,81	6,36	9,07	8,12	7,01	5,28	5,23	7,42	5,72	-2,11
X	11	12	13	14	15	16	17	18	19	20
Y	3,57	0,06	4,96	3,88	3,04	11,18	4,93	0,58	3,99	5,55

$\sigma = 2,2$

Вариант 3.

X	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
---	---	-----	-----	-----	-----	-----	-----	-----	-----	-----

Y	0,47	0,38	0,56	0,70	0,75	0,86	0,53	0,80	0,99	0,84
X	1	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9
Y	0,93	0,88	0,89	0,99	1,01	0,85	0,96	0,89	0,84	0,64

$\sigma = 0,1$

#### Вариант 4.

X	1	2	3	4	5	6	7	8	9	10
Y	5,98	7,43	0,98	-0,75	0,34	-2,84	1,79	6,04	5,09	8,78
X	11	12	13	14	15	16	17	18	19	20
Y	10,39	6,73	4,39	6,53	10,08	12,51	8,75	21,91	26,93	27,73

$\sigma = 2$

#### Вариант 5.

X	0,2	0,4	0,6	0,8	1	1,2	1,4	1,6	1,8	2
Y	2,21	1,55	2,53	3,21	3,01	3,23	3,93	2,91	2,99	4,33
X	2,2	2,4	2,6	2,8	3	3,2	3,4	3,6	3,8	4
Y	3,91	3,70	4,16	3,68	3,73	3,92	2,94	2,91	2,29	1,09

$\sigma = 0,5$

#### Вариант 6.

X	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5
Y	7,49	4,79	4,72	3,99	4,65	4,70	5,85	4,38	5,30	4,16
X	5,5	6	6,5	7	7,5	8	8,5	9	9,5	10
Y	5,37	7,75	8,63	12,04	11,12	7,24	9,22	15,11	15,54	17,70

$\sigma = 1$

#### Вариант 7.

X	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
Y	1,78	1,65	0,96	1,04	1,26	1,09	1,35	0,95	0,93	0,77
X	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2
Y	0,86	0,82	1,22	1,25	1,29	1,50	1,35	1,52	1,97	2,40



$\sigma = 0,15$

Вариант 8.

X	1	2	3	4	5	6	7	8	9	10
Y	7,37	10,35	9,69	10,25	12,23	14,91	15,06	14,27	11,89	13,31
X	11	12	13	14	15	16	17	18	19	20
Y	13,51	12,68	12,83	14,38	13,33	12,34	11,86	14,71	13,56	12,52

$\sigma = 1,3$

Вариант 9.

X	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
		—	—		—				—	
Y	-8,74	5,33	6,17	-5,17	2,22	1,73	2,43	1,94	0,61	1,82
X	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2
Y	2,69	2,23	3,07	5,73	5,01	4,37	4,41	8,80	7,97	7,29

$\sigma = 1,7$

Вариант 10.

X	1	2	3	4	5	6	7	8	9	10
Y	2,26	3,82	3,50	3,47	3,90	4,56	4,24	3,56	2,29	2,51
X	11	12	13	14	15	16	17	18	19	20
Y	2,32	1,77	1,67	2,16	1,74	1,42	1,39	2,73	2,63	2,69

$\sigma = 0,5$

## ЛАБОРАТОРНАЯ РАБОТА 5

### Тема 5. Комплексная оценка финансового состояния предприятия

#### **Цель работы:**

Оценка риска банкротства предприятия на основе комплексного показателя

#### **1. Методика оценки**

Методика предусматривает выполнение следующих шагов:

1. **Идентификация лингвистических переменных** «Состояние предприятия»  $E_j$  (например, «предельное неблагополучие», «неблагополучие», «среднего качества», «относительное благополучие» и «предельное благополучие»), «Риск банкротства»  $G_j$  (например, «запредельный», «высокий», «пограничный», «приемлемый» и «незначительный»), а также установление однозначного соответствия между введенными переменными. Элементами носителя (множества  $G$ ) является показатель степени банкротства  $g$ , принимающий значения от 0 до 1.

2. **Выбор финансовых показателей  $X_i$** , которые наилучшим образом характеризуют отдельные аспекты деятельности предприятия и влияют на оценку риска банкротства (см. табл. 5.1).

Таблица 5.1.

#### *Характеристика финансовых показателей*

Шифр показателя $X_i$	Наименование показателя $X_i$	Описание
$X_1$	Коэффициент автономии	Собственный капитал / Валюта баланса
$X_2$	Коэффициент промежуточной ликвидности	[Краткосрочная дебиторская задолженность + Краткосрочные финансовые вложения + Денежные средства] / Краткосрочные обязательства
$X_3$	Коэффициент абсолютной ликвидности	[Денежные средства + Краткосрочные финансовые вложения] / Краткосрочные обязательства
$X_4$	Оборачиваемость активов	Выручка за период / Средняя стоимость активов
$X_5$	Рентабельность основ-	Прибыль от реализации / Затраты на про-

	ной деятельности	изводство продукции
X <sub>6</sub>	Рентабельность активов	Чистая прибыль / Средняя балансовая стоимость активов

- X<sub>1</sub> – коэффициент автономии (коэффициент концентрации собственных средств, который показывает долю собственных средств в стоимости имущества предприятия),

- X<sub>2</sub> – коэффициент промежуточной ликвидности (отражает способность компании погашать свои текущие обязательства в случае возникновения сложностей с реализацией продукции),

- X<sub>3</sub> – коэффициент абсолютной ликвидности (характеризует способность компании погашать текущие обязательства за счёт денежных средств, средств на расчетных счетах и краткосрочных финансовых вложений; наиболее важен для поставщиков товарно–материальных ресурсов и для банков, кредитующих предприятие),

- X<sub>4</sub> – оборачиваемость активов (отражает скорость оборота денежных средств вложенных в имущество),

- X<sub>5</sub> – рентабельность основной деятельности (определяет, сколько чистой прибыли получено с 1 рубля затрат на производство),

- X<sub>6</sub> – рентабельность активов (показывает сколько предприятие имеет чистой прибыли с рубля вложенного в капитал).

В качестве примера рассматривается финансовое состояние предприятия ОАО «ФСК ЕЭС» за период 2009 – 2011 г.г., которое характеризуется следующими финансовыми показателями (табл. 5.2).

Таблица 5.2.

*Значения финансовых показателей предприятия ОАО «ФСК ЕЭС»*

Шифр показателя X <sub>i</sub>	Наименование показателя X <sub>i</sub>	Значение X <sub>i</sub> в период I (2009 г.) (X <sub>I,i</sub> )	Значение X <sub>i</sub> в период II (2010 г.) (X <sub>II,i</sub> )	Значение X <sub>i</sub> в период III (2011 г.) (X <sub>III,i</sub> )
X <sub>1</sub>	Коэффициент автономии	0,92	0,877	0,876
X <sub>2</sub>	Коэффициент промежуточной ликвидности	5,013	2,685	3,911
X <sub>3</sub>	Коэффициент абсолютной ликвидности	1,406	1,09	1,045
X <sub>4</sub>	Оборачиваемость активов	0,134	0,125	0,137
X <sub>5</sub>	Рентабельность основной деятельности	0,081	0,229	0,359
X <sub>6</sub>	Рентабельность активов	0,09	-0,088	0,072

Затем для каждого показателя вводится лингвистическая переменная «Уровень показателя  $X_j$ », с терм–множеством «очень низкий», «низкий», «средний», «высокий», «очень высокий».

### 3. *Определение системы весов показателей.*

Для характеристики показателей воспользуемся экспертным методом.

В качестве членов экспертной группы было привлечено 8 специалистов одного из филиалов ОАО "ФСК ЕЭС".

На основе полученных данных о значимости показателей (табл. 5.3) в результате экспертного опроса была дана оценка согласованности мнений экспертов с помощью коэффициента конкордации ( $W$ ).

Таблица 5.3.

*Оценки экспертов*

	$\mathcal{E}_1$	$\mathcal{E}_2$	$\mathcal{E}_3$	$\mathcal{E}_4$	$\mathcal{E}_5$	$\mathcal{E}_6$	$\mathcal{E}_7$	$\mathcal{E}_8$
$X_1$	6	6	6	6	5	6	6	5
$X_2$	2	2	2	2	2	1	2	1
$X_3$	1	1	1	1	1	2	1	2
$X_4$	5	5	5	5	6	5	5	6
$X_5$	4	4	4	3	3	4	4	4
$X_6$	3	3	3	4	4	3	3	3

В общем виде формула расчета коэффициента конкордации имеет следующий вид:

$$W = \frac{12 \sum_{j=1}^m d_j^2}{n^2 (m^3 - m)},$$

где  $n$  – количество экспертов,

$m$  – количество параметров (оцениваемых объектов),

$d_j^2$  – отклонение суммы рангов по  $j$ -ому параметру от среднего значения рангов:

$$d_j^2 = \left( \sum_{i=1}^n R_{ji} - \frac{n(m+1)}{2} \right)^2,$$

где  $R_{ji} \in \{1, 2, \dots, m\}$  – ранг  $j$ -го параметра  $i$  – го эксперта ( $i = 1, 2, \dots, n$ ).

Результат обработки экспертных оценок значимости приведен в табл. 5.4.

Таблица 5.4.

*Оценка согласованности мнений экспертов*

<b>Коэффициент</b>	<b>Сумма рангов, <math>R_{ji}</math></b>	$d_j^2$	$1/R_{ji}$	<b>Оценка важности <math>R_{ji}/n</math></b>
X <sub>1</sub>	46	324	0,022	6
X <sub>2</sub>	14	196	0,071	2
X <sub>3</sub>	10	324	0,100	1
X <sub>4</sub>	42	196	0,024	5
X <sub>5</sub>	30	4	0,033	4
X <sub>6</sub>	26	4	0,038	3
Итого	168	1048	0,289	
W=0.94				

Таким образом, следует отметить, что степень согласованности мнений экспертов составляет 0,94 (что больше критического значения 0,75), что является достаточной согласованностью.

4. Установление соответствия между значениями показателя степени риска ( $g$ ) и нечеткими подмножествами множества  $G$  (см. табл. 5.5).

Таблица 5.5.

*Классификация степени риска банкротства*

<b>Интервал значений <math>g</math></b>	<b>Классификация уровня параметра</b>	<b>Функция принадлежности</b>
$0 \leq g \leq 0.15$	риск банкротства незначителен	1
$0.15 < g < 0.25$	риск банкротства незначителен	$\mu_5 = 10 \times (0.25 - g)$
	приемлемый риск банкротства	$1 - \mu_5 = \mu_4$
$0.25 \leq g \leq 0.35$	приемлемый риск банкротства	1
$0.35 < g < 0.45$	приемлемый риск банкротства	$\mu_4 = 10 \times (0.45 - g)$
	риск банкротства пограничный	$1 - \mu_4 = \mu_3$
$0.45 \leq g \leq 0.55$	риск банкротства пограничный	1

$0.55 < g < 0.65$	риск банкротства пограничный	$\mu_3 = 10 \times (0.65 - g)$
	риск банкротства высокий	$1 - \mu_3 = \mu_2$
$0.65 \leq g \leq 0.75$	риск банкротства высокий	1
$0.75 < g < 0.85$	риск банкротства высокий	$\mu_2 = 10 \times (0.85 - g)$
	запредельный риск банкротства	$1 - \mu_2 = \mu_1$
$0.85 \leq g \leq 1.0$	запредельный риск банкротства	1

5. Оценка границ интервалов значений показателей. Построение функций принадлежности.

Пусть экспертному сообществу в качестве исходной информации предлагается интервал по каждому показателю. Необходимо в рамках этого интервала выполнить внутреннее разбиение на 5 уровней (например, «очень низкий», «низкий», «средний», «высокий», «очень высокий»).

Данные, полученные в результате экспертного опроса о показателе рентабельности деятельности представлены в табл. 5.6 – 5.11.

### Вариант 1

Таблица 5.6

Данные опроса экспертов по показателю  $X_1$

	Э <sub>1</sub>	Э <sub>2</sub>	Э <sub>3</sub>	Э <sub>4</sub>	Э <sub>5</sub>	Э <sub>6</sub>	Э <sub>7</sub>	Э <sub>8</sub>
«очень низкий»	0; 0,1	0; 0,1	0; 0,2	0; 0,15	0; 0,1	0; 0,1	0; 0,1	0; 0,1
«низкий»	0,1; 0,3	0,1; 0,3	0,2; 0,4	0,15; 0,4	0,1; 0,3	0,1; 0,3	0,1; 0,3	0,1; 0,3
«средний»	0,3; 0,5	0,3; 0,5	0,4; 0,5	0,4; 0,45	0,3; 0,5	0,3; 0,5	0,3; 0,5	0,3; 0,45
«высокий»	0,5; 0,8	0,5; 0,8	0,5; 0,7	0,45; 0,6	0,5; 0,7	0,5; 0,6	0,5; 0,6	0,45; 0,75
«очень высокий»	0,8; 1	0,8; 1	0,7; 1	0,6; 1	0,7; 1	0,6; 1	0,6; 1	0,75; 1

### Вариант 2

Таблица 5.7

Данные опроса экспертов по показателю  $X_2$

	Э <sub>1</sub>	Э <sub>2</sub>	Э <sub>3</sub>	Э <sub>4</sub>	Э <sub>5</sub>	Э <sub>6</sub>	Э <sub>7</sub>	Э <sub>8</sub>
«очень низкий»	0; 0,2	0; 0,2	0; 0,4	0; 0,35	0; 0,4	0; 0,3	0; 0,3	0; 0,2
«низкий»	0,2; 0,6	0,2; 0,6	0,4; 0,7	0,35; 0,7	0,4; 0,65	0,3; 0,7	0,3; 0,6	0,2; 0,6
«средний»	0,6; 1	0,6; 1	0,7; 0,9	0,7; 0,8	0,65; 1	0,7; 0,9	0,6; 0,8	0,6; 0,8
«высокий»	1; 1,5	1; 1,8	0,9; 1,5	0,8; 1,7	1; 1,6	0,9; 1,5	0,8; 1,5	0,8; 1,8
«очень высокий»	1,5; ∞	1,8; ∞	1,5; ∞	1,7; ∞	1,6; ∞	1,5; ∞	1,5; ∞	1,8; ∞

### Вариант 3

Таблица 5.8

*Данные опроса экспертов по показателю  $X_3$*

	$\mathcal{E}_1$	$\mathcal{E}_2$	$\mathcal{E}_3$	$\mathcal{E}_4$	$\mathcal{E}_5$	$\mathcal{E}_6$	$\mathcal{E}_7$	$\mathcal{E}_8$
«очень низкий»	0; 0,08	0; 0,05	0; 0,06	0; 0,05	0; 0,05	0; 0,06	0; 0,08	0; 0,05
«низкий»	0,08; 0,2	0,05; 0,2	0,06; 0,12	0,05; 0,1	0,05; 0,15	0,06; 0,1	0,08; 0,18	0,05; 0,2
«средний»	0,2; 0,4	0,2; 0,4	0,12; 0,32	0,1; 0,3	0,15; 0,3	0,1; 0,3	0,18; 0,4	0,2; 0,4
«высокий»	0,4; 0,8	0,4; 1	0,32; 0,5	0,3; 0,6	0,3; 0,6	0,3; 0,5	0,4; 0,8	0,4; 0,8
«очень высокий»	0,8; $\infty$	1; $\infty$	0,5; $\infty$	0,6; $\infty$	0,6; $\infty$	0,5; $\infty$	0,8; $\infty$	0,8; $\infty$

### Вариант 4

Таблица 5.9

*Данные опроса экспертов по показателю  $X_4$*

	$\mathcal{E}_1$	$\mathcal{E}_2$	$\mathcal{E}_3$	$\mathcal{E}_4$	$\mathcal{E}_5$	$\mathcal{E}_6$	$\mathcal{E}_7$	$\mathcal{E}_8$
«очень низкий»	0; 0,12	0; 0,15	0; 0,12	0; 0,12	0; 0,12	0; 0,15	0; 0,15	0; 0,15
«низкий»	0,12; 0,25	0,15; 0,25	0,12; 0,2	0,12; 0,2	0,12; 0,2	0,15; 0,25	0,15; 0,25	0,15; 0,25
«средний»	0,25; 0,35	0,25; 0,45	0,2; 0,4	0,2; 0,4	0,2; 0,45	0,25; 0,35	0,25; 0,4	0,25; 0,35
«высокий»	0,35; 0,85	0,45; 1	0,4; 0,8	0,4; 0,8	0,45; 0,6	0,35; 0,65	0,4; 0,8	0,35; 0,75
«очень высокий»	0,85; $\infty$	1; $\infty$	0,8; $\infty$	0,8; $\infty$	0,6; $\infty$	0,65; $\infty$	0,8; $\infty$	0,75; $\infty$

### Вариант 5

Таблица 5.10

*Данные опроса экспертов по показателю  $X_5$*

	$\mathcal{E}_1$	$\mathcal{E}_2$	$\mathcal{E}_3$	$\mathcal{E}_4$	$\mathcal{E}_5$	$\mathcal{E}_6$	$\mathcal{E}_7$	$\mathcal{E}_8$
«очень низкий»	-1; -0,06	-1; -0,08	-1; -0,03	-1; -0,03	-1; -0,02	-1; -0,02	-1; -0,01	-1; -0,02
«низкий»	-0,06; 0,04	-0,08; 0,01	-0,03; 0,05	-0,03; 0,04	-0,02; 0,07	-0,02; 0,07	-0,01; 0,08	-0,02; 0,08
«средний»	0,04; 0,1	0,01; 0,1	0,05; 0,15	0,04; 0,14	0,07; 0,17	0,07; 0,15	0,08; 0,18	0,08; 0,18
«высокий»	0,1; 0,24	0,1; 0,3	0,15; 0,4	0,14; 0,4	0,17; 0,4	0,15; 0,45	0,18; 0,45	0,18; 0,48
«очень высокий»	0,24; 1	0,3; 1	0,4; 1	0,4; 1	0,4; 1	0,45; 1	0,45; 1	0,48; 1

## Вариант 6

Таблица 5.11

*Данные опроса экспертов по показателю  $X_6$*

	$\mathfrak{A}_1$	$\mathfrak{A}_2$	$\mathfrak{A}_3$	$\mathfrak{A}_4$	$\mathfrak{A}_5$	$\mathfrak{A}_6$	$\mathfrak{A}_7$	$\mathfrak{A}_8$
«очень низкий»	-1; 0	-1; 0	-1; 0	-1; 0	-1; 0	-1; 0	-1; 0	-1; 0
«низкий»	0; 0,01	0; 0,015	0; 0,006	0; 0,01	0; 0,01	0; 0,01	0; 0,012	0; 0,008
«средний»	0,01; 0,1	0,015; 0,1	0,006; 0,06	0,01; 0,06	0,01; 0,06	0,01; 0,1	0,012; 0,1	0,008; 0,08
«высокий»	0,1; 0,3	0,1; 0,4	0,06; 0,24	0,06; 0,3	0,06; 0,3	0,1; 0,4	0,1; 0,4	0,08; 0,24
«очень высокий»	0,3; $\infty$	0,4; $\infty$	0,24; $\infty$	0,3; $\infty$	0,3; $\infty$	0,4; $\infty$	0,4; $\infty$	0,24; $\infty$

Результатом опроса является 8 интервалов вещественной оси  $[a_i; b_i]$ ,  $i = 1, \dots, 8$ .

$$\text{Определим } A = \min_i \{a_i\}, \quad B = \min_i \{\max(a_i), \min(b_i)\}$$

$$C = \max_i \{\max(a_i), \min(b_i)\}, \quad D = \max_i \{b_i\}.$$

Тогда четыре пары чисел –  $(A; 0)$ ,  $(B; 1)$ ,  $(C; 1)$ ,  $(D; 0)$  – являются множеством вершин трапецевидной функции принадлежности. Например, для значения «очень низкий» показателя  $X_5$  получим точки:  $(-1; 0)$ ,  $(-1; 1)$ ,  $(-0,01; 1)$ ,  $(-0,01; 0)$ ; для значения «низкий» –  $(-0,08; 0)$ ,  $(-0,01; 1)$ ;  $(0,01; 1)$ ;  $(0,08; 1)$ .

В результате мы получим опорные точки трапецевидных функций принадлежности. В качестве примера приведена часть заполненной таблицы 5.12 для показателя  $X_5$ .

Таблица 5.12

*Опорные точки трапецевидных функций принадлежности для показателя  $X_5$*

Опорные точки	Нечеткое подмножество				
	«очень низкий»	«низкий»	«средний»	«высокий»	«очень высокий»
$a_1$	-1	-0,08			
$a_2$	-1	-0,01			
$a_3$	-0,08	0,01			
$a_4$	-0,01	0,08			



*Примечание. Студент заполняет таблицу 7.N для своего показателя в соответствии со своим номером варианта N.*

Таким образом, выбранные показатели на основании предварительного экспертного анализа получили следующую классификацию (табл. 5.13):

Таблица 5.13.

*Результаты классификации параметров  $X_1 - X_6$*

Шифр показателя	Т-числа $\{\gamma\}$ для значений лингвистической переменной "Величина параметра":				
	"очень низкий"	"низкий"	"средний"	"высокий"	"очень высокий"
$X_1$					
$X_2$					
$X_3$					
$X_4$					
$X_5$	(-1; -1; -0,08; -0,01; 0,08; 0,01)	(-0,08; -0,01; 0,01; 0,08)			
$X_6$					

*Примечание. Студент заполняет строку с номером N таблицы 8 для своего показателя в соответствии со своим номером варианта N. Затем расчеты всех вариантов студенты передают друг другу для полного заполнения таблицы 8.*

В табл. 8 приведена классификация показателей с использованием трапециевидных чисел вида  $(a_1, a_2, a_3, a_4)$ , где  $a_1$  и  $a_4$  – абсциссы нижнего основания, а  $a_2$  и  $a_3$  – абсциссы верхнего основания трапеции. Верхнее основание трапеции соответствует уверенности эксперта в правильности своей классификации, а ее ребра задают интервал неуверенности.

#### 6. Распознавание уровня показателей.

Распознавание уровня показателя – самый деликатный вопрос данного метода. Если значение фактора точно попадают в рассматриваемый интервал, то значение равно единице для данного качественного уровня и нулю для всех остальных уровней. Если значение лежит в зоне неуверенности, то для двух смежных классов формируются значения, сумма которых равна единице. Вычисление значений идет по правилу вычисления ординаты наклонного ребра трапециевидной функции принадлежности по заданной абсциссе точки на нижнем основании трапеции.

На основе табл. 8 построим правила распознавания уровней показателей  $X_1 - X_6$ .

Итак, проведем вычисления на примере  $X_6$  – показателя рентабельности активов за 2010 год. Воспользуемся табл. 8 «Результаты классификации параметров  $X_1 - X_6$ » для определения степени оценочной уверенности принадлежности к этому уровню.

Графическая интерпретация представлена на рис. 5.1, где отображен переход от уровня «средний» к уровню «высокий».



Рис. 5.1. Переход от уровня «средний» к уровню «высокий»

Так как фактическое значение, равное 0,072 попадает в переходные интервал, то принадлежность уровню и степени оценочной уверенности этой принадлежности можно определить из уравнения:

$$COU_{\text{средний}} = \frac{-100x+10}{4} = \frac{-100 \cdot 0,072+10}{4} = 0,7$$

$$COU_{\text{высокий}} = \frac{100x-6}{4} = \frac{100 \cdot 0,072-6}{4} = 0,3$$

где  $COU_{\text{средний}}$  – степень оценочной уверенности принадлежности к «среднему» уровню на основании линейной функции принадлежности;

$COU_{\text{высокий}}$  – степень оценочной уверенности принадлежности к «высокому» уровню на основании линейной функции принадлежности

Таким образом,  $COU_{\text{средний}} = 0,7$ ;  $COU_{\text{высокий}} = 0,3$ .

Аналогичным образом происходит распознавание уровней остальных показателей по критерию.

Результатом проведенной классификации является табл. 5.14:

Таблица 5.14.

Классификация уровня показателей  $X_1$ –  $X_6$  за 2008–2010гг.

Год	Показатель $X_i$	Значение $\{\square\}$				
		$\square_1(x_{1,i})$	$\square_2(x_{1,i})$	$\square_3(x_{1,i})$	$\square_4(x_{1,i})$	$\square_5(x_{1,i})$
2008	$X_1$					
	$X_2$					
	$X_3$					
	$X_4$					
	$X_5$					
	$X_6$					
2009	$X_1$					
	$X_2$					
	$X_3$					
	$X_4$					
	$X_5$					
	$X_6$					
2010	$X_1$					
	$X_2$					
	$X_3$					
	$X_4$					
	$X_5$					
	$X_6$	0	0	0,7	0,3	0

*Примечание.* Студент заполняет строку с номером  $N$  таблицы 9 для своего показателя в соответствии со своим номером варианта  $N$  для каждого период (2008, 2009, 2010 гг.). Затем расчеты всех вариантов студенты передают друг другу для полного заполнения таблицы 9.

### 7. Определение степени риска банкротства. Анализ результатов.

В соответствии с результатом распознавания количественное значение степени риска банкротства предприятия  $g$  определяется по формуле двойной свертки Недосекина:

$$g = \sum_{j=1}^5 g_j \sum_{i=1}^N r_i \lambda_{ij}, \quad \text{где } g_j = 0.9 - 0.2(j-1)$$

Здесь  $r_i = 1/N = 1/6$  – веса показателей.

С учетом вышеизложенного, оценка степени риска банкротства по формуле дает  $g_I = 0,373$ ,  $g_{II} = 0,396$ ,  $g_{III} = 0,288$ , т.е. наблюдается сокращение риска банкротства.

Лингвистическое распознавание значений  $g$  по данным определяет с большей степенью риска банкротства предприятия ОАО «ФСК ЕЭС» (функция принадлежности равна 0,77) как приемлемая и с меньшей степенью риска, как пограничная (функция принадлежности равна 0,23). Степень риска банкротства предприятия во втором периоде распознается примерно с одинаковой степенью соответствия как приемлемая (0,54) и как пограничная (0,46). В третьем периоде степень риска банкротства предприятия распознается уверенно как приемлемая (функция принадлежности равна 1).

## **СОДЕРЖАНИЕ ОТЧЕТА ПО ЛАБОРАТОРНЫМ РАБОТАМ**

1. Титульный лист (Приложение А).
2. Название и цель лабораторной работы.
3. Вариант задания.
4. Детальное описание выполненных действий.
5. Обязательное представление формул, расчетов и скриншотов из программы, выполняющей анализ.
6. Выводы по работе.

## СПИСОК ЛИТЕРАТУРЫ

1. Силич В.А., Силич И.П. Теория систем и системный анализ: Учебное пособие. – Томск: Томский политехнический университет, 2010. – 281 с.
2. Маслов А.В. Теория систем и системный анализ / А.В. Маслов. – Юрга: Изд-во ЮТИ ТПУ, 2011. – 238 с.
3. Черников Ю.Г. Системный анализ и исследование операций: Учебное пособие для вузов. – М.: Издательство Московского государственного горного университета, 2006. – 370 с. – [Электронный ресурс] – Режим доступа: <http://ezproxy.ha.tpu.ru:2117/view/book/3512/>.
4. Анфилатов В.С. и др. Системный анализ в управлении: Учебное пособие / В.С. Анфилатов, А.А. Емельянов, А.А. Кукушкин; Под ред. А.А. Емельянова. – М.: Финансы и статистика, 2009. – 368 с.

**Приложение А**

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ**  
Федеральное государственное бюджетное образовательное учреждение  
высшего профессионального образования  
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

**ЮРГИНСКИЙ ТЕХНОЛОГИЧЕСКИЙ ИНСТИТУТ (ФИЛИАЛ)  
ФЕДЕРАЛЬНОГО ГОСУДАРСТВЕННОГО БЮДЖЕТНОГО ОБРАЗОВА-  
ТЕЛЬНОГО УЧРЕЖДЕНИЯ ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРА-  
ЗОВАНИЯ «НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Направление 230700 Прикладная информатика  
Уровень: магистрант

Кафедра: информационных систем

**Генерация случайных чисел с заданным законом распределения**

**Дисциплина:** Системный анализ, управление и обработка информации в ана-  
литической экономике

Выполнил: Студент гр. <u>17В30</u>	_____	<u>Т.В. Внуковский</u>
	(Подпись)	
	_____	
	(Дата)	
Проверил: Преподаватель	_____	И.О. Фамилия
	(Подпись)	
	_____	
	(Дата)	

**Юрга – 2016**

Учебное издание

Составитель МИЦЕЛЬ Артур Александрович

## **СИСТЕМНЫЙ АНАЛИЗ, УПРАВЛЕНИЕ И ОБРАБОТКА ИНФОРМАЦИИ В АНАЛИТИЧЕСКОЙ ЭКОНОМИКЕ**

Методические указания к выполнению лабораторных работ  
по дисциплине «Системный анализ, управление и обработка информации в аналитической экономике»  
для магистрантов, обучающихся по направлению 09.04.03  
«Прикладная информатика»

Печатается в редакции автора–составителя

**Отпечатано в Издательстве ЮТИ ТПУ в полном соответствии  
с качеством предоставленного оригинал–макета**