Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования

Томский государственный университет систем управления и радиоэлектроники

(ТУСУР)

# ПРИКЛАДНАЯ МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Лабораторный практикум

2015

**Прикладная математическая статистика**

Лабораторный практикум. —Томск: ТУСУР. – 2015. – 72с.

Составитель **А.А. Мицель**

В пособии приводится описание пяти лабораторных работ по основным разделам прикладной математической статистики – генерация случайных чисел с заданным законом распределения, оценка закона распределения на основе выборочных данных, дисперсионный анализ данных, корреляционный анализ случайных данных, линейная регрессия. Работы выполняются с помощью пакета EXCEL. Приводятся примеры выполнения работ.

Учебное пособие предназначено для магистрантов направления 01.04.02 «Прикладная математика и информатика», обучающихся по магистерской программе «Математическое и программное обеспечение вычислительных комплексов, систем и компьютерных сетей». Представляет интерес для инженеров, аспирантов, преподавателей, ученых, занимающихся вопросами статистической обработки данных.

## ОГЛАВЛЕНИЕ

# ЛАБОРАТОРНАЯ РАБОТА 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.

## 1.1. 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 - $\overline{x}$ 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.

e) Построить график эмпирической функцию распределения и сравнить с графиком теоретического распределения *F* (для построения графиков использовать не менее 40 точек).

f) Найти основные выборочные характеристики - $\overline{x}$ и $s^2$ и сравнить их с математическим ожиданием и дисперсией теоретического распределения *F*.

## 1.2. Пример выполнения задания

### Задание 1

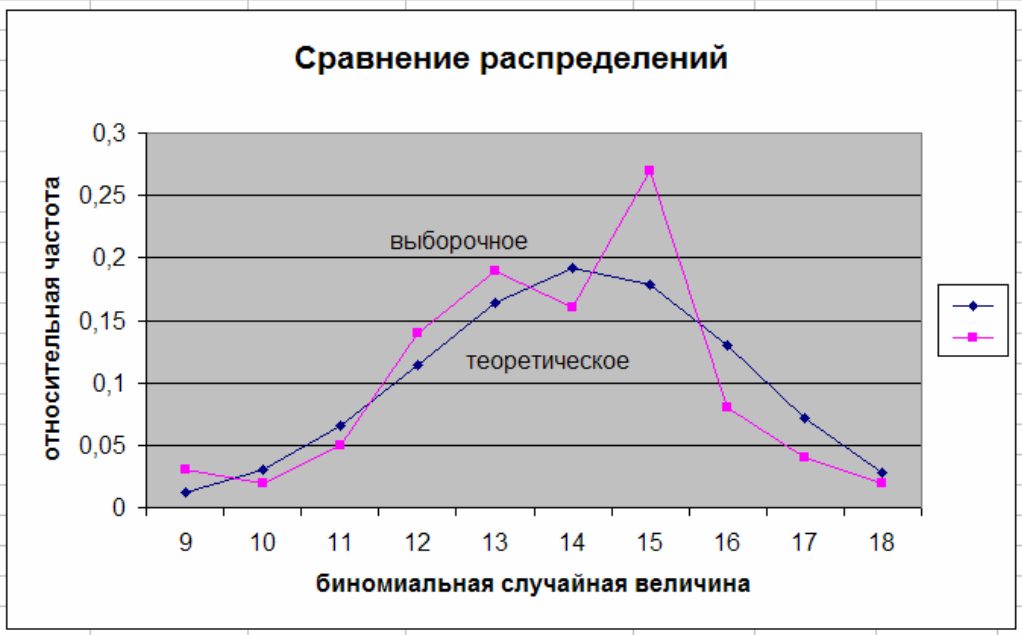
**Биномиальное распределение с параметрами n = 20 и p=0,7 .**

Количество интервалов: 9

| Выборка | Вариационный ряд | | Квадрат выборки | Ряд с абсолютными частотами | | Ряд с относительными частотами | | Ряд с накопленными частотами | |
|---|---|---|---|---|---|---|---|---|---|
| 15 | 9 | | 225 | 9 | 3 | 9 | 0,03 | 9 | 0 |
| 16 | 9 | | 256 | 10 | 2 | 10 | 0,02 | 10 | 0,05 |
| 10 | 9 | | 100 | 11 | 5 | 11 | 0,05 | 11 | 0,1 |
| 15 | 10 | | 225 | 12 | 14 | 12 | 0,14 | 12 | 0,24 |
| 14 | 10 | | 196 | 13 | 19 | 13 | 0,19 | 13 | 0,43 |
| 15 | 11 | | 225 | 14 | 16 | 14 | 0,16 | 14 | 0,59 |
| 13 | 11 | | 169 | 15 | 27 | 15 | 0,27 | 15 | 0,86 |
| 13 | 11 | | 169 | 16 | 8 | 16 | 0,08 | 16 | 0,94 |
| 14 | 11 | | 196 | 17 | 4 | 17 | 0,04 | 17 | 0,98 |
| 9 | 11 | | 81 | 18 | 2 | 18 | 0,02 | 18 | 1 |
| 14 | 12 | | 196 | | | | | | |
| 15 | 12 | | 225 | | | | | | |
| 14 | 12 | | 196 | | | | | | |
| 12 | 12 | | 144 | | | | | | |
| 13 | 12 | | 169 | | | | | | |
| 14 | 12 | | 196 | | | | | | |
| 13 | 12 | | 169 | | | | | | |
| 15 | 12 | | 225 | | | | | | |
| 13 | 12 | | 169 | | | | | | |
| 11 | 12 | | 121 | | | | | | |
| 16 | 12 | | 256 | | | | | | |
| 14 | 12 | | 196 | | | | | | |
| 15 | 12 | | 225 | | | | | | |
| 14 | 12 | | 196 | | | | | | |
| 12 | 13 | | 144 | | | | | | |
| 9 | 13 | | 81 | | | | | | |
| 15 | 13 | | 225 | | | | | | |
| 12 | 13 | | 144 | | | | | | |
| 14 | 13 | | 196 | | | | | | |
| 11 | 13 | | 121 | | | | | | |
| 13 | 13 | | 169 | | | | | | |
| 14 | 13 | | 196 | | | | | | |
| 18 | 13 | | 324 | | | | | | |
| 15 | 13 | | 225 | | | | | | |
| 14 | 13 | | 196 | | | | | | |
| 15 | 13 | | 225 | | | | | | |
| 12 | 13 | | 144 | | | | | | |
| 14 | 13 | | 196 | | | | | | |
| 15 | 13 | | 225 | | | | | | |
| 15 | 13 | | 225 | | | | | | |
| 15 | 13 | | 225 | | | | | | |
| 15 | 13 | | 225 | | | | | | |
| 15 | 13 | | 225 | | | | | | |
| 17 | 14 | | 289 | | | | | | |
| 13 | 14 | | 169 | | | | | | |
| 15 | 14 | | 225 | | | | | | |
| 12 | 14 | | 144 | | | | | | |
| 15 | 14 | | 225 | | | | | | |
| 15 | 14 | | 225 | | | | | | |
| 11 | 14 | | 121 | | | | | | |
| 14 | 14 | | 196 | | | | | | |
| 15 | 14 | | 225 | | | | | | |
| 15 | 14 | | 225 | | | | | | |
| 13 | 14 | | 169 | | | | | | |
| 11 | 14 | | 121 | | | | | | |
| 16 | 14 | | 256 | | | | | | |
| 15 | 14 | | 225 | | | | | | |
| 15 | 14 | | 225 | | | | | | |
| 17 | 14 | | 289 | | | | | | |
| 14 | 15 | | 196 | | | | | | |
| 13 | 15 | | 169 | | | | | | |
| 12 | 15 | | 144 | | | | | | |
| 15 | 15 | | 225 | | | | | | |
| 11 | 15 | | 121 | | | | | | |
| 13 | 15 | | 169 | | | | | | |
| 12 | 15 | | 144 | | | | | | |
| 13 | 15 | | 169 | | | | | | |
| 16 | 15 | | 256 | | | | | | |
| 14 | 15 | | 196 | | | | | | |
| 17 | 15 | | 289 | | | | | | |
| 18 | 15 | | 324 | | | | | | |
| 13 | 15 | | 169 | | | | | | |
| 13 | 15 | | 169 | | | | | | |
| 12 | 15 | | 144 | | | | | | |
| 10 | 15 | | 100 | | | | | | |
| 13 | 15 | | 169 | | | | | | |
| 14 | 15 | | 196 | | | | | | |
| 13 | 15 | | 169 | | | | | | |
| 12 | 15 | | 144 | | | | | | |
| 13 | 15 | | 169 | | | | | | |
| 13 | 15 | | 169 | | | | | | |
| 9 | 15 | | 81 | | | | | | |
| 12 | 15 | | 144 | | | | | | |
| 12 | 15 | | 144 | | | | | | |
| 15 | 15 | | 225 | | | | | | |
| 12 | 15 | | 144 | | | | | | |
| 16 | 16 | | 256 | | | | | | |
| 15 | 16 | | 225 | | | | | | |
| 15 | 16 | | 225 | | | | | | |
| 15 | 16 | | 225 | | | | | | |
| 15 | 16 | | 225 | | | | | | |
| 17 | 16 | | 289 | | | | | | |
| 12 | 16 | | 144 | | | | | | |
| 13 | 16 | | 169 | | | | | | |
| 12 | 17 | | 144 | | | | | | |
| 16 | 17 | | 256 | | | | | | |
| 16 | 17 | | 256 | | | | | | |
| 16 | 17 | | 256 | | | | | | |
| 14 | 18 | | 196 | | | | | | |
| 13 | 18 | | 169 | | | | | | |



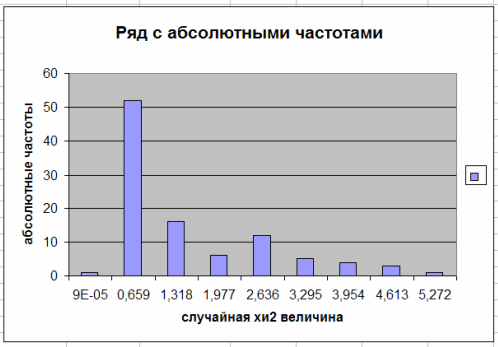
| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | | |
| | | | | | | | | |
| | факториал 20! | 2,4329E+18 | | | | | | |
| | | | | | | | | |
| Числовые характеристики выборки | | | | | биномиального распределения с параметрами n=20 и p=0,7 | | относительные частоты | |
| выборочное среднее | второй начальный выборочный момент | выборочная дисперсия | несмещенная выборочная дисперсия | | | | | |
| | | | | | 9 | 0,012007 | 0,03 | |
| 13,78 | 193,38 | 3,4916 | 3,526868687 | | 10 | 0,030817 | 0,02 | |
| M(Y)=n*p=14 | | | D(Y)=n*p*(1-p)=4,2 | | 11 | 0,06537 | 0,05 | |
| | | | | | 12 | 0,114397 | 0,14 | |
| | | | | | 13 | 0,164262 | 0,19 | |
| | | | | | 14 | 0,191639 | 0,16 | |
| | | | | | 15 | 0,178863 | 0,27 | |
| | | | | | 16 | 0,130421 | 0,08 | |
| | | | | | 17 | 0,071604 | 0,04 | |
| | | | | | 18 | 0,027846 | 0,02 | |

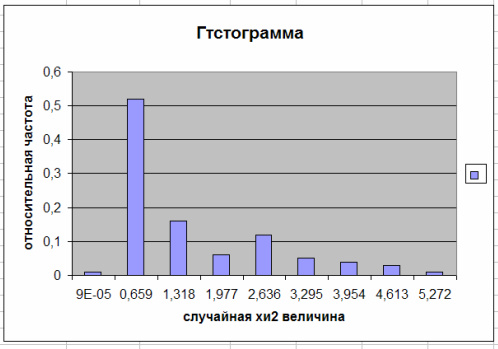


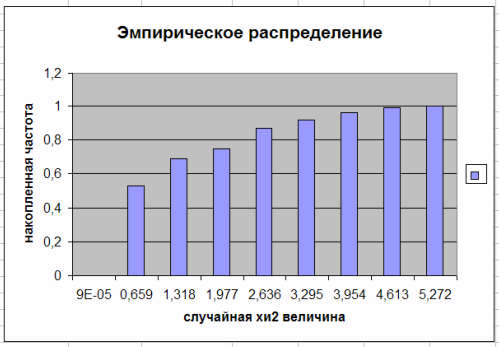

## Задание 2

ХИ - квадрат распределение с одной степенью свободы

| | | | | Количество интервалов | Шаг | | | | | | 0,659014 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | 9 | 0,6590141 | | | | | | |
| Стандартное нормальное N(0,1) | ХИ - квадрат (Выборка 1) | | Вариационный ряд (Выборка2) | Квадрат выборки | Ряд с абсолютными частотами | | Ряд с относительными частотами | | Ряд с накопленными частотами | | |
| 0,242051 | 0,058589 | | 9,36448E-05 | 0,00343261 | 9,364E-05 | 1 | 9,36448E-05 | 0,01 | 9,36E-05 | 0 | |
| 0,546748 | 0,298934 | | 0,000316393 | 0,08936144 | 0,6591077 | 52 | 0,659107708 | 0,52 | 0,659108 | 0,53 | |
| -1,89259 | 3,581881 | | 0,000415695 | 12,8298693 | 1,3181218 | 16 | 1,318121772 | 0,16 | 1,318122 | 0,69 | |
| -0,39369 | 0,154994 | | 0,003134711 | 0,02402319 | 1,9771358 | 6 | 1,977135835 | 0,06 | 1,977136 | 0,75 | |
| -0,57956 | 0,335889 | | 0,004988468 | 0,11282163 | 2,6361499 | 12 | 2,636149899 | 0,12 | 2,63615 | 0,87 | |
| -1,20086 | 1,442071 | | 0,0074712 | 2,0795694 | 3,295164 | 5 | 3,295163962 | 0,05 | 3,295164 | 0,92 | |
| -0,49782 | 0,247826 | | 0,009960946 | 0,06141761 | 3,954178 | 4 | 3,954178026 | 0,04 | 3,954178 | 0,96 | |
| -0,45761 | 0,209403 | | 0,016400305 | 0,04384951 | 4,6131921 | 3 | 4,613192089 | 0,03 | 4,613192 | 0,99 | |
| 1,499548 | 2,248643 | | 0,017018379 | 5,05639509 | 5,2722062 | 1 | 5,272206153 | 0,01 | 5,272206 | 1 | |
| -0,48797 | 0,238118 | | 0,017321677 | 0,05670023 | | | | | | | |
| 0,132538 | 0,017566 | | 0,017566425 | 0,00030858 | | | | | | | |
| -0,89304 | 0,797516 | | 0,018019537 | 0,63603143 | | | | | | | |
| -0,64942 | 0,421744 | | 0,019435043 | 0,17786764 | | | | | | | |
| 0,086436 | 0,007471 | | 0,029828225 | 5,5819E-05 | | | | | | | |
| 0,009677 | 9,36E-05 | | 0,030232094 | 8,7693E-09 | | | | | | | |
| 1,122175 | 1,259278 | | 0,043133221 | 1,58578066 | | | | | | | |
| -0,13941 | 0,019435 | | 0,044607128 | 0,00037772 | | | | | | | |
| 0,611899 | 0,37442 | | 0,058588515 | 0,14019057 | | | | | | | |
| -0,62849 | 0,395006 | | 0,061877799 | 0,15602966 | | | | | | | |
| 0,558333 | 0,311736 | | 0,06615388 | 0,09717925 | | | | | | | |
| -1,72835 | 2,987191 | | 0,068287249 | 8,92330902 | | | | | | | |
| -1,5898 | 2,527454 | | 0,088513172 | 6,38802206 | | | | | | | |
| -0,13045 | 0,017018 | | 0,122498205 | 0,00028963 | | | | | | | |
| -1,15957 | 1,344606 | | 0,127333339 | 1,80796633 | | | | | | | |
| 0,207685 | 0,043133 | | 0,133047959 | 0,00186047 | | | | | | | |
| -1,09614 | 1,201516 | | 0,13442326 | 1,44363967 | | | | | | | |
| -1,94412 | 3,779594 | | 0,150668939 | 14,2853276 | | | | | | | |
| -0,99643 | 0,992875 | | 0,154994156 | 0,98580035 | | | | | | | |
| -0,05599 | 0,003135 | | 0,167716313 | 9,8264E-06 | | | | | | | |
| -0,01779 | 0,000316 | | 0,179180371 | 1,001E-07 | | | | | | | |
| 0,409532 | 0,167716 | | 0,209402737 | 0,02812876 | | | | | | | |
| 0,423297 | 0,17918 | | 0,229965323 | 0,03210561 | | | | | | | |
| -0,75537 | 0,570586 | | 0,236189236 | 0,3255689 | | | | | | | |
| -1,57933 | 2,494278 | | 0,238118104 | 6,22142458 | | | | | | | |
| 0,785324 | 0,616733 | | 0,247825768 | 0,38035991 | | | | | | | |
| -1,13943 | 1,298298 | | 0,266618188 | 1,68557652 | | | | | | | |
| -0,13161 | 0,017322 | | 0,267251374 | 0,00030004 | | | | | | | |
| -0,9524 | 0,907075 | | 0,29227935 | 0,8227844 | | | | | | | |
| -0,6975 | 0,48651 | | 0,298933839 | 0,23669221 | | | | | | | |
| 1,448761 | 2,09891 | | 0,311735865 | 4,40542143 | | | | | | | |
| 0,866453 | 0,750741 | | 0,315342739 | 0,5636114 | | | | | | | |
| 1,952276 | 3,811381 | | 0,327876897 | 14,5266251 | | | | | | | |
| 0,625516 | 0,391271 | | 0,33588931 | 0,15309276 | | | | | | | |
| -1,31291 | 1,723721 | | 0,362474085 | 2,9712147 | | | | | | | |
| -0,48599 | 0,236189 | | 0,374420313 | 0,05578536 | | | | | | | |
| 0,257204 | 0,066154 | | 0,391270704 | 0,00437634 | | | | | | | |
| 1,117307 | 1,248376 | | 0,395005896 | 1,55844231 | | | | | | | |
| -0,57261 | 0,327877 | | 0,421743571 | 0,10750326 | | | | | | | |
| -1,46781 | 2,154468 | | 0,433513368 | 4,64173423 | | | | | | | |
| -1,48257 | 2,198019 | | 0,436272348 | 4,83128894 | | | | | | | |
| -1,25215 | 1,567872 | | 0,486510233 | 2,45822162 | | | | | | | |
| -1,49275 | 2,228313 | | 0,57058645 | 4,96538093 | | | | | | | |
| 2,104434 | 4,428644 | | 0,616733257 | 19,6128858 | | | | | | | |
| -0,56155 | 0,315343 | | 0,69764319 | 0,09944104 | | | | | | | |
| -0,35 | 0,122498 | | 0,750740567 | 0,01500581 | | | | | | | |
| -0,66051 | 0,436272 | | 0,797515789 | 0,19033356 | | | | | | | |
| -0,12806 | 0,0164 | | 0,90707464 | 0,00026897 | | | | | | | |
| -1,39273 | 1,939702 | | 0,914443643 | 3,76244521 | | | | | | | |
| 0,172708 | 0,029828 | | 0,956484228 | 0,00088972 | | | | | | | |
| -0,38816 | 0,150669 | | 0,95962669 | 0,02270113 | | | | | | | |
| 0,248752 | 0,061878 | | 0,992874791 | 0,00382886 | | | | | | | |
| 0,366638 | 0,134423 | | 1,047426305 | 0,01806961 | | | | | | | |
| 1,545623 | 2,388949 | | 1,140860206 | 5,7070781 | | | | | | | |
| 0,978 | 0,956484 | | 1,153660645 | 0,91486208 | | | | | | | |
| -1,52133 | 2,314431 | | 1,201515571 | 5,35659115 | | | | | | | |
| -0,51635 | 0,266618 | | 1,248375871 | 0,07108526 | | | | | | | |
| -0,60206 | 0,362474 | | 1,259277831 | 0,13138746 | | | | | | | |
| -0,17387 | 0,030232 | | 1,298297545 | 0,00091398 | | | | | | | |
| 1,074086 | 1,153661 | | 1,315781472 | 1,33093288 | | | | | | | |
| -1,65061 | 2,724499 | | 1,344606385 | 7,42289315 | | | | | | | |
| -1,79897 | 3,236298 | | 1,44207122 | 10,473626 | | | | | | | |
| 1,741437 | 3,032602 | | 1,456518517 | 9,19667546 | | | | | | | |
| -0,35684 | 0,127333 | | 1,567871686 | 0,01621378 | | | | | | | |
| 0,479547 | 0,229965 | | 1,723721178 | 0,05288405 | | | | | | | |
| 0,020389 | 0,000416 | | 1,939702352 | 1,728E-07 | | | | | | | |
| -2,12665 | 4,522654 | | 1,997713097 | 20,4544004 | | | | | | | |
| 1,702692 | 2,899161 | | 2,087566882 | 8,40513522 | | | | | | | |
| -1,92037 | 3,687824 | | 2,09890958 | 13,6000453 | | | | | | | |
| 0,83525 | 0,697643 | | 2,129643243 | 0,48670602 | | | | | | | |
| -1,45933 | 2,129643 | | 2,154468433 | 4,53538034 | | | | | | | |
| -0,36476 | 0,133048 | | 2,198019323 | 0,01770176 | | | | | | | |
| -1,4134 | 1,997713 | | 2,228313472 | 3,99085762 | | | | | | | |
| 2,134802 | 4,557381 | | 2,248642943 | 20,769722 | | | | | | | |
| -0,97961 | 0,959627 | | 2,314431064 | 0,92088338 | | | | | | | |
| -1,06811 | 1,14086 | | 2,388949163 | 1,30156201 | | | | | | | |
| 0,099805 | 0,009961 | | 2,494278368 | 9,922E-05 | | | | | | | |
| -0,95627 | 0,914444 | | 2,52745367 | 0,83620718 | | | | | | | |
| 1,023438 | 1,047426 | | 2,724498696 | 1,09710187 | | | | | | | |
| 0,070629 | 0,004988 | | 2,899161123 | 2,4885E-05 | | | | | | | |
| 0,261318 | 0,068287 | | 2,987190824 | 0,00466315 | | | | | | | |
| 0,658417 | 0,433513 | | 3,032602095 | 0,18793384 | | | | | | | |
| 0,134237 | 0,01802 | | 3,236298189 | 0,0003247 | | | | | | | |
| 1,444841 | 2,087567 | | 3,581880688 | 4,35793549 | | | | | | | |
| -0,2112 | 0,044607 | | 3,687823925 | 0,0019898 | | | | | | | |
| 0,297512 | 0,088513 | | 3,779593572 | 0,00783458 | | | | | | | |
| -2,29613 | 5,272206 | | 3,811381003 | 27,7961577 | | | | | | | |
| -0,51696 | 0,267251 | | 4,428643789 | 0,0714233 | | | | | | | |
| -1,14708 | 1,315781 | | 4,522654136 | 1,73128088 | | | | | | | |
| 0,540629 | 0,292279 | | 4,557381047 | 0,08542722 | | | | | | | |
| 1,206863 | 1,456519 | | 5,272206153 | 2,12144619 | | | | | | | |

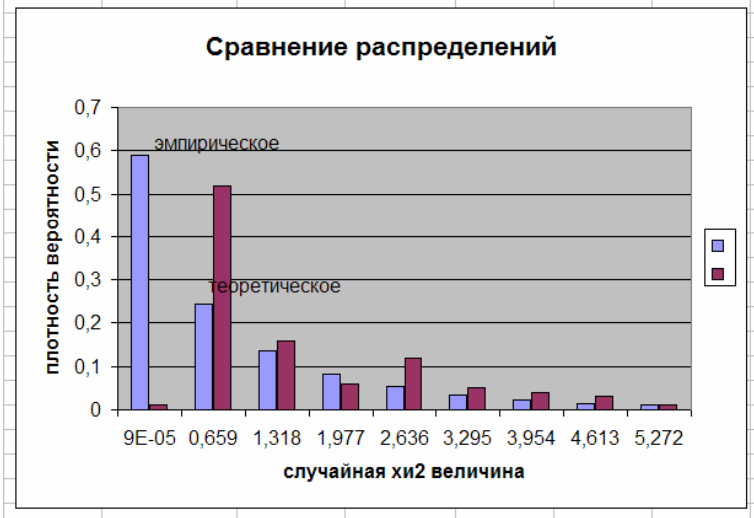


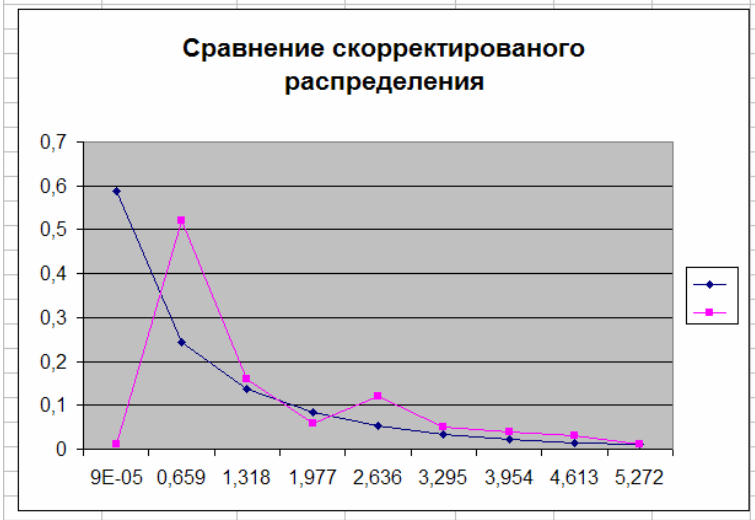



2,718281828

| Числовые характеристики выборки | | | | | Формула ХИ-квадрат распределения с одной степенью свободы в точках x+(шаг/2) | | относительные частоты |
|---|---|---|---|---|---|---|---|
| выборочное среднее | второй начальный выборочный момент | выборочная дисперсия | несмещенная выборочная дисперсия | | | | |
| 1,11365 | 2,85146383 | 1,611248 | 1,6275232 | | 9,36E-05 | 0,589311 | 0,01 |
| | | | | | 0,659108 | 0,244749 | 0,52 |
| M(ХИ2)=1 | | | D(ХИ2)=2 | | 1,318122 | 0,136365 | 0,16 |
| | | | | | 1,977136 | 0,082897 | 0,06 |
| | | | | | 2,63615 | 0,052585 | 0,12 |
| | | | | | 3,295164 | 0,034213 | 0,05 |
| | | | | | 3,954178 | 0,022637 | 0,04 |
| | | | | | 4,613192 | 0,015158 | 0,03 |
| | | | | | 5,272206 | 0,010241 | 0,01 |



# Приложение 1 к лабораторной работе 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 $f(x) = 1/(\pi(1 + x^2)), \quad x \in R$.

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 к лабораторной работе 1

# Генерация случайных чисел в 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”.

# Приложение 3 к лабораторной работе 1

# Статистические функции пакета EXEL

## 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:

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

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**:** ПУАССОН (x; λ; 0) = $\lambda^x e^{-\lambda} / x!$: ПУАССОН (x; λ; 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: $P = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}$ = ГИПЕРГЕОМЕТ(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 (σ).

***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 (1), то возвращается плотности распределения от аргумента $x$.

Таким образом:

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

НОРМРАСП(x; a; σ; 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. НОРМСТРАСП(z) $= P(\xi < z) = F_{\xi}(z)$, где $\xi \in N_{0,1}$, $F_{\xi}(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$.

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

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

*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 НОРМСТОБР($\alpha$) возвращает значение $\tau_\alpha$, для которого $P(\xi < \tau_\alpha) = \alpha, \quad \xi \in N_{0,1}$

***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. 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. 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 $P(\xi < \tau_{1-\alpha/2}) = 1 - \alpha/2, \quad \xi \in T_n$.

***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. 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 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.

- **FРАСП**(**x**; **степени свободы1**; **степени свободы2**)

Возвращает вероятность того, что случайная величина, распределенная по закону Фишера примет значение, большее, чем $x$, т.е. FРАСП(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 свободы1, степени свободы2*** — число степеней свободы, характеризующих распределение.

- **FРАСПОБР**(**вероятность**; **степени свободы1**; **степени свободы2**)

возвращает критическую точку распределения Фишера для заданной вероятности, то есть FРАСПОБР($\alpha, n1, n2$)=$t_{\alpha}$, 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.

# Приложение 4 к лабораторной работе 1

## Примеры использования функций EXCEL

**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:

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

Найдем приближенное значение вероятности, используя локальную формулу Муавра-Лапласа. Согласно этой формуле, вероятность $P_n(m) \approx \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{npq}} e^{-\frac{(m-np)^2}{2npq}}$, 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: $P_{24}(3) \approx$ 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 $P_n(m) \approx \frac{\lambda^m e^{-\lambda}}{m!}$, $\lambda = np$, т.е. приближенно равна вероятности пуассоновского 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: $P_{24}(3) \approx$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 передачи 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:

$P_{200}(0 \le m \le 2)$ =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** равном **ЛОЖЬ (0)**. Таким образом: $P_{200}(0 \le m \le 2) = P_{200}(0) + P_{200}(1) + P_{200}(2) \approx$

$\approx$НОРМРАСП$(0;\ 2;\ \sqrt{1{,}98};\ 0)$+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: $P_{200}(0 \le m \le 2) \approx$ПУАССОН$(2;\ 200 \cdot 0{,}01;\ 1)$=0,676676.

Заметим, что погрешность, полученная при использования 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 $(0 \le m \le 2)$.

Найдем приближенное значение вероятности, используя теперь интегральную формулу Муавра-Лапласа. Согласно этой формуле, вероятность $P_n(m_1 \le m \le m_2) \approx \Phi\left(\frac{m_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{m_1 - np}{\sqrt{npq}}\right)$, 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 $P_n(m_1 \le m \le m_2) \approx F_\xi(m_2) - F_\xi(m_1)$. Значения функции распределения нормальной величины возвращает функция **НОРМРАСП,** при значении параметра **интегральная** равном **ИСТИНА (1)**. Таким образом: $P_n(0 \le m \le 2) \approx$ 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: $P_{10000}(m > 5100) = 1 - P_{10000}(m \le 5100) =$ 1- БИНОМРАСП(5100;10000;1/2;1)=0,022213.

Найдем приближенное значение вероятности, используя интегральную формулу Муавра-Лапласа. Согласно этой формуле, вероятность $P_n(m_1 \le m \le m_2) \approx \Phi\left(\frac{m_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{m_1 - np}{\sqrt{npq}}\right)$, 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 $P_n(m_1 \le m \le m_2) \approx F_\xi(m_2) - F_\xi(m_1)$.

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

$P_{10000}(m > 5100) = 1 - P_{10000}(m \le 5100) \approx 1 - F_\xi(5100) =$ НОРМРАСП(

$5100; 10000 \cdot 1/2; \sqrt{10000 \cdot 1/2 \cdot 1/2}; 1)$ )=0,02275.

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

$P_{10000}(m > 5100) = 1 - P_{10000}(m \le 5100) \approx 1 - F_\xi(5100 + 1/2) =$

=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

$P(1 < \xi < 4) = F_\xi(4) - F_\xi(1) =$

=НОРМРАСП( $4; 3; \sqrt{7}; 1) -$ 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 НОРМОБР( $\beta; a; \sigma$ ) =

$\tau_\beta$, для которого $P(\xi < \tau_\beta) = \beta, \quad \xi \in N_{a,\sigma^2}$. Таким образом, квантиль уровня 0,85 равна $\tau_{0,85}$ = 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 $t_{0,07} = \tau_{0,93}$ = 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: $\tau_{0,975}$ = 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

# Оценка закона распределения на основе выборочных данных

***Цель работы:***

Оценка закона распределения генеральной совокупности на основе выборочных данных.

## 2.1. Необходимые теоретические сведения

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

Пусть $\{X_1, X_2, \ldots, X_n\}$ 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 $\rho(\vec{X}) = \sum_{i=1}^{m} \frac{(n_i - np_i)^2}{np_i}$, 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 $n \to \infty$

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

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

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

где $\tau_{1-\alpha}$-квантиль распределения $\chi^2_{m-1}$.

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 $i \in \{1,2,\ldots,m\}$. Поэтому, следует стремиться к как можно большему 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 $np_i \geq 5$.

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

Пусть $\{X_1, X_2, \ldots, X_n\}$ 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, $i \in \{1,2,\ldots,m\}$. 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

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

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

$$\rho(\vec{X}) = \sum_{i=1}^{m} \frac{\left(n_i - np_i^*\right)^2}{np_i^*} \Rightarrow \chi^2_{m-l-1}. \qquad (3)$$

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

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

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

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

Поэтому желательно уточнить оценки, найденные другим способом (методом максимального правдоподобия или методом моментов) путем минимизации $\rho(\vec{X})$.

### 2.1.3. Метод моментов оценки параметров распределения

Идея этого метода заключается в приравнивании теоретических и эмпирических моментов.

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

$$\mu_i(\theta_1, ..., \theta_k) = E[x^i] = \begin{cases} \int x^i f(x, \theta_1, ..., \theta_k) dx, & \text{если } x \text{ непрерывная величина} \\ \sum_j x_j^i p(x_j, \theta_1, ..., \theta_k), & \text{если } x \text{ дискретная величина} \end{cases}$$

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

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

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

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

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

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

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

$$\hat{\theta}_1 = \hat{\theta}_1(x_1,...,x_n),$$
$$\hat{\theta}_2 = \hat{\theta}_2(x_1,...,x_n),$$
$$.....,$$
$$\hat{\theta}_k = \hat{\theta}_k(x_1,...,x_n).$$

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

**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

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

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.

## 2.1.4. Метод максимального правдоподобия

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

$$L(\theta,x)=\begin{cases}\prod_{j=1}^{n} f(x_j,\theta_1,...,\theta_k), & \textit{если } x \textit{ непрерывная величина}\\ \prod_{j=1}^{n} p(x_j,\theta_1,...,\theta_k), & \textit{если } x \textit{ дискретная величина}\end{cases}$$

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

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

**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

$L(\alpha,x)=\prod_{j=1}^{n}\alpha e^{-\alpha x_j}=\alpha^n e^{-\alpha\sum_{j=1}^{n}x_j}$. Из условие 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, полученной 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

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

$$L(\alpha,\beta)=\begin{cases}\frac{1}{(\beta-\alpha)^n}, & \alpha\le\tilde{x}_1\le,\tilde{x}_2\le\ldots\le\tilde{x}_n\le\beta\\ 0, & \widetilde{X}\in[\alpha,\beta]\end{cases}$$, где $\widetilde{X}=(\tilde{x}_1,\ldots,\tilde{x}_n)$ – упорядоченная выборка (вариационный ряд).

Проанализируем неравенства $\alpha\le\tilde{x}_1\le,\tilde{x}_2\le\ldots\le\tilde{x}_n\le\theta$. 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).

## 2.1.5. Метод наименьших квадр*атов*

Пусть дана табличная функция $y(x)$

| $x$ | $x_1$ | $x_2$ | $\cdots$ | $x_n$ |
|---|---|---|---|---|
| $y$ | $y_1$ | $y_2$ | $\cdots$ | $y_n$ |

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

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

Параметры $\theta_1,...,\theta_k$ будем подбирать таким образом, чтобы расхождение табличной функции с функцией $f(\theta_1,...,\theta_k;x)$ было минимальным. Для этого построим функционал $F(\theta_1,...,\theta_k)=\sum_{j=1}^{n}\left(y_i-f(\theta_1,...,\theta_k;x_j)\right)^2$ и найдем его минимум.

Необходимое условие минимума имеет вид

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

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

**Пример 5.** Пусть дана независимая выборка из $X=(x_1,x_2,...,x_n)$ из распределения $P_\theta$. Разобьем весь диапазон данных $[x_{\min},x_{\max}]$ на $m$ интервалов и построим гистограмму. Обозначим середины интервалов $\tilde{x}_1,\tilde{x}_2,...,\tilde{x}_m$. 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 $f(\theta_1,...,\theta_k;x)$ так, чтобы расхождение между гистограммой и функцией $f(\theta,x)$ было минимально. В результате мы приходим к методу наименьших квадратов.

## 2.2. 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 $f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}$, $x\in R$, где параметры $a$ и $\sigma>0$ - 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: $m_1=\int\limits_{-\infty}^{\infty}x\frac{1}{\sigma\sqrt{2}}e^{-\frac{|x-a|\sqrt{2}}{\sigma}}dx=a$,

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

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

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

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

3) $F$ - распределение $\chi^2$ с $k$ степенями свободы, где параметр $k$ – неизвестен;

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) $F$ - распределение с плотностью $f(x)=\frac{2a}{(1+ax)^3}$, $x\in[0,\infty)$, где параметр $a>0$ неизвестен.

7) $F$ - распределение Стьюдента $t_k$ с $k$ степенями свободы, где параметр $k$ – неизвестен.

**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 $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$, выборочную дисперсию $s^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$, выборочный коэффициент асимметрии $\overline{A}=\frac{1}{(n-1)s^3}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^3$ и выборочный коэффициент эксцесса $\overline{E}=\frac{1}{(n-1)s^4}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^4-3$.

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 $\overline{A}$ и $\overline{E}$ 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 $\rho(\vec{X}) = \sum_{i=1}^{m} \frac{\left(n_i - np_i^*\right)^2}{np_i^*}$)

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

## 2.3. 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:

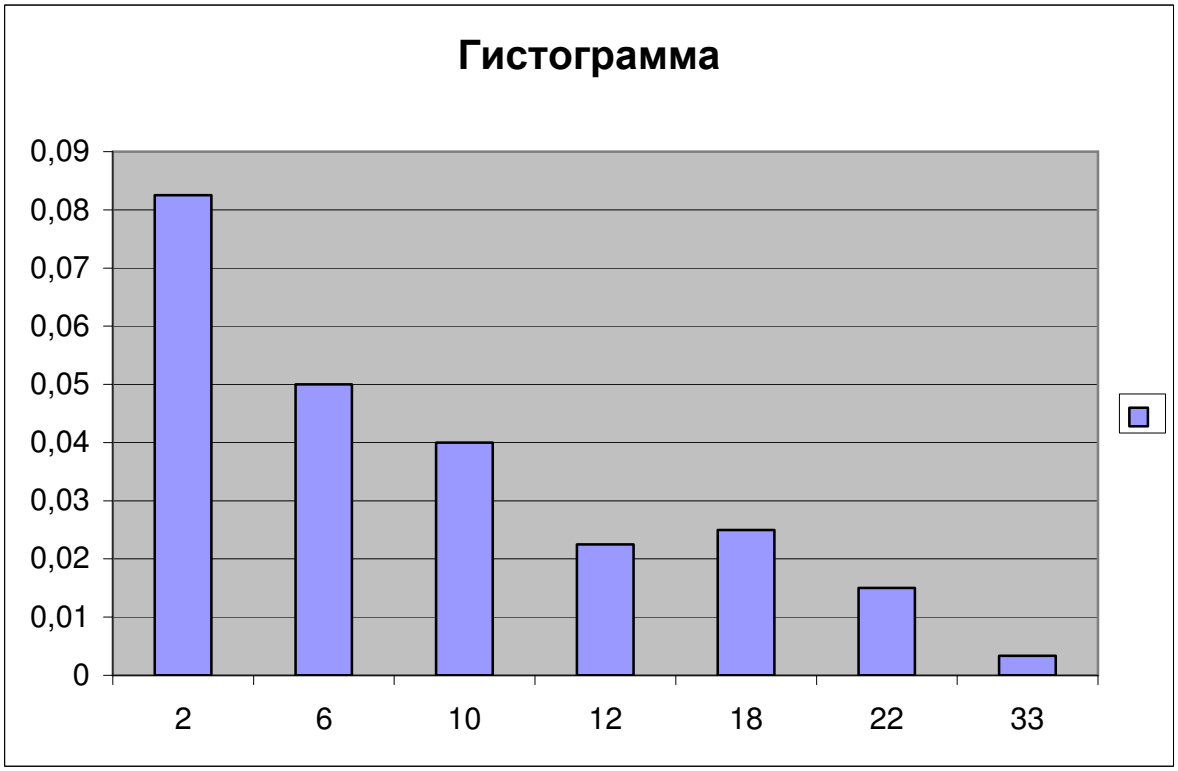
| Интервал $\tilde{x}_{i-1} - \tilde{x}_i$ | 0 - 6 | 6 - 12 | 12 - 18 | 18 - 24 | 24 - 30 | 30 - 36 | 36 - 42 |
|---|---|---|---|---|---|---|---|
| Частота | 46 | 23 | 14 | 11 | 3 | 0 | 3 |

| $n_i$ | | | | | | | |
|---|---|---|---|---|---|---|---|

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 |

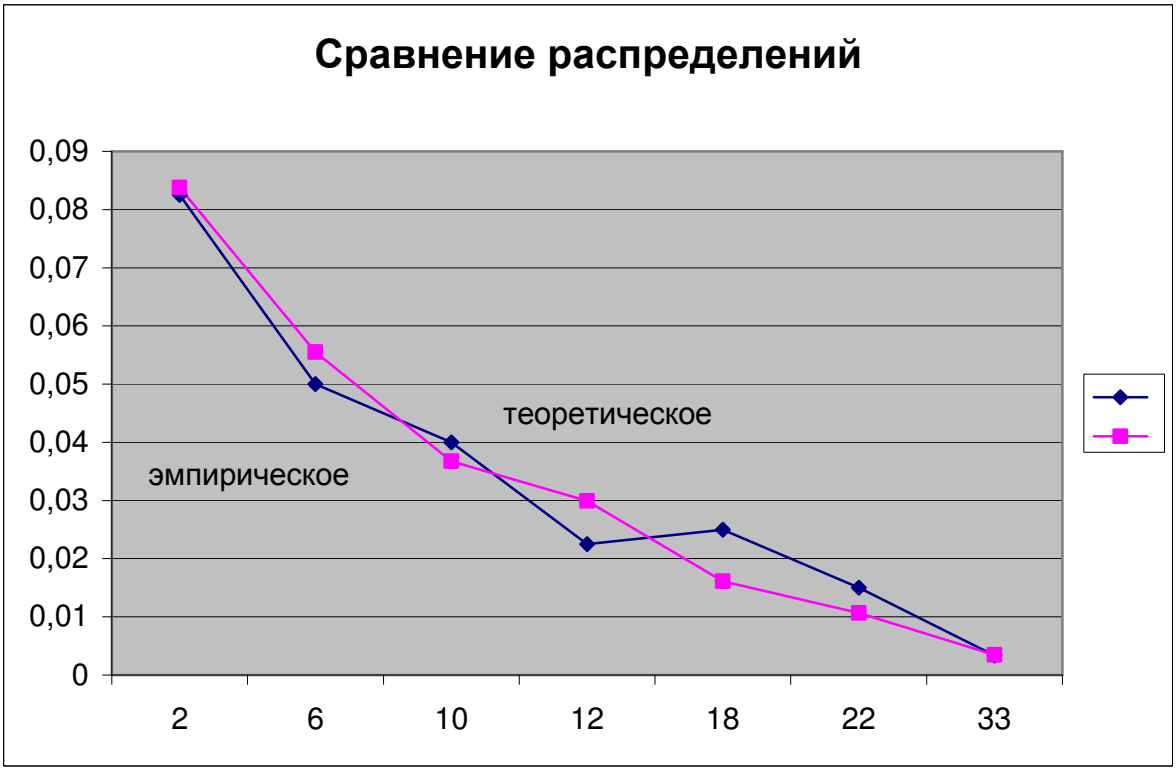
Соответствующая гистограмма приведена на рисунке.




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 |




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

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

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

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 $\lambda^* = 0{,}103$: $p_i^* = F(\tilde{x}_i) - F(\tilde{x}_{i-1}),\quad i = 1{,}7$, $F(x) = 1 - e^{-\lambda^* x}$ - 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:

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

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): $\rho_{кр} = 11{,}07$. (Критическую точку заданного уровня можно получить, например, используя функцию пакета EXCEL – ХИ2ОБР).

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

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

# Приложение к лабораторной работе 2

## Варианты заданий

Вариант 1.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 5,905 | 2,089 | -0,54 | -1,76 | -0,32 | 2,653 | 2,591 | -0,77 | -8,72 | -6,09 |
| 2,752 | 8,509 | 3,658 | 3,583 | -0,91 | 0,96 | -3,7 | 6,593 | -2,09 | 0,058 |
| 1,013 | -9,79 | 0,862 | 1,67 | 4,889 | 1,696 | 6,856 | -6,17 | 1,606 | 4,305 |
| -2,78 | -0,23 | 1,763 | -6,66 | -5,1 | 12,8 | -2,02 | -1,87 | -5,89 | 4,256 |
| 5,554 | -2,16 | 4,859 | -0,39 | 3,725 | -2,08 | 5,382 | -0,66 | 5,4 | 5,062 |
| 0,673 | -1,57 | 5,876 | 4,409 | 3,775 | 1,657 | -0,52 | -7,42 | 11,83 | 9,073 |
| 7,574 | 5,853 | 12,65 | 1,74 | 0,566 | 6,627 | 2,047 | 3,326 | -6,31 | 1,257 |
| 4,689 | -1,38 | -3,85 | 9,836 | 6,439 | 6,108 | 6,025 | -0,55 | 6,035 | -3,47 |
| 2,015 | -1,23 | 5,246 | 3,273 | 7,755 | 4,351 | -0,19 | 6,378 | 4,582 | 5,176 |
| 13 | -2,34 | -7,11 | 4,14 | -2,56 | -1,05 | 17,58 | 3,33 | 2,854 | 0,502 |

Вариант 2. $\mu = 5,\ \sigma = 5$

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 11,57 | 16,66 | -1,8 | 2,817 | 8,19 | 3,297 | 6,599 | 9,933 | 4,618 | 5,749 |
| 4,484 | 3,538 | 9,919 | 15,14 | 16,08 | 3,546 | 1,867 | 10,39 | 4,792 | 5,92 |
| 3,517 | 2,766 | -4,37 | 9,782 | 7,036 | -0,67 | 4,359 | -5,5 | -3,58 | 2,481 |
| 3,621 | 2,673 | -3,18 | 6,431 | 1,579 | 0,016 | -0,03 | 11,86 | -5,38 | 0,475 |
| 12,69 | 3,897 | 9,73 | 3,518 | 10,26 | 11,75 | 0,76 | 10,46 | -2,09 | 13,61 |
| 8,939 | 2,018 | 2,174 | -1,26 | 7,532 | 4,902 | 9,171 | -1,37 | -2,85 | 4,996 |
| 18,58 | 1,85 | -5,37 | -1,97 | 3,929 | 10,1 | 8,714 | 2,423 | 3,798 | 6,924 |
| 9,802 | 2,864 | 9,058 | 7,484 | 0,482 | 6,848 | 4,001 | 0,367 | 4,394 | -1,69 |
| -7,31 | 6,267 | 8,759 | 13,07 | 0,468 | 0,8 | 5,101 | 3,599 | 0,057 | 6,368 |
| 4,374 | 4,74 | 12,03 | 12,15 | 7,708 | 0,165 | 2,382 | -4,08 | 10,82 | 3,183 |
| 11,57 | 16,66 | -1,8 | 2,817 | 8,19 | 3,297 | 6,599 | 9,933 | 4,618 | 5,749 |

Вариант 3.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 5,834 | 4,553 | 5,137 | 4,795 | 5,551 | 3,882 | 5,396 | 4,171 | 4,32 | 4,763 |
| 4,848 | 4,049 | 6,429 | 2,418 | 5,556 | 4,58 | 6,551 | 3,297 | 4,812 | 4,764 |
| 6,645 | 5,743 | 5,584 | 5,988 | 4,938 | 4,836 | 6,421 | 6,496 | 4,378 | 4,514 |
| 3,64 | 4,595 | 4,757 | 5,152 | 3,684 | 6,371 | 4,729 | 2,867 | 5,954 | 4,37 |
| 3,503 | 5,389 | 4,972 | 7,599 | 5,121 | 5,801 | 5,066 | 6,875 | 3,533 | 5,73 |
| 4,452 | 5,072 | 6,532 | 4,174 | 6,16 | 3,592 | 4,709 | 5,009 | 4,174 | 6,608 |
| 5,418 | 5,76 | 6,265 | 6,254 | 3,489 | 5,932 | 6,627 | 4,303 | 4,342 | 4,605 |
| 5,625 | 6,619 | 5,25 | 2,918 | 4,729 | 5,234 | 5,999 | 5,277 | 3,761 | 5,45 |
| 6,173 | 5,783 | 4,515 | 3,626 | 6,41 | 5,683 | 4,688 | 4,226 | 5,862 | 4,82 |
| 5,722 | 5,224 | 5,777 | 4,264 | 4,273 | 4,646 | 5,003 | 4,203 | 3,643 | 5,644 |

Вариант 4.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1,048 | -3,67 | 2,758 | 7,659 | 7,184 | 10,8 | -10,5 | 1,276 | 6,584 | -2,54 |
| -0,52 | -6,48 | -7,68 | -1,94 | -0,91 | -9,94 | 0,013 | 0,668 | 2,401 | 0,813 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0,953 | 0,795 | 8,075 | 1,751 | 1,435 | 0,24 | 12,69 | 5,359 | 16,3 | -0,36 |
| 10,26 | -5,91 | 3,866 | 5,544 | 12,25 | 1,753 | 0,196 | 4,454 | 0,754 | 4,834 |
| -4,74 | -1,27 | -5,26 | 0,822 | 1,907 | 2,08 | 0,968 | 14,62 | -6,87 | -0,73 |
| -14,3 | 8,762 | -3,68 | -0,36 | 4,834 | 3,574 | 5,404 | 4,105 | -4,26 | -2,7 |
| 4,539 | 3,031 | -1,74 | 1,253 | 2,391 | 3,944 | 2,413 | -1,59 | 11,98 | 3,655 |
| 2,21 | 5,181 | 5,341 | -0,28 | -1,65 | 6,675 | -3,2 | -5,53 | 4,618 | 4,29 |
| 14,72 | 8,736 | 7,83 | 2,333 | 2,006 | 3,523 | 1,927 | -2,36 | -7,11 | 5,174 |
| 3,486 | 4,201 | 2,655 | -2,2 | 7,424 | 1,009 | -1,23 | -1,14 | 0,573 | 0,479 |

Вариант 5.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 2,42 | 0,788 | 0,66 | 5,459 | 2,396 | 0,917 | 0,377 | 2,752 | 0,453 | 1,511 |
| 0,352 | 4,798 | 0,044 | 3,608 | 5,629 | 2,412 | 1,659 | 0,726 | 1,764 | 1,278 |
| 2,422 | 1,383 | 0,032 | 7,04 | 0,254 | 3,33 | 1,434 | 2,754 | 0,917 | 0,813 |
| 1,593 | 2,598 | 0,794 | 0,334 | 6,728 | 1,69 | 2,73 | 0,646 | 0,553 | 0,65 |
| 0,646 | 2,423 | 1,323 | 1,381 | 0,483 | 1,808 | 0,268 | 1,532 | 0,546 | 0,283 |
| 1,212 | 2,21 | 0,949 | 2,988 | 0,101 | 1,786 | 3,697 | 0,449 | 0,799 | 0,569 |
| 5,504 | 0,343 | 0,723 | 0,051 | 2,92 | 1,767 | 4,216 | 0,161 | 2,974 | 2,194 |
| 5,238 | 0,616 | 0,31 | 0,311 | 0,246 | 0,138 | 2,806 | 1,484 | 0,038 | 3,409 |
| 1,494 | 0,733 | 0,591 | 0,807 | 2,051 | 4,206 | 1,263 | 4,708 | 0,31 | 0,107 |
| 3,186 | 1,558 | 0,604 | 3,291 | 2,212 | 5,687 | 0,376 | 2,334 | 1,395 | 3,449 |

Вариант 6.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 12,01 | 2,624 | 12,25 | 4,08 | 6,967 | 10,87 | 8,041 | 1,358 | 11,33 | 11,23 |
| 2,443 | 1,02 | 3,705 | 1,513 | 9,882 | 4,937 | 5,467 | 3,006 | 3,774 | 12,47 |
| 6,24 | 5,373 | 12,1 | 9,412 | 5,302 | 12,83 | 6,902 | 7,364 | 8,433 | 5,991 |
| 4,473 | 5,995 | 12,29 | 5,812 | 4,649 | 7,701 | 4,319 | 8,52 | 3,752 | 5,938 |
| 5,407 | 6,141 | 11,24 | 3,921 | 8,604 | 9,861 | 6,624 | 6,298 | 5,548 | 2,335 |
| 3,084 | 1,435 | 5,629 | 9,076 | 3,956 | 10,85 | 2,228 | 3,806 | 1,473 | 1,944 |
| 7,821 | 4,522 | 3,858 | 4,755 | 0,761 | 11,33 | 9,863 | 9,687 | 6,851 | 4,336 |
| 3,084 | 4,554 | 2,377 | 16,67 | 4,094 | 5,525 | 6,794 | 4,622 | 6,387 | 2,501 |
| 8,887 | 5,228 | 10,69 | 10,92 | 5,932 | 5,79 | 8,326 | 3,489 | 11,07 | 5,821 |
| 4,778 | 6,231 | 3,806 | 1,759 | 4,977 | 10,17 | 3,962 | 3,712 | 3,978 | 6,596 |

Вариант 7.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0,392 | 0,162 | 0,318 | 0,601 | 0,301 | 0,033 | 0,303 | 0,05 | 0,234 | 0,007 |
| 0,435 | 0,347 | 0,162 | 0,028 | 0,095 | 0,081 | 0,006 | 0,352 | 0,196 | 0,178 |
| 0,062 | 0,018 | 0,388 | 0,006 | 0,013 | 0,291 | 0,084 | 0,329 | 0,181 | 0,514 |
| 0,193 | 0,243 | 0,322 | 0,089 | 0,606 | 0,105 | 0,244 | 0,053 | 0,213 | 0,029 |
| 0,221 | 0,634 | 0,126 | 0,034 | 0,494 | 0,059 | 0,185 | 0,061 | 0,062 | 0,154 |
| 0,209 | 0,237 | 0,141 | 0,137 | 0,705 | 0,162 | 0,051 | 0,024 | 0,294 | 0,059 |
| 0,201 | 0,524 | 0,123 | 0,004 | 0,04 | 0,297 | 0,144 | 0,666 | 0,101 | 0,378 |
| 0,112 | 0,345 | 0,486 | 0,081 | 0,944 | 0,308 | 1,122 | 0,088 | 0,126 | 0,162 |
| 0,167 | 0,197 | 0,44 | 0,019 | 0,01 | 0,117 | 0,023 | 0,022 | 0,004 | 0,109 |
| 0,266 | 0,369 | 2E-04 | 0,247 | 0,217 | 0,034 | 0,091 | 0,557 | 0,07 | 0,109 |

Вариант 8.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 3,146 | 2,302 | 3,789 | 4,697 | 4,654 | 4,875 | 2,043 | 3,222 | 4,59 | 2,416 |
| 2,735 | 2,136 | 2,097 | 2,492 | 2,659 | 2,051 | 2,855 | 3,029 | 3,661 | 3,072 |
| 3,116 | 3,067 | 4,731 | 3,398 | 3,278 | 2,912 | 4,927 | 4,42 | 4,974 | 2,769 |
| 4,855 | 2,16 | 4,115 | 4,45 | 4,918 | 3,399 | 2,901 | 4,251 | 3,054 | 4,327 |
| 2,223 | 2,595 | 2,192 | 3,075 | 3,461 | 3,534 | 3,12 | 4,958 | 2,122 | 2,692 |
| 2,015 | 4,778 | 2,301 | 2,77 | 4,327 | 4,039 | 4,427 | 4,173 | 2,255 | 2,397 |
| 4,268 | 3,88 | 2,521 | 3,214 | 3,657 | 4,135 | 3,665 | 2,543 | 4,911 | 4,061 |
| 3,586 | 4,39 | 4,417 | 2,787 | 2,534 | 4,6 | 2,345 | 2,179 | 4,285 | 4,215 |
| 4,959 | 4,777 | 4,712 | 3,635 | 3,502 | 4,025 | 3,469 | 2,437 | 2,114 | 4,389 |
| 4,015 | 4,195 | 3,754 | 2,457 | 4,677 | 3,133 | 2,601 | 2,617 | 3,002 | 2,975 |

Вариант 9.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0,0768 | 0,3993 | 0,0053 | 0,6394 | 0,5534 | 0,0084 | 0,0906 | 0,0455 | 0,0652 | 0,0141 |
| 0,581 | 0,2629 | 0,2624 | 0,2419 | 0,1884 | 0,0338 | 0,0082 | 0,09 | 0,2067 | 0,0293 |
| 0,002 | 0,0229 | 0,0063 | 0,9865 | 0,0192 | 0,6063 | 0,1188 | 1,7134 | 0,0218 | 0,0947 |
| 0,0226 | 0,7813 | 0,0624 | 0,071 | 0,0848 | 0,0138 | 0,1625 | 1,7929 | 0,1949 | 0,0108 |
| 0,0106 | 0,1345 | 0,1473 | 0,0371 | 0,2595 | 0,0204 | 0,0975 | 0,1808 | 0,1811 | 0,0641 |
| 0,3089 | 0,1209 | 0,7202 | 0,0553 | 0,0096 | 0,0458 | 0,4529 | 0,1155 | 0,2791 | 0,0311 |
| 0,01 | 0,0547 | 0,0053 | 1,0792 | 0,0381 | 0,6262 | 0,0044 | 0,0981 | 0,0344 | 0,3127 |
| 1,7188 | 0,0196 | 0,1464 | 1,5181 | 0,0419 | 0,2705 | 0,0299 | 0,0678 | 0,0155 | 0,0996 |
| 0,0149 | 0,0559 | 0,1738 | 0,2022 | 0,0301 | 0,084 | 0,0451 | 0,0083 | 0,0752 | 0,587 |
| 0,0942 | 0,0718 | 0,4102 | 0,0504 | 0,0067 | 0,3332 | 0,1083 | 0,0194 | 0,0679 | 0,0133 |

Вариант 10.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| -0,428 | 5,396 | -1,678 | -0,768 | -0,847 | -1,328 | 2,5249 | -2,425 | -1,025 | -2,896 |
| -0,276 | -1,278 | -0,092 | -0,965 | -6,615 | -0,158 | -0,335 | -0,417 | -0,602 | -0,094 |
| 0,3655 | -1,974 | 0,5474 | -1,439 | 2,7882 | -1,289 | 0,2115 | 3,2261 | 0,3779 | 0,1321 |
| -0,57 | 0,466 | 0,6778 | 1,1292 | -0,279 | 0,3128 | -0,414 | 0,5483 | 1,6146 | 2,321 |
| 4,3788 | 3,1857 | -0,876 | -0,748 | -0,189 | 1,5779 | 0,6836 | 0,3412 | 2,2469 | 1,018 |
| 0,3568 | -0,611 | -0,393 | -1,069 | 0,9291 | -3,991 | 3,0778 | -1,794 | 1,269 | -1,446 |
| -0,114 | 1,0855 | 0,7176 | 0,291 | -5,601 | -1,458 | 0,282 | -1,314 | 0,7293 | -0,61 |
| -0,283 | 0,4941 | 0,1782 | -0,034 | 3,6203 | 0,5758 | 1,0648 | -0,198 | -0,015 | -0,518 |
| -0,688 | 0,5781 | 0,3784 | 2,9263 | -0,722 | 0,3285 | -0,791 | -0,951 | -2,324 | 0,3049 |
| -0,496 | -1,62 | 0,0911 | -0,639 | -0,298 | -0,18 | -3,717 | 0,3879 | 0,699 | 0,1266 |

# ЛАБОРАТОРНАЯ РАБОТА 3

# Дисперсионный анализ данных

***Цель работы:***

Выполнить однофакторный и двухфакторный анализ данных.

## 3.1. Практическое задание

### 3.1.1. Однофакторный параметрический анализ

1) Сгенерировать средствами пакета EXCEL 8 выборок из 10 значений случайной величины с нормальным законом распределения $N(\mu, \sigma^2)$. Варианты значений параметров $\mu, \sigma^2$ приведены в приложении 1. Здесь $k = 8$ – количество уровней фактора $A$, $n_1 = n_2 = ...n_8 = n = 10$ – объемы выборок.

2) Выполнить следующие расчеты:

- вычислить средние значения по уровням $\bar{x}_i = \frac{1}{n}\sum_{j=1}^{n} x_{ji}$;
- вычислить общее среднее для всей выборки $\bar{x} = \frac{1}{k \cdot n}\sum_{i=1}^{k}\sum_{j=1}^{n} x_{ji}$;
- вычислить общую выборочную дисперсию $s_0^2 = \frac{1}{N-1}\sum_{i=1}^{k}\sum_{j=1}^{n}(x_{ji} - \bar{x})^2,\ N = k \cdot n$;
- вычислить выборочные дисперсии по уровням $s_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_{ji} - \bar{x}_i)^2$;
- проверить однородность дисперсий $s_i^2$, используя статистику Кохрана (см. [1], стр. 418) $g = \frac{\max\limits_{1 \le i \le k} s_i^2}{\sum\limits_{i=1}^{k} s_i^2},\ k = 8$. Гипотеза об однородности дисперсий принимается, если выполняется неравенство $g \le g_\alpha(k, n);\ k = 8,\ n = 10$, где $\alpha$ – доверительная вероятность (положить $\alpha = 0{,}95$). Значения $g_\alpha(k, n)$ приведены в таблице 8 (см. статистические таблицы).

- вычислить дисперсию, характеризующую фактор случайности $s_{сл}^2 = \frac{1}{k}\sum_{i=1}^{k} s_i^2$, где $k$ – количество уровней фактора $A$ ($k = 8$);
- вычислить дисперсию фактора $A$: $s_A^2 = \frac{n}{k-1}\sum_{i=1}^{k} (\bar{x}_i - \bar{x})^2$;

3) Проверить нулевую гипотезу равенства средних значений на различных уровнях фактора $A$:

$$H_0 : m_1 = m_2 = ... = m_k = m,$$

используя критерий $F = \frac{s_A^2}{s_{сл}^2}$. Если $F = \frac{s_A^2}{s_{сл}^2} \le F_{0,05}(7; 72)$, то гипотеза $H_0$ принимается, следовательно, номер уровня фактора $A$ не влияет на исследуемый процесс $X$.

## 3.1.2. Однофакторный непараметрический анализ

### *3.1.2.1. Анализ на основе критерия Краскела-Уоллеса (произвольные альтернативы)*

1) Сгенерировать средствами пакета EXCEL 8 выборок из 10 значений случайной величины с биномиальным законом распределения $B(n, p)$. Варианты значений параметров $n, p$ приведены в приложении 1. Здесь $k = 8$ – количество уровней фактора $A$, $n_1 = n_2 = ...n_8 = n = 10$ – объемы выборок.
2) Заменить данные $\{x_{ji}\}$ (наблюдения) их рангами $\{r_{ji}\}$ путем упорядочивания табличных данных $\{x_{ji}\}$ в порядке возрастания.
3) Для каждой обработки $i$ (уровня фактора, столбца таблицы) вычислить:
   - суммарный и средний ранги $R_i = \sum_{j=1}^{n_i} r_{ji}$ и $\bar{R}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} r_{ji}$;
   - средний по всей совокупности ранг $\hat{R} = \frac{N+1}{2}$, где $N = \sum_{i=1}^{k} n_i$ – общее число наблюдений ($N = 80$);

- вычислить статистику *Краскела-Уоллеса*

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{n} n_i \left( \overline{R}_i - \frac{N+1}{2} \right)^2 , \; n_1 = n_2 = ... n_8 = 10 .$$

4) Проверить, если ли среди $x_{ji}$ совпадающие значения. Если имеются совпадающие значения, то при ранжировании и переходе к $r_{ji}$ надо использовать средние ранги (например, если 2 значения (5 и 5) занимают ранги 11, 12, то средний ранг (11,5) надо присвоить им обоим).
5) Вычислить модифицированную статистику $H'$:

$H' = \frac{H}{1 - \left( \sum_{j=1}^{m} T_j / (N^3 - N) \right)}$, где $m$ — число групп совпадающих наблюдений;

$T_j = r_j^3 - t_j$ ($t_j$ — число совпадающих наблюдений в группе $j$).

6) Проверить гипотезу $H_0$ о том, что расхождение наблюдений в сериях опытов для различных уровней факторов можно объяснить только случайными причинами. На статистическом языке это предположение означает, что все данные таблицы $x_{ji}$ принадлежат одному и тому же распределению. Если $H \le \chi^2_\alpha (v = 7)$ (или $H' \le \chi^2_\alpha (v = 7)$), то гипотеза $H_0$ принимается.

### *3.1.2.2.Анализ на основе критерия Джонкхиера (альтернативы с упорядочиванием)*

1) Для табличных данных, полученных в задании 2.1, дДля каждой пары уровней $u$ и $v$, где $1 \le u < v \le k$, вычислить по выборкам с номерами $u$ и $v$ статистики Манна-Уитни: $U(u,v) = \sum_{i=1}^{n} \sum_{i=1}^{n} \varphi(x_i, y_j)$, где $x$ и $y$ – два сравниваемых столбца таблицы анализируемых данных. Здесь

$$\varphi(x_i, y_j) = \begin{cases} 0, & если\, x_i > y_j; \\ \frac{1}{2}, & если\, x_i = y_j; \\ 1, & если\, x_i < y_j. \end{cases}$$

2) Вычислить статистику Джонкхиера $I = \sum U(u,v)$ для $1 \le u < v \le k$.
3) Вычислить среднее и дисперсию статистики Джонкхиера

$MI = \frac{1}{4} \left( N^2 - \sum_{i=1}^{k} n_j \right)$; $DI = \frac{1}{72} \left[ N^2 (2N+3) - \sum_{j=1}^{k} n_j^2 (2n_j + 3) \right]$, где $n_j$ – количество наблюдений в каждом уровне ($n_1 = n_2 = ... n_8 = 10$), а затем нормированную величину $I^* = \frac{I - MI}{\sqrt{DI}}$.

4) Проверить гипотезу $H_0$ о том, что расхождение наблюдений в сериях опытов для различных уровней факторов можно объяснить только случайными причинами. Если $\left|I^*\right| \le u_\alpha,\ \alpha = 0{,}975$, то гипотеза $H_0$ принимается.

### 3.1.3. Двухфакторный параметрический анализ

1) Сгенерировать средствами пакета EXCEL 10 выборок из 30 значений случайной величины с нормальным законом распределения $N(\mu_m, \sigma_m^2),\ m = 1,...,10$. Варианты значений параметров $\mu_m, \sigma_m^2$ приведены в приложении 2. Из этих данных подготовить первый столбец таблицы, составленный из средних выборочных значений: $x_1 = \frac{1}{30}\sum_{i=1}^{30} x_i^{(1)}$ – по первой выборке; $x_2 = \frac{1}{30}\sum_{i=1}^{30} x_i^{(2)}$ – по второй выборке и т.д. $x_{10} = \frac{1}{30}\sum_{i=1}^{30} x_i^{(10)}$ – по десятой выборке. Этот столбец будет соответствовать первому уровню фактора $A$. Каждое значение $x_i,\ i = 1,...,10$ в этом столбце соответствует фактору $B$.

Аналогично формируются 2-й, 3-й и т.д. 8-й столбцы таблицы, которые будут соответствовать 2-му, 3-му и т.д. 8-му уровням фактора $A$. В результате мы получим таблицу, подобную таблице задания №1, в которой $k = 8$ – количество уровней фактора $A$, $n_1 = n_2 = ...n_8 = 10$ – уровни фактора $B$.

2) Выполнить следующие расчеты:

- вычислить средние значения по уровням фактора $A$ $\bar{x}_i = \frac{1}{n}\sum_{j=1}^{n} x_{ji}$;
- вычислить средние значения по уровням фактора $B$ $\bar{x}_j = \frac{1}{n}\sum_{i=1}^{k} x_{ji}$;
- вычислить общее среднее для всей выборки $\bar{x} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n} x_{ji}^2$;
- вычислить выборочные дисперсии по уровням $s_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_{ji} - \bar{x}_i)^2$;

- вычислить дисперсию, характеризующую фактор случайности

  $s_{сл}^2 = \frac{1}{(n-1)(k-1)} \sum_{i=1}^{k} \sum_{j=1}^{n} \left( x_{ji} - \bar{x}_i - \bar{x}_j + \bar{x} \right)^2$, где $k$ – количество уровней фактора $A$ $(k = 8)$;

- вычислить дисперсию фактора $A$: $s_A^2 = \frac{n}{n-1} \sum_{i=1}^{k} (\bar{x}_i - \bar{x})^2$;

3) Проверить нулевую гипотезу равенства средних значений на различных уровнях фактора $A$:

$$H_0 : \alpha_1 = \alpha_2 = ... = \alpha_k ,$$

используя критерий $F = \frac{s_A^2}{s_{сл}^2}$. Если $F = \frac{s_A^2}{s_{сл}^2} \le F_{0,05}(7; 72)$, то гипотеза $H_0$ принимается, следовательно, фактор $A$ не влияет на исследуемый процесс $X$.

## 3.1.4. Двухфакторный непараметрический анализ

### *3.1.4.1. Анализ на основе критерия Фридмана (произвольные альтернативы)*

1) Сгенерировать средствами пакета EXCEL 10 выборок из 30 значений случайной величины с биномиальным законом распределения $B(n_m,\ p_m),\ m = 1,...,10$. Варианты значений параметров $n_m,\ p_m$ приведены в приложении 2. Из этих данных подготовить первый столбец таблицы, составленный из средних выборочных значений: $x_1 = \frac{1}{30} \sum_{i=1}^{30} x_i^{(1)}$ – по первой выборке; $x_2 = \frac{1}{30} \sum_{i=1}^{30} x_i^{(2)}$ – по второй выборке и т.д. $x_{10} = \frac{1}{30} \sum_{i=1}^{30} x_i^{(10)}$ – по десятой выборке. Этот столбец будет соответствовать первому уровню фактора $A$. Каждое значение $x_i,\ i = 1,...,10$ в этом столбце соответствует фактору $B$.

Аналогично формируются 2-й, 3-й и т.д. 8-й столбцы таблицы, которые будут соответствовать 2-му, 3-му и т.д. 8-му уровням фактора $A$. В результате мы получим таблицу, подобную таблице задания 2.1, в которой $k = 8$ – количество уровней фактора $A$, $n_1 = n_2 = ...n_8 = 10$ – уровни фактора $B$.

1)

2) Заменить данные $\{x_{ji}\}$ (наблюдения), полученные в задании №3 их рангами $\{r_{ji}\}$ путем упорядочивания табличных данных $\{x_{ji}\}$ в порядке возрастания. При этом ранги подсчитываются по каждой строке отдельно (т.е. по каждому фактору $B_j$)

3) Для каждого столбца таблицы вычислить:

- суммарный и средний ранги $R_i = \sum_{j=1}^{n_i} r_{ji}$ и $\overline{R}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} r_{ji}$ ; $n_i = n = 10$.
- вычислить статистику *Фридмана* $S = \frac{12n}{k(k+1)}\sum_{i=1}^{k}\left(\overline{R}_i - \overline{R}\right)^2$.

3) Проверить гипотезу $H_0$ о том, что расхождение наблюдений в сериях опытов для различных уровней факторов можно объяснить только случайными причинами. Если $S \le \chi^2_{0,05}(7)$, то гипотеза $H_0$ принимается.

### *3.1.4.2. Анализ на основе критерия Пейджа (альтернативы с упорядочиванием)*

1) Упорядочить по возрастанию суммарные ранги $R_i = \sum_{j=1}^{n_i} r_{ji}$, полученные в задании 4.1.

2) Вычислить статистику Пейджа $L = \sum_{i=1}^{k} i \cdot R_i = 1 \cdot R_1 + 2 \cdot R_2 + 3 \cdot R_3 + ... + k \cdot R_k$.

3) Проверить гипотезу $H_0$ о том, что расхождение наблюдений в сериях опытов для различных уровней факторов можно объяснить только случайными причинами. Если $L_{набл} \le L_q(k,n)$, то гипотеза $H_0$ принимается.

# Приложение к лабораторной работе 3

## Варианты заданий

### Варианты задания 1

| Номер варианта | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\mu$ | 0 | 0,5 | 1,0 | 2,0 | 2,5 | 3,0 | 3,5 | 4,0 | 4,5 | 5,0 |
| $\sigma$ | 0,2 | 0,2 | 0,2 | 0,5 | 0,5 | 0,5 | 1 | 1 | 1 | 1,5 |

### Варианты задания 2

| Номер варианта | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | 20 | 30 | 40 | 35 | 25 | 100 | 90 | 80 | 70 | 50 |
| $p$ | 0,4 | 0,5 | 0,7 | 0,4 | 0,8 | 0,3 | 0,6 | 0,7 | 0,1 | 0,5 |

### Варианты задания 3

| $m$ | 1 | | 2 | | 3 | | 4 | | 5 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\mu_m$ | $\sigma_m$ | $\mu_m$ | $\sigma_m$ | $\mu_m$ | $\sigma_m$ | $\mu_m$ | $\sigma_m$ | $\mu_m$ | $\sigma_m$ |
| 1 | 1 | 0,5 | 1,4 | 0,5 | 1 | 1,5 | 2 | 0,5 | 1 | 1,5 |
| 2 | 1,2 | 0,6 | 2,2 | 0,6 | 1,2 | 1,6 | 2,2 | 0,6 | 1,2 | 2,6 |
| 3 | 1,6 | 0,7 | 1,6 | 0,7 | 1,6 | 0,7 | 1,6 | 0,7 | 1,6 | 1,7 |
| 4 | 1,3 | 0,5 | 1,3 | 0,5 | 1,3 | 0,5 | 2,3 | 0,5 | 1,3 | 0,5 |
| 5 | 1,7 | 0,7 | 2,5 | 0,7 | 1,7 | 0,7 | 1,7 | 0,7 | 1,7 | 1,7 |
| 6 | 1.4 | 0,8 | 1.4 | 0,8 | 1.4 | 1,8 | 1.4 | 0,8 | 1.4 | 0,8 |
| 7 | 1,5 | 0,4 | 1,5 | 0,4 | 1,5 | 0,4 | 1,5 | 0,4 | 1,5 | 1,4 |
| 8 | 1,9 | 0,4 | 2,9 | 0,4 | 1,9 | 0,4 | 2,9 | 0,4 | 1,9 | 0,4 |
| 9 | 1,6 | 0,5 | 1,6 | 0,5 | 1,6 | 0,5 | 1,6 | 0,5 | 1,6 | 1,5 |
| 10 | 1,8 | 0,9 | 1,8 | 0,9 | 1,8 | 0,9 | 2,8 | 0,9 | 1,8 | 0,9 |
| $m$ | 6 | | 7 | | 8 | | 9 | | 10 | |
| | $\mu_m$ | $\sigma_m$ | $\mu_m$ | $\sigma_m$ | $\mu_m$ | $\sigma_m$ | $\mu_m$ | $\sigma_m$ | $\mu_m$ | $\sigma_m$ |
| 1 | 1 | 0,5 | 1,4 | 0,05 | 0 | 1,5 | 1 | 1,5 | 2 | 1,5 |
| 2 | 3,2 | 0,6 | 2,2 | 0,6 | 0,2 | 1,6 | 1,2 | 0,6 | 1,2 | 2,6 |
| 3 | 1,6 | 0,7 | 1,6 | 0,7 | 1,6 | 0,7 | 1,6 | 0,7 | 1,6 | 1,7 |
| 4 | 1,3 | 0,5 | 1,3 | 0,5 | 0,3 | 0,5 | 2,3 | 0,5 | 1,3 | 0,5 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 3,7 | 0,7 | 2,5 | 0,07 | 1,7 | 0,7 | 1,7 | 0,7 | 2,7 | 1,7 |
| 6 | 1.4 | 0,8 | 1.4 | 0,8 | 1.4 | 1,8 | 1.4 | 1,8 | 0.4 | 0,8 |
| 7 | 1,5 | 0,4 | 1,5 | 0,4 | 0,5 | 0,4 | 1,5 | 0,4 | 1,5 | 1,4 |
| 8 | 1,9 | 0,4 | 2,9 | 0,4 | 1,9 | 0,4 | 2,9 | 0,4 | 1,9 | 0,4 |
| 9 | 3,6 | 0,5 | 1,6 | 0,5 | 1,6 | 0,5 | 1,6 | 0,5 | 1,6 | 1,5 |
| 10 | 1,8 | 0,9 | 1,8 | 0,09 | 0,8 | 0,9 | 2,8 | 0,9 | 1,8 | 0,9 |

## Варианты задания 4

| $m$ | 1 | | 2 | | 3 | | 4 | | 5 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $n_m$ | $p_m$ | $n_m$ | $p_m$ | $n_m$ | $p_m$ | $n_m$ | $p_m$ | $n_m$ | $p_m$ |
| 1 | 30 | 0,5 | 35 | 0,5 | 20 | 0,4 | 30 | 0,1 | 30 | 0,5 |
| 2 | 40 | 0,6 | 45 | 0,6 | 30 | 0,6 | 40 | 0,2 | 100 | 0,6 |
| 3 | 50 | 0,7 | 55 | 0,7 | 35 | 0,4 | 50 | 0,3 | 50 | 0,7 |
| 4 | 60 | 0,5 | 65 | 0,5 | 60 | 0,2 | 60 | 0,4 | 60 | 0,5 |
| 5 | 70 | 0,7 | 75 | 0,7 | 70 | 0,7 | 70 | 0,5 | 30 | 0,7 |
| 6 | 80 | 0,8 | 85 | 0,8 | 80 | 0,8 | 80 | 0,6 | 40 | 0,8 |
| 7 | 90 | 0,4 | 95 | 0,4 | 90 | 0,4 | 90 | 0,4 | 90 | 0,3 |
| 8 | 20 | 0,4 | 25 | 0,4 | 20 | 0,4 | 20 | 0,3 | 20 | 0,4 |
| 9 | 100 | 0,5 | 100 | 0,5 | 90 | 0,3 | 100 | 0,2 | 100 | 0,2 |
| 10 | 85 | 0,9 | 85 | 0,9 | 85 | 0,9 | 85 | 0,9 | 85 | 0,9 |
| $m$ | 6 | | 7 | | 8 | | 9 | | 10 | |
| | $n_m$ | $p_m$ | $n_m$ | $p_m$ | $n_m$ | $p_m$ | $n_m$ | $p_m$ | $n_m$ | $p_m$ |
| 1 | 10 | 0,5 | 15 | 0,55 | 40 | 0,4 | 15 | 0,1 | 30 | 0,55 |
| 2 | 20 | 0,6 | 45 | 0,65 | 30 | 0,6 | 45 | 0,2 | 100 | 0,65 |
| 3 | 30 | 0,7 | 55 | 0,75 | 50 | 0,4 | 55 | 0,3 | 50 | 0,75 |
| 4 | 40 | 0,5 | 65 | 0,55 | 60 | 0,2 | 65 | 0,4 | 60 | 0,55 |
| 5 | 50 | 0,7 | 75 | 0,7 | 70 | 0,7 | 75 | 0,5 | 30 | 0,7 |
| 6 | 40 | 0,8 | 85 | 0,8 | 80 | 0,8 | 85 | 0,6 | 40 | 0,8 |
| 7 | 30 | 0,4 | 95 | 0,4 | 50 | 0,4 | 95 | 0,4 | 90 | 0,4 |
| 8 | 20 | 0,4 | 25 | 0,45 | 20 | 0,4 | 25 | 0,3 | 20 | 0,45 |
| 9 | 10 | 0,5 | 100 | 0,5 | 30 | 0,3 | 100 | 0,2 | 100 | 0,5 |
| 10 | 85 | 0,9 | 85 | 0,9 | 10 | 0,9 | 85 | 0,9 | 85 | 0,9 |

# ЛАБОРАТОРНАЯ РАБОТА 4

# Корреляционный анализ случайных данных

***Цель работы:***

Рассчитать параметрические и непараметрические коэффициенты корреляции.

## 4.1. Практическое задание

### 4.1.1. Вычисление параметрических коэффициентов корреляции

1) Сгенерировать средствами пакета EXCEL 5 выборок из 10 значений случайной величины с нормальным законом $N(\mu, \sigma^2)$. Эти 5 выборок будем использовать в качестве независимых признаков $(x_1, x_2, x_3, x_4, x_5)$. Варианты значений параметров $\mu, \sigma^2$ приведены в приложении 1.

2) Рассчитать зависимый признак $y_i = a_0 + a_1 \cdot x_{1i} + a_2 \cdot x_{2i} + a_3 \cdot x_{3i} + a_4 \cdot x_{4i} + a_5 \cdot x_{5i} + \varepsilon_i$, $i = 1,...,10$. Здесь $a_0, a_1, ..., a_5$ – параметры (их значения приведены в приложении 1); $\varepsilon_i$ – случайная погрешность с нормальным законом распределения $\varepsilon_i \sim N(0, \sigma_i^2)$, где $\sigma_i = 0,2 \cdot M(\bar{y})$.

#### *4.1.1.1. Парные коэффициенты корреляции*

3) Выполнить следующие расчеты:

- вычислить выборочные средние и дисперсии зависимого признака $y$ и независимых признаков $x_1, x_2, x_3, x_4, x_5$, а также средние значения произведений $\overline{y \cdot x_j}$ по формулам:
- $s_y^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)^2$; $s_{x_j}^2 = \frac{1}{n}\sum_{i=1}^{n} x_{ij}^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_{ij}\right)^2$; $j = 1,...,5$;
- $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$;; $\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$; $\overline{y \cdot x}_j = \frac{1}{n}\sum_{i=1}^{n} y_i \cdot x_{ij}$ $\overline{y \cdot x}_j = \frac{1}{n}\sum_{i=1}^{n} y_i \cdot x_{ij}$. Здесь $x_{ij} - i$-е значение случайной величины $x$ из $j$-й выборки.

- скорректировать коэффициент корреляции (так как $n<15$) по формуле $r^{*}=r\left(1+\frac{1-r^2}{2(n-3)}\right)$.
- вычислить парные коэффициенты корреляции между зависимым признаком $y$ и независимыми признаками $x_1, x_2, x_3, x_4, x_5$ (коэффициенты корреляции Пирсона) по формулам: $r_{yx_j}=\frac{\overline{y\cdot x_j}-\overline{y}\cdot\overline{x}_j}{s_y\cdot s_{x_j}}$.
- вычислить $t$-статистики $t_j=\sqrt{\frac{r_{yx_j}^2(n-2)}{1-r_{yx_j}^2}}$.

4) Проверить гипотезу о значимости коэффициентов корреляции. Если $t_j>t_{0,05}(n-2=8)$, то коэффициент $r_{yx_j}$ значимый, и, следовательно, связь между $y$ и $x_j$ статистически значима.

### *4.1.1.2 Множественный коэффициент корреляции*

5) Вычислить парные коэффициенты корреляции между зависимыми признаками $r_{x_ix_j}=\frac{\overline{x_i\cdot x_j}-\overline{x_i}\cdot\overline{x_j}}{s_{x_i}\cdot s_{x_j}};\ j>i,\ i=1,...,5$, где $\overline{x_i\cdot x_j}=\frac{1}{n}\sum_{k=1}^{n}x_{ki}\cdot x_{kj};\ j>i;\ i=1,...,5$

Вычислить множественный коэффициент корреляции между результирующим признаком $Y$ и факторными признаками $x_1, x_1, ..., x_5$ по формуле

$$R_{y/x_1,x_2,...,x_m}=\sqrt{1-\frac{|\rho|}{|\rho_1|}}, \tag{8}$$

где $|\rho|$ – определитель матрицы парной корреляции

$$\rho=\begin{pmatrix}1 & \rho_{yx_1} & \rho_{yx_2} & \rho_{yx_2} & \rho_{yx_4} & \rho_{yx_5}\\ \rho_{x_1y} & 1 & \rho_{x_1x_2} & \rho_{x_1x_3} & \rho_{x_1x_4} & \rho_{x_1x_5}\\ \rho_{x_2y} & \rho_{x_2x_1} & 1 & \rho_{x_2x_3} & \rho_{x_2x_3} & \rho_{x_2x_5}\\ \rho_{x_3y} & \rho_{x_3x_1} & \rho_{x_3x_2} & 1 & \rho_{x_3x_4} & \rho_{x_3x_5}\\ \rho_{x_4y} & \rho_{x_4x_1} & \rho_{x_4x_2} & \rho_{x_4x_3} & 1 & \rho_{x_4x_5}\\ \rho_{x_5y} & \rho_{x_5x_1} & \rho_{x_5x_2} & \rho_{x_5x_3} & \rho_{x_5x_4} & 1\end{pmatrix}; \tag{9}$$

$|\rho_1|$ – алгебраическое дополнение элемента $\rho_{11}$.

6) Вычислить скорректированный коэффициент корреляции:

$$\widetilde{R}_{y/x_1,x_2,\ldots,x_k} = \sqrt{1-(1-R^2)\frac{n-1}{n-k-1}},$$

7) Вычислить статистику Фишера $F = \dfrac{\frac{1}{2}R^2_{y/x_1,\ldots,x_5}}{\frac{1}{n-6}\left(1-R^2_{y/x_1,\ldots,x_5}\right)}$;

8) Проверить гипотезу о значимости множественного коэффициента корреляции. Если $F > F_{0,05}(v_1 = 5,\ v_2 = n-6)$, то множественный коэффициент корреляции считается значимым.

## 4.1.2. Вычисление непараметрических коэффициентов корреляции

1) Сгенерировать средствами пакета EXCEL 5 выборок из 10 значений случайной величины с биномиальным законом распределения $B(n,\ p)$. Эти 5 выборок будем использовать в качестве независимых признаков $(x_1, x_2, x_3, x_4, x_5)$. Варианты значений параметров $n,\ p$ приведены в приложении 2.
2) Проранжировать выборки, не упорядочивая их.

### *4.1.2.1. Коэффициент ранговой корреляции Спирмана*

3) Вычислить:

- величины $d_k^2(i,j) = \left(R_{k,x_i} - R_{k,x_j}\right)^2;\ i \neq j,\ i,j = 1,\ldots,5;\ k = 1,\ldots,10;$
- коэффициенты ранговой корреляции Спирмана $\rho_{x_i/y_j} = 1 - \dfrac{6\sum_{k=1}^{n} d_k^2(i,j)}{n(n^2-1)}$;
- $t$-статистики $t_{ij} = \rho_{x_i/y_j}\sqrt{\dfrac{n-2}{1-\rho^2_{x_i/y_j}}}$

4) Проверить значимость коэффициентов корреляции. Если $t_{ij} > t_{0,05}(n-2)$, то коэффициент $\rho_{x_i / y_j}$ считается значимым.

#### *4.1.2.1.2 Коэффициент ранговой корреляции Кендалла*

5) Рассматриваются все комбинации пар столбцов исходной таблицы данных – (1;2), (1;3), (1;4), (1;5), (2;3), (2;4), (2;5), (3;4), (3;5), (4;5). Первый столбец обозначим $X$, второй – $Y$.

- В каждой паре столбцов значения первого столбца $X$ упорядочиваются по возрастанию, а значения второго столбца $Y$ располагаются в порядке, соответствующем значениям $X$
- для каждого ранга $Y$ определяется число следующих за ним значений рангов, превышающих его по величине. Суммируя эти числа, определяем величину $P$ (число последовательностей) — меру соответствия последовательностей рангов $X$ и $Y$ (см. пример в лекции);
- для каждого ранга $Y$ определяется число следующих за ним рангов, меньших его величины. Суммируя величины, получаем величину $Q$ (число инверсий);
- определяется разность по всем членам ряда $S = P - Q$ и вычисляется $\tau$. Связь между признаками можно признать статистически значимой, если значение коэффициента корреляции $|\tau| > \tau_\alpha = u_\alpha \sqrt{\frac{2(2n+5)}{9n(n-1)}}$.

#### *4.1.2.1.3 Коэффициент конкордации (множественный коэффициент ранговой корреляции)*

6)Проранжировать столбцы исходной таблицы $\{x_{ji}\}$ (наблюдения) их рангами $\{r_{ji}\}$ не упорядочивая табличные данные.

7)Для каждой $j$ -й строки таблицы вычислить:

8)сумму рангов $R_j = \sum_{i=1}^{5} r_{ji}$ и квадрат суммы $R_j^2$;

9)сумму рангов по всей совокупности ранг $\widetilde{R} = \sum_{j=1}^{10} R_j$ и $\widetilde{R^2} = \sum_{j=1}^{10} R_j^2$;

10)вычислить коэффициент конкордации $W = \dfrac{12\left(\widetilde{R^2} - \dfrac{\widehat{R}^2}{n}\right)}{m^2(n^3 - n)}$, $m = 5,\ n = 10$.

11) Проверить значимость связи между признаками. Если $W > W_\alpha$, где $W_\alpha = \dfrac{1}{m(n-1)}\chi_\alpha^2(n-1)$, то с вероятностью $\alpha$ корреляция между признаками признается значимой. Если среди последовательностей рангов есть совпадения, то коэффициент конкордации следует вычислять по формуле

$$W = \frac{12\left(\widetilde{R^2} - \left(\dfrac{\widehat{R}}{n}\right)^2\right)}{m^2(n^2-1) - m\sum_{j=1}^{m} T_j},$$

где $T_j = t_j^3 - t_j$, $t_j$ – количество совпавших рангов в $j$-й последовательности.
Совпавшим рангам присваиваются средние ранги.

# Приложение к лабораторной работе 4

## Варианты заданий

### Варианты задания 1

| Номер варианта | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\mu$ | 0 | 0,5 | 1,0 | 2,0 | 2,5 | 3,0 | 3,5 | 4,0 | 4,5 | 5,0 |
| $\sigma$ | 0,2 | 0,2 | 0,2 | 0,5 | 0,5 | 0,5 | 1 | 1 | 1 | 1,5 |
| $a_0$ | 1 | 2 | 3 | 4 | 5 | 1 | -2 | 3 | -4 | 5 |
| $a_1$ | 2 | 3 | 4 | 5 | 6 | -2 | 3 | -4 | 5 | -6 |
| $a_2$ | 3 | 4 | 5 | 6 | 7 | 3 | -4 | 5 | -6 | 7 |
| $a_3$ | 4 | 5 | 6 | 7 | 8 | -4 | 5 | -6 | 7 | -8 |
| $a_4$ | 5 | 6 | 7 | 8 | 9 | 5 | -6 | 7 | -8 | 9 |
| $a_5$ | 6 | 7 | 8 | 9 | 10 | -6 | 7 | -8 | 9 | -10 |

### Варианты задания 2

| Номер варианта | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | 20 | 30 | 40 | 35 | 25 | 100 | 90 | 80 | 70 | 50 |
| $p$ | 0,4 | 0,5 | 0,7 | 0,4 | 0,8 | 0,3 | 0,6 | 0,7 | 0,1 | 0,5 |

# ЛАБОРАТОРНАЯ РАБОТА 5

# Линейная регрессия

***Цель работы:***

Оценка уравнения линейной регрессии на основе выборочных данных

## 5.1. Необходимые сведения из теории

### 5.1.1. Построение модели парной регрессии

Рассмотрим линейную по коэффициентам модель парной регрессии:

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

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

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

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

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

$$Y = XB + \varepsilon,$$

где $Y = (y_1, y_2 ..., y_n)^T$ - вектор значений переменной $y$;

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

$\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T$ – 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 $x_1, x_2, ..., x_k$:

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

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

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

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

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

### 5.1.2. Оценка погрешности регрессии

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

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

$\hat{y}_i=\beta_0^*+\beta_1^*x_i+\ldots\beta_k^*x_k$. Качество модели также можно оценить с использованием оценки коэффициента детерминации: $R^2=\dfrac{\sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2}{\sum_{i=1}^{n}(y_i-\overline{y})^2}$.

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

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

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 $t_j=\dfrac{\beta_j^*}{\sqrt{s^2\left[(X^TX)^{-1}\right]_{jj}}}$, которая при истинности гипотезы $H_0$: $\beta_j=0$, 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 $P\left(x > \left|t_j\right|\right)$. Степень значимости параметров распределения качественно определяется по уровню значимости: не значимые ($\alpha \geq 0{,}100$), 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 $\tilde{t}_j = \dfrac{\beta_j^* - \beta_j}{\sqrt{s^2\left[(X^T X)^{-1}\right]_{jj}}}$, имеющие распределение Стьюдента с $n-k-1$ степенью свободы. Для уровня значимости $\alpha$ доверительный интервал рассчитывается по формуле $\beta_j^* \pm t_\alpha \sqrt{s^2\left[(X^T X)^{-1}\right]_{jj}}$, где $t_\alpha$ – квантиль распределение Стьюдента с $n-k-1$ степенью свободы.

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

## 5.2. 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. 1)


Рис. 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

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

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, соответствующую $X$ (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 ИТОГОВ

| *Регрессионная статистика* | |
|---|---|
| Множественный R | 0,852379622 |
| R-квадрат | 0,72655102 |
| Нормированный R-квадрат | 0,694380552 |
| Стандартная ошибка | 1,820336831 |
| Наблюдения | 20 |

| | *Коэффициенты* | *Стандартная ошибка* | *t-статистика* | *P-Значение* |
|---|---|---|---|---|
| Y-пересечение | -0,963028418 | 1,354297293 | -0,711090853 | 0,486670901 |
| Переменная X 1 | 2,604940094 | 0,594036759 | 4,385149663 | 0,000403854 |
| Переменная X 2 | -0,167559332 | 0,054953956 | -3,049085893 | 0,007253372 |

Здесь в первой таблице:

1. Множественный R – корень квадратный из коэффициента детерминации;
2. R-квадрат – коэффициент детерминации;
3. Нормированный R-квадрат – это скорректированная величина коэффициента детерминации, вычисляемая по формуле $\hat{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k-1}$;
4. Стандартная ошибка – значение $s = \sqrt{s^2}$, где $s^2$ - оценка дисперсии предсказания $\sigma^2$;
5. Наблюдения – объем выборки

Во второй 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 $P\left(x > \left|t_j\right|\right)$.

Соответствующий график предсказанных значений в сравнении с исходными данными имеет вид (рис. 2):

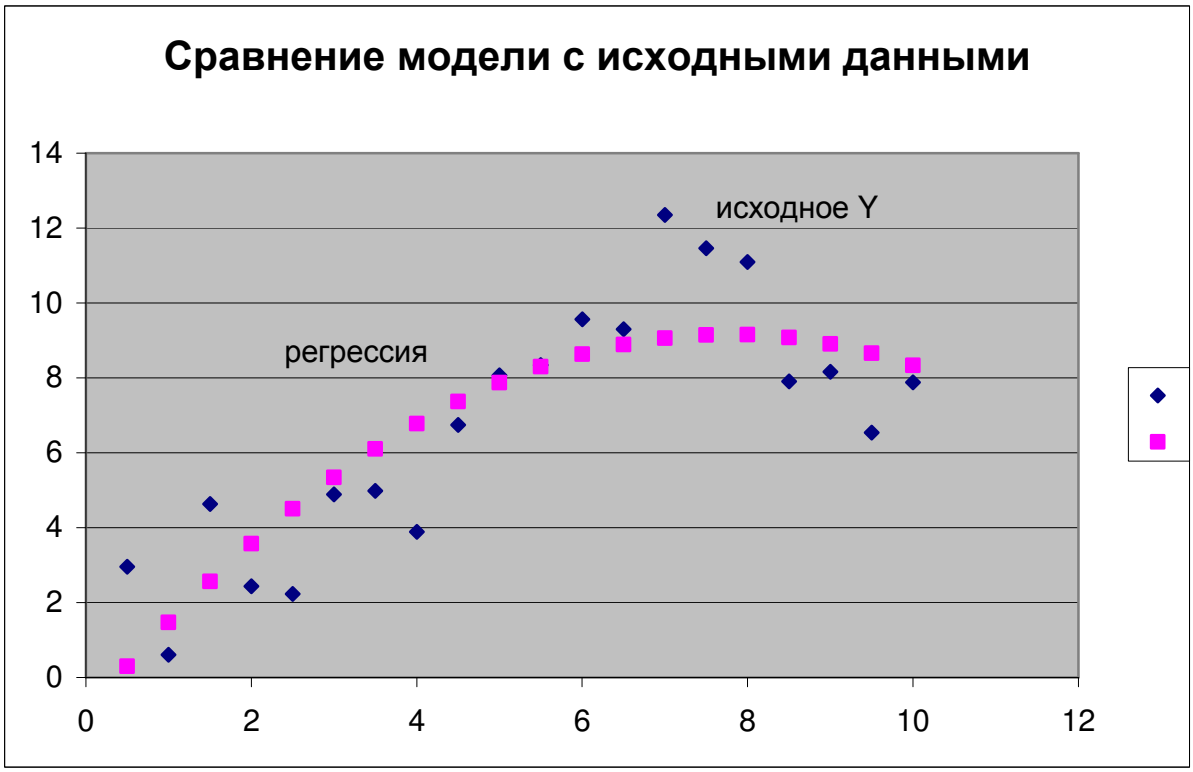



Рис. 2. Сравнение модели $y = \beta_0 + \beta_1 x + \beta_2 x^2$ с исходными данными

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

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$$

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

ВЫВОД ИТОГОВ

| *Регрессионная статистика* | |
|---|---|
| Множественный R | 0,925507816 |
| R-квадрат | 0,856564717 |
| Нормированный R-квадрат | 0,829670602 |
| Стандартная ошибка | 1,358958106 |
| Наблюдения | 20 |

| | *Коэффициенты* | *Стандартная ошибка* | *t-статистика* | *P-Значение* |
|---|---|---|---|---|
| Y-пересечение | 3,176148907 | 1,484433212 | 2,13963746 | 0,048142193 |
| Переменная X 1 | -1,619155418 | 1,194561911 | -1,355438679 | 0,194104204 |
| Переменная X 2 | 0,814063749 | 0,261005991 | 3,118946605 | 0,006612096 |
| Переменная X 3 | -0,062325275 | 0,016365816 | -3,808259635 | 0,001545564 |

График показан на рис. 3

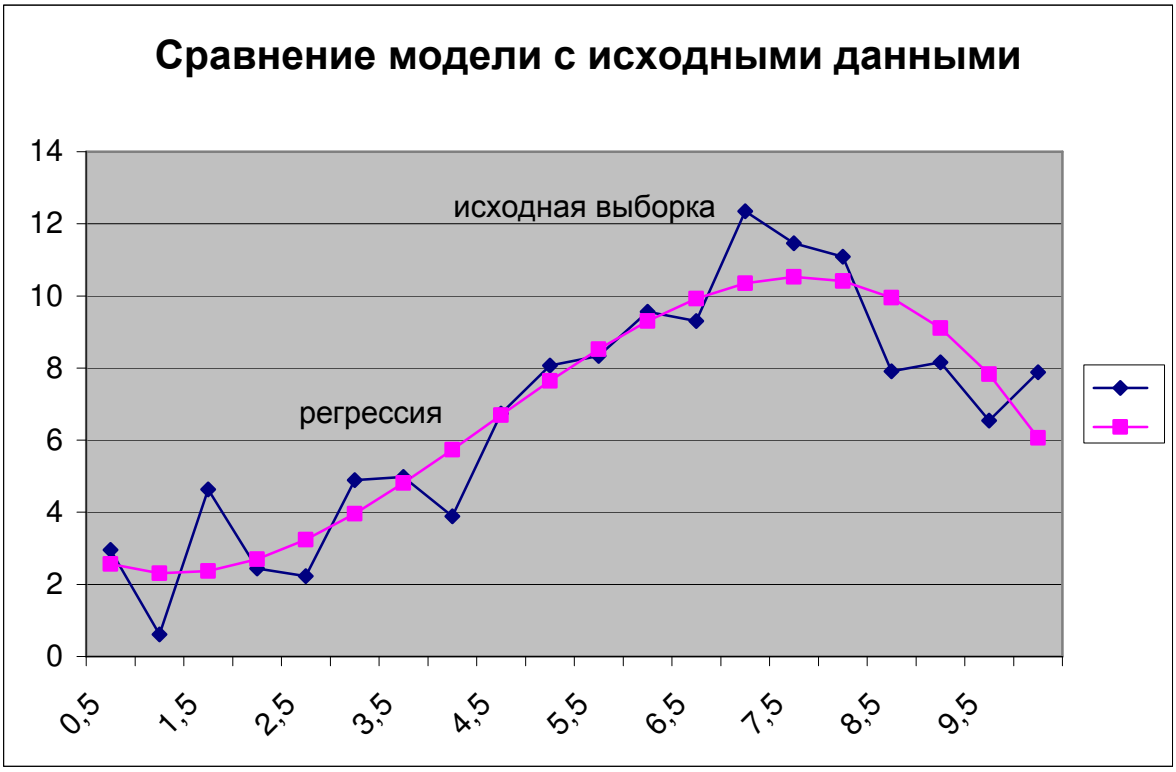



Рис. 3. Сравнение модели $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$ с исходными данными

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 значимо 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:

$$y = \beta_0 + \beta_2 x^2 + \beta_3 x^3$$

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 | 0,051733578 |
| Переменная X 1 | 0,469599546 | 0,060941974 | 7,70568321 | 6,05722E-07 |
| Переменная X 2 | -0,041727702 | 0,006223514 | -6,70484584 | 3,6971E-06 |

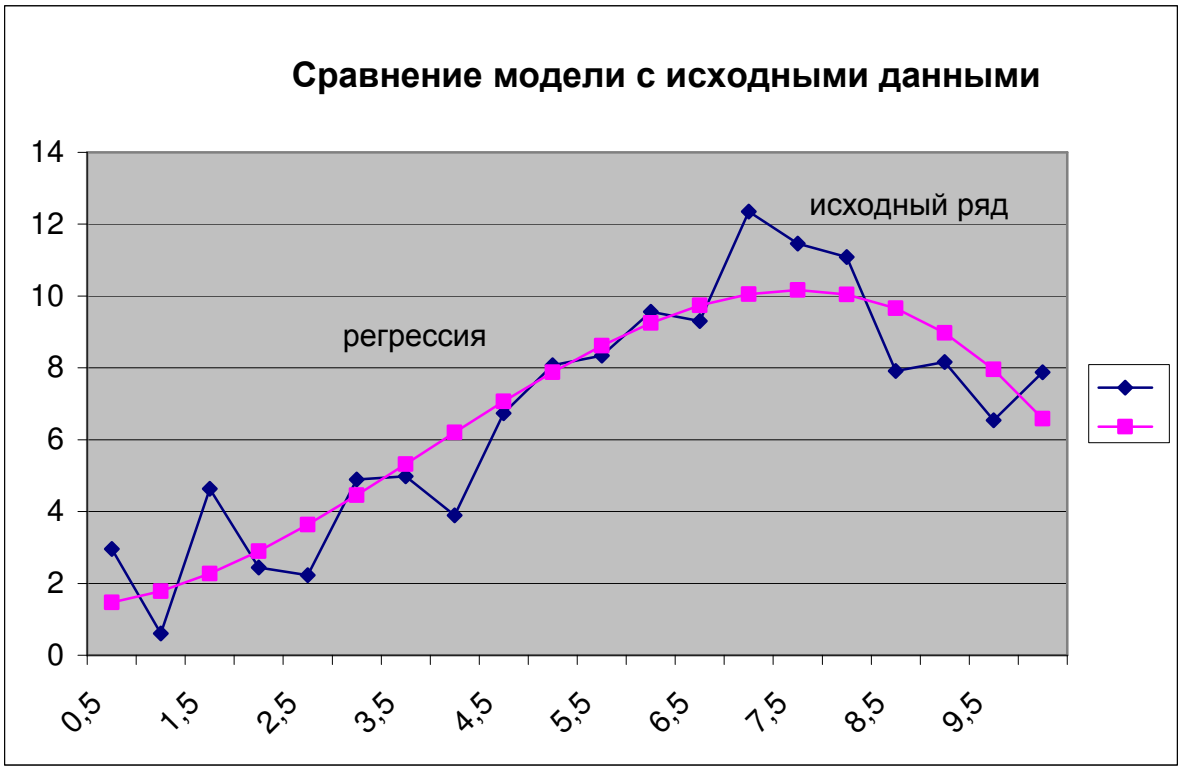



Рис. 4. Сравнение модели $y = \beta_0 + \beta_2 x^2 + \beta_3 x^3$ с исходными данными

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 |
| Наблюдения | 20 |

| | *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 | 0,006163641 |
| Переменная X 3 | -3,032825609 | 0,646493647 | -4,691191664 | 0,000245237 |

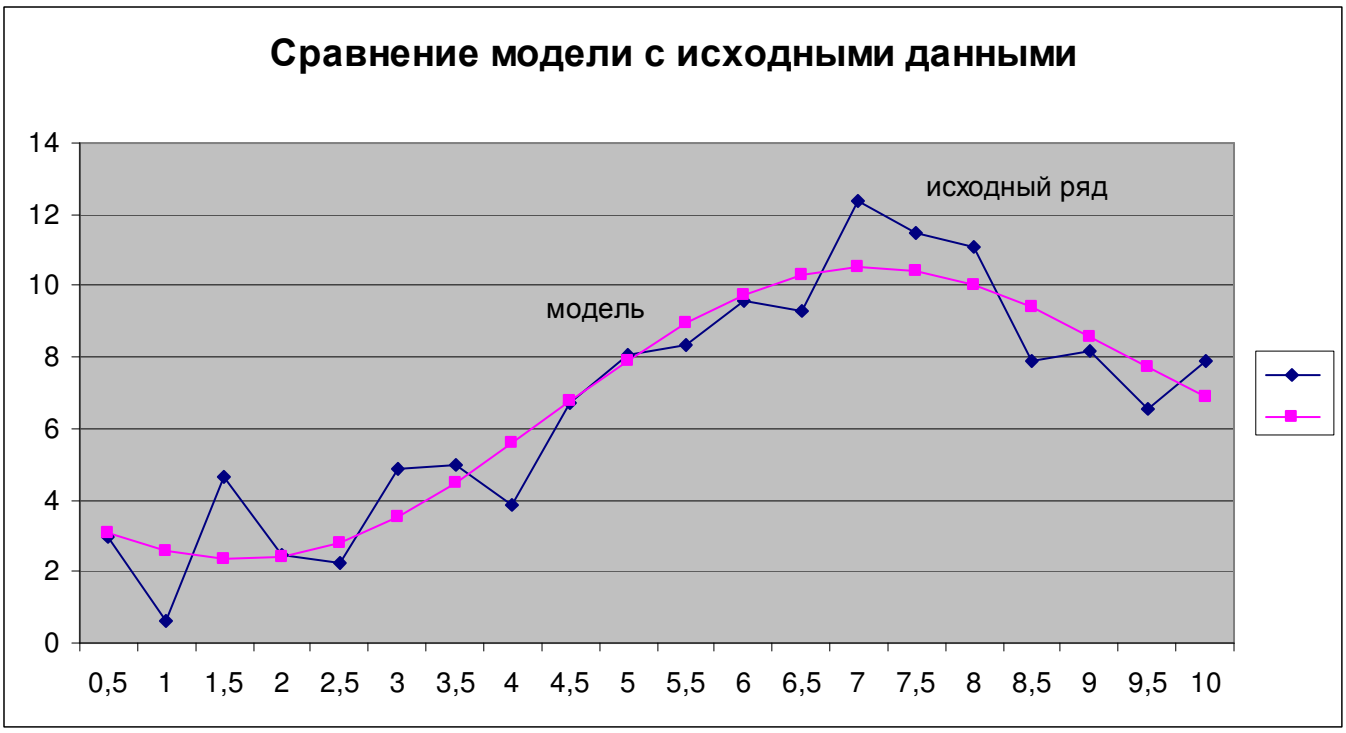



Рис. 5. Сравнение модели $y = \beta_0 + \beta_1 x + \beta_2 cos(2\pi x/10) + \beta_3 sin(2\pi x/10)$ с исходными данными

Заметим, что мы получили значение стандартной ошибки меньше заданой величины с. 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 $y = \beta_0 + \beta_1 x + \beta_2 Cos(2\pi x/10) + \beta_3 Cos(2\pi x/10)$.

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 условного среднего $\tilde{y} = M(Y \mid X = x)$ для тех же значений $X$, что приведены в выборке, найдем по формуле $\left[ \left( X^T B^* \right)_j \pm t_\alpha s \sqrt{\left( X (X^T X)^{-1} X^T \right)_{jj}} \right]$ (см. рис 6), где $t_\alpha$ квантиль распределение Стьюдента с $n-k-1$ степенью свободы (доверительный уровень возьмем 0,67, тогда $\alpha = 0,33$ и $t_\alpha = 1,0047$).

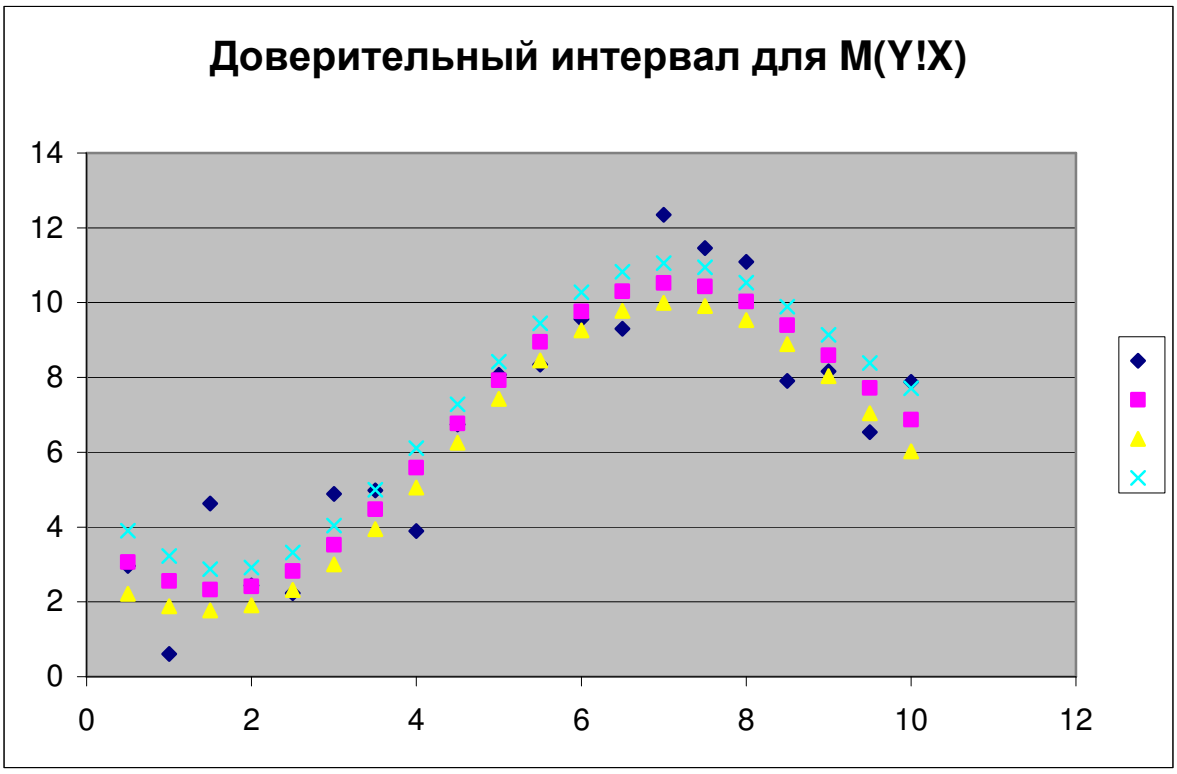



Рис. 6. Доверительный интервал для M(Y | X)

Доверительный интервал для значений $y$ рассчитываем по формуле $y_j = \left[ \left( X^T B^* \right)_j \pm t_\alpha s \left( 1 + \sqrt{\left( X (X^T X)^{-1} X^T \right)_{jj}} \right) \right]$. Доверительный интервал для значений $y$, полученный по этим формулам, отображен на рис 7.

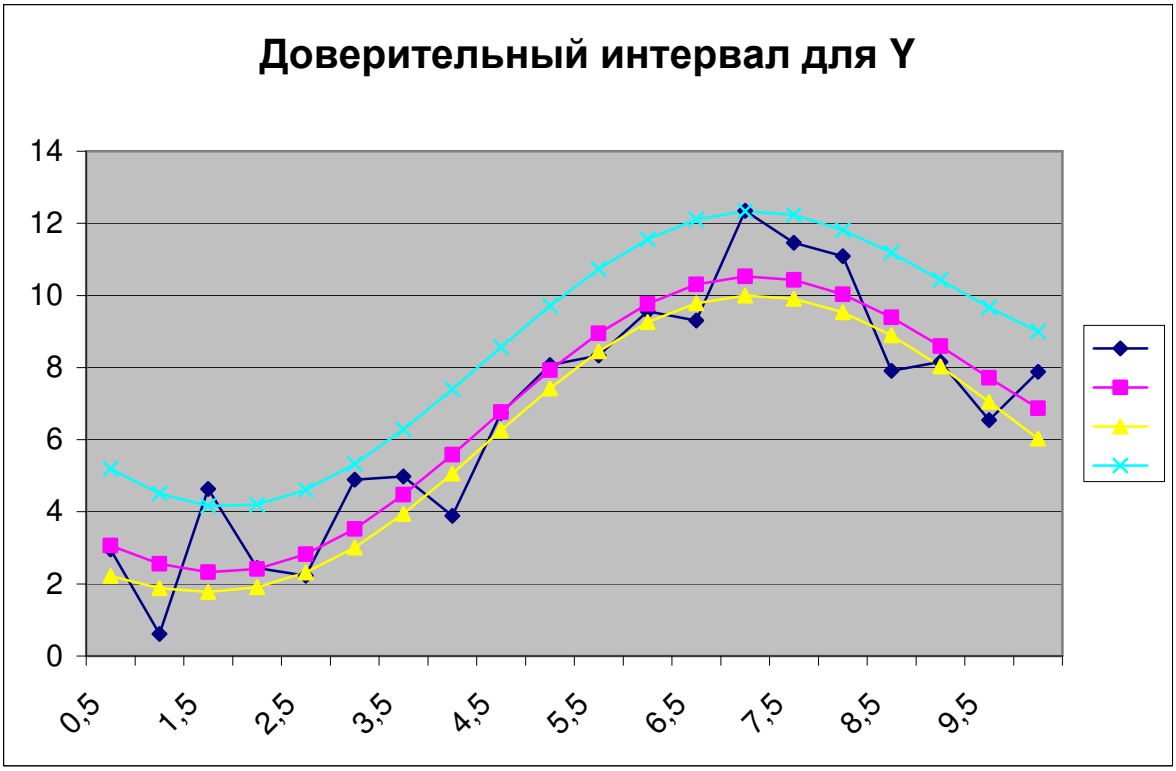



Рис. 7. Доверительный интервал для Y

## 5.3. Практическое 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 (дисперсию 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 $\tilde{y} = M(Y \mid X = x)$ и значений $y$ при каждом наблюдаемом значении $x$ (отобразить графически).

# Приложение к лабораторной работе 5

## 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 |

$\sigma = 0{,}5$

# Литература

1. Кобзарь А.И. Прикладная математическая статистика. – М.: ФИЗМАТЛИТ, 2012. – 813с.

2) Свешников А.А. Прикладные методы теории вероятностей. — Санкт-Петербург: Лань, 2012. – 480 с. [Электронный ресурс]. – Режим доступа: http://e.lanbook.com/books/element.php?pl1_cid=25&pl1_id=3184

3) Туганбаев А.А., Крупин В.Г. Теория вероятностей и математическая статистика — Санкт-Петербург: Лань, 2011. – 320 с. [Электронный ресурс]. – Режим доступа: http://e.lanbook.com/books/element.php?pl1_cid=25&pl1_id=652

4) Белов А.А., Баллод Б.А., Елизарова Н.Н. Теория вероятностей и математическая статистика: учебник для вузов. – Ростов н/Д: Феникс, 2008. – 318 с. (3 экз. в библиотеке ТУСУР)