

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Рукович Александр Владимирович

Должность: Директор

Дата подписания: 31.10.2014

Уникальный программный ключ:

f45eb7c44954caac05ea704f52eb8d7d6b5cb96aebd9b4bda094afdaaf0763f

Министерство образования и науки Российской Федерации
Технический институт (филиал) федерального государственного
автономного образовательного учреждения высшего профессионального
образования «Северо-Восточный федеральный университет
имени М.К. Аммосова» в г. Нерюнгри

**МЕТОДИЧЕСКИЕ УКАЗАНИЯ
К ЛАБОРАТОРНЫМ РАБОТАМ
ПО ДИСЦИПЛИНЕ
«СТАТИСТИЧЕСКИЙ ПАКЕТ ПРОГРАММ STATISTICA»**

ДЛЯ СТУДЕНТОВ НАПРАВЛЕНИЙ ОБУЧЕНИЯ
09.03.03 "ПРИКЛАДНАЯ ИНФОРМАТИКА",
ПРОФИЛЬ "ПРИКЛАДНАЯ ИНФОРМАТИКА В ЭКОНОМИКЕ"
01.03.02 "ПРИКЛАДНАЯ МАТЕМАТИКА И ИНФОРМАТИКА",
ПРОФИЛЬ – ОБЩИЙ
(очная форма обучения)

Нерюнгри
Издательство ТИ (ф) СВФУ
2015

Утверждено учебно-методическим советом ТИ (ф) ФГАОУ ВПО «СВФУ»

Составитель:

В.В. Юданова, старший преподаватель кафедры математики и информатики

Рецензент:

В.М. Самохина, к.п.н., зав. кафедрой математики и информатики

Подготовлено на кафедре математики и информатики

Печатается в авторской редакции

Методические указания по курсу «Статистический пакет программ Statistica» составлены для студентов направлений обучения 09.03.03 "Прикладная информатика", профиль "Прикладная информатика в экономике", 01.03.02 "Прикладная математика и информатика", профиль – общий. Подробно разобраны этапы проведения корреляционного, дисперсионного и регрессионного анализов в программе Statistica на примерах выполнения практических заданий и приведены варианты заданий для самостоятельного выполнения на лабораторных работах по данному курсу.

Методические указания предназначены для студентов, преподавателей и аспирантов, интересующихся статистическими компьютерными исследованиями.

© Технический институт (ф) СВФУ, 2015

ОГЛАВЛЕНИЕ

Введение	4
Раздел 1. Предварительный анализ и обработка данных в программе Statistica	5
Лабораторная работа 1. Моделирование распределений случайных величин и вычисление вероятностей	5
Практические задания к лабораторной работе 1	7
Лабораторная работа 2. Построение таблицы частот	8
Практические задания к лабораторной работе 2	11
Лабораторная работа 3. Проверка статистических гипотез. Критерии нормальности	11
Практические задания к лабораторной работе 3	14
Раздел 2. Методы корреляционного и дисперсионного анализа в программе Statistica	17
Лабораторная работа 4. Корреляционный анализ	17
Практические задания к лабораторной работе 4	21
Лабораторная работа 5. Однофакторный дисперсионный анализ	21
Практические задания к лабораторной работе 5	23
Раздел 3. Регрессионный анализ в программе Statistica	26
Лабораторная работа 6. Простая парная регрессия	26
Практические задания к лабораторной работе 6	30
Лабораторная работа 7. Множественная парная регрессия	30
Практические задания к лабораторной работе 7	33
Лабораторная работа 8. Пошаговая регрессия	33
Практические задания к лабораторной работе 8	36
Список литературы	37
Приложение	38

ВВЕДЕНИЕ

Система Statistica является одним из наиболее распространенных профессиональных пакетов статистической обработки данных в среде Windows. Пакет Statistica состоит из отдельных программ – модулей, каждый из которых содержит конкретный метод обработки данных, например, описательная статистика, регрессионный анализ и т.д. В данных методических указаниях рассматривается работа с теми модулями, которые наиболее часто используются в классическом статистическом исследовании.

Структура изложенных материалов представлена в виде трех разделов: Предварительный анализ данных, Методы корреляционного и дисперсионного анализа и Регрессионный анализ. В содержательной части методических указаний каждый раздел содержит подробное описание применения того или иного статистического метода исследования на примере решения различного вида задач. Также по каждой теме раздела приведены практические задания по вариантам для самостоятельного выполнения на лабораторных работах по дисциплине «Статистический пакет программ Statistica».

Методические указания предназначены для студентов направлений обучения 09.03.03 «Прикладная информатика» и 01.03.02 «Прикладная математика и информатика» и также могут быть использованы преподавателями ВУЗов, аспирантами интересующихся статистическими компьютерными исследованиями.

РАЗДЕЛ 1. ПРЕДВАРИТЕЛЬНЫЙ АНАЛИЗ И ОБРАБОТКА ДАННЫХ В ПРОГРАММЕ STATISTICA

Лабораторная работа 1

Тема: Моделирование распределений случайных величин и вычисление вероятностей

Цель: закрепить знания по использованию методов получения выборки из генеральной совокупности с заданным законом распределения и по вычислению критических значений заданного распределения с помощью вероятностного калькулятора.

Генерация случайных чисел

Генерация случайных чисел означает получение последовательности случайных чисел, что, в частности, требуется для моделирования объектов, имеющих случайную природу, по известному распределению вероятностей.

$Rnd(x)$ – функция генерации случайных чисел, равномерно распределенных на отрезке $[0, x)$; $Rnd(b-a)+a$ генерирует числа, распределенные равномерно на отрезке $[a, b]$.

Пример: Сгенерировать выборку из 15 случаев согласно нормального закона распределения с заданными параметрами: μ (математическое ожидание) = 2, σ (стандартное отклонение) = 3.

Ход работы

1. Создайте файл данных в программе Statistica содержащий одну переменную Var1 и 15 случаев наблюдений.
2. Откройте окно спецификации переменной Var1 в котором введите в поле Name (имя): Normal, Type (тип): Number (число), Decimal places (дес. разр.): 4, и в поле Long name (Длинное имя (метка, связь или формула): =Vnormal(rnd(1);2;3) и нажмите Ок.
3. Проверьте в файле данных заполнение столбца Normal случайными вещественными числами из $[0; 1]$ с точностью до 4 знаков после запятой.

Вероятностный калькулятор

Вероятностный калькулятор или калькулятор вероятностных распределений – позволяет на основе заданных пользователем различных параметров того или иного распределения вычислить его критические значения.

Пример: Вычислить критические значения нормального распределения при различных значениях параметров: математическое ожидание, стандартное отклонение и доверительная вероятность и сделать вывод об изменении графика плотности распределения.

Ход работы

1. Запустите вероятностный калькулятор: Statistics (Анализ) – Basic Statistics/Tables (Основные статистики и таблицы) – Probability calculator (Вероятностный калькулятор).
2. В диалоговом окне калькулятора вероятностных распределений выберите – Distribution (распределение): Z (нормальное) и введите параметры: mean (среднее): 1, st.dev. (стандартное отклонение): 0, p (доверительная вероятность) введите значение 0,5; установите галку Inverse (обратная ф.р.) (рис. 1.1.).

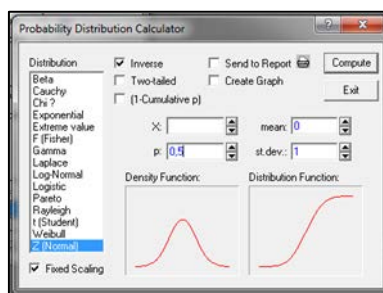


Рис. 1.1. Вероятностный калькулятор: работа с нормальным распределением

3. Нажмите кнопку Compute (Вычислить) – в строке Z появится соответствующий квантиль (аналогично при изменении Z команда Compute (Вычислить) изменит значение для p).
4. Выберите опцию Create Graph (Создать график) и нажмите кнопку Compute (Вычислить), сохраните полученные графики функции плотности и функции распределения.
5. Увеличивая и уменьшая значение mean (среднее), следите за изменением графиков. Сохраните полученные графики.
6. Сделайте вывод: как изменяется график плотности нормального распределения по оси ординат при изменении среднего значения.

Пример: Вычислить критическое значение χ^2 (хи-квадрат) – распределения при следующих параметрах: число степеней свободы 4, доверительная вероятность 0,95.

Ход работы

1. В диалоговом окне калькулятора вероятностных распределений выберите: Distribution (Распределение): Chi I (хи-квадрат-распределение) и введите параметры хи-квадрат-распределения: df (число степеней свободы): 4, p (вероятность): 0,95, установите галочки Inverse (обратная ф. р.) и 1 – cumulative p (1 – p) (или укажите p: 0,05 и снимите галочку 1 – cumulative p (1 – p)) (рис. 1.2.).

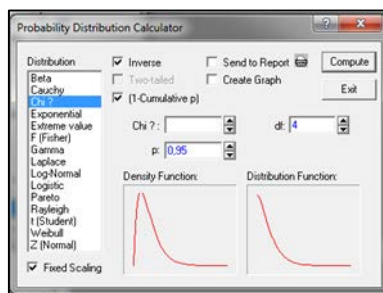


Рис. 1.2. Вероятностный калькулятор: работа с распределением χ^2

2. Нажмите кнопку Compute (Вычислить) – в строке Chi I появится соответствующее критическое значение (аналогично при изменении Chi I команда Compute (Вычислить) изменит значение для p).
3. Проверьте полученное значение по таблице распределения χ^2 .

Практические задания к лабораторной работе 1

Генерация случайных чисел

1. Сформировать выборку из 10 случайных чисел, лежащих в диапазоне от 0 до 1.
2. Сформировать выборку из 20 случайных чисел, лежащих в диапазоне от 5 до 20.
3. Бросают несимметричную монету, для которой вероятность успеха, – выпадения герба, составляет $p = 0,3$. Построить случайную последовательность выпадения герба (или 1) или решки (или 0) на 20 испытаний.
4. Составить случайную последовательность для выпадения шестерки при подбрасывании игральной кубика в 20 испытаниях.
5. Сгенерировать выборки согласно следующим законам распределения $R[0;2]$, $N(2;0.25)$, $E(1/7)$. (привести таблицу к размерам $3v \times 50c$ и задать имена переменных Var1 – Rav, Var2 – Normal, Var3 – Expon).

Вероятностный калькулятор

6. Известно, что в некоторой стране рост взрослых мужчин приближенно имеет нормальное распределение со средним 176,6 см и стандартным отклонением 7,63. Произвольным образом выбирается мужчина. Какова вероятность, что рост этого наугад выбранного мужчины не больше 185 см и не меньше 175 см?
7. Составьте таблицу критических значений хи-квадрат-распределения с заданным числом степеней свободы ($k = 19$) для вероятностей 0,01; 0,05; 0,1; 0,9; 0,95; 0,99. Проверьте полученные значения проверьте по таблице распределения χ^2
8. Составьте таблицу критических значений t-распределение Стьюдента с заданным числом степеней свободы ($k = 10$) для вероятностей 0,01; 0,025; 0,05; 0,1; 0,25; 0,4. Проверьте полученные значения проверьте по таблице t- распределения.
9. Составьте таблицу критических значений F – распределения с заданным числом степеней свободы ($k_1 = 10$, $k_2 = 15$) для вероятностей 0,01; 0,025; 0,05; 0,1. Проверьте полученные значения проверьте по таблице F – распределения.
10. Для случайной величины X , имеющей нормальное распределение $N(1, 2)$, т.е. μ (математическое ожидание) = 1, σ (стандартное отклонение) = 2, составить таблицу значений вероятностей следующих событий: $P(X < 2)$; $P(X \geq 3)$; $P(0 < X < 3)$; $P(|X-1| < 1)$; $P(|X-2| > 1)$. Проверить полученные значения для вероятности событий $P(X < 2)$ и $P(X \geq 3)$ по таблице значений функции Лапласа, используя формулы:

вают на k -непересекающихся интервалов. Затем определяют **частоты** – количество n_i элементов выборки, попавших в i -ый интервал (элемент, совпадающий с правой границей интервала попадает в следующий интервал). Наряду с частотами также подсчитывают:

		Frequency table: Знач (Spreadsheet4)			
From	To	Count	Cumulative Count	Percent	Cumulative Percent
10,10000	<=x<12,10000	2	2	3,63636	3,6364
12,10000	<=x<14,10000	4	6	7,27273	10,9091
14,10000	<=x<16,10000	8	14	14,54545	25,4545
16,10000	<=x<18,10000	12	26	21,81818	47,2727
18,10000	<=x<20,10000	16	42	29,09091	76,3636
20,10000	<=x<22,10000	10	52	18,18182	94,5455
22,10000	<=x<24,10000	3	55	5,45455	100,0000
24,10000	<=x<26,10000	0	55	0,00000	100,0000
Missing		0	55	0,00000	100,0000

Рис. 1.3. Таблица частот: размер шага 2

6. Для наглядного представления выборки постройте гистограмму и полигон частот: в диалоговом окне Frequency tables (Таблицы частот) нажмите кнопку Histograms (Гистограммы). Сохраните полученный график в отчет Отчет_лаб2.str (рис. 1.4.).

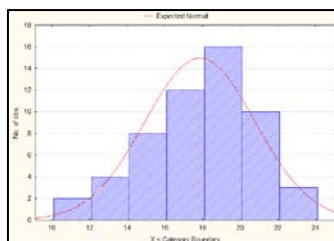


Рис. 1.4. Гистограмма и полигон частот

7. Постройте процентную огиву (многоугольник частот) – это ломаная, вершины которой имеют абсциссы, совпадающие с правыми границами интервалов группировки и ординаты, совпадающие со значениями накопленных частот в процентах для соответствующих интервалов: в построенной таблице частот, выделите столбец Cumulative percent (кумулятивные проценты) – вызовите контекстное меню – Grafs of blok data (График для блока данных) – Line plot: Entire columns. Полученный график сохраните в отчет Отчет_лаб2.str (рис. 1.5.).



Рис. 1.5. Процентная огива или многоугольник частот

8. В отчете по построенной процентной огиве сделайте вывод: какие значения не превышают 50% всех наблюдений и сколько процентов составляют значения, не превышающие число 16,1.

Практические задания к лабораторной работе 2

По заданной выборке выполнить следующие задания (выполнение задания сохранять в отчет):

1. Построить вариационный ряд.
2. Вычислить оценки математического ожидания, моды и медианы, несмещенную оценку дисперсии, размах выборки.
3. Построить таблицу частот для сгруппированной выборки (количество интервалов равно 4, длину интервала определить).
4. Построить гистограмму и полигон частот.
5. Построить огиву и сделать по ней вывод.

Варианты заданий

№	Значения выборки																			
1	1	2	2	4	3	3	1	1	4	2	1	3	2	1	1	2	2	5	6	7
2	0	6	2	3	5	8	3	2	1	9	4	4	9	1	3	2	6	1	2	4
3	2	4	1	3	1	2	2	2	5	1	5	1	4	2	0	4	3	3	1	0
4	2	4	5	3	1	7	5	4	5	7	6	6	5	2	2	1	4	7	2	1
5	4	4	8	5	9	3	9	3	3	3	7	5	3	6	7	8	4	6	5	9
6	8	4	9	3	7	3	4	9	6	9	5	8	4	7	7	4	3	5	4	9
7	3	7	9	5	7	9	9	5	2	6	7	4	3	9	8	8	8	6	5	6
8	14	10	9	6	9	6	6	8	10	9	11	7	8	9	7	12	8	7	8	13
9	8	3	3	10	4	10	8	12	5	11	5	5	6	6	4	11	7	3	4	11
10	7	8	12	11	8	8	12	10	8	12	5	11	9	10	11	10	12	8	7	9

Лабораторная работа 3

Тема: Проверка статистических гипотез. Критерии нормальности.

Цель: закрепить знания, умения и навыки по использованию критериев Колмогорова-Смирного, Лиллиефорса, Шапиро-Уилкса и критерия χ^2 (хи-квадрат) для проверки соответствия данных нормальному закону распределения.

Критерии нормальности дают возможность проверить имеют ли исходные данные нормальный закон распределения. Проверка закона распределения необходима, т.к. большинство методов математической статистики (корреляционный, регрессионный, дискриминантный анализ и др.) разработаны применительно к случаям нормального распределения.

Пример: Проверить гипотезу о нормальном распределении для выборки из 55 наблюдений с помощью критериев Колмогорова-Смирного, Лиллиефорса и Шапиро-Уилкса:

18,3	15,4	17,2	19,2	23,3	18,1	21,9
15,3	16,8	13,2	20,4	16,5	19,7	20,5
14,3	20,1	16,8	14,7	20,8	19,5	15,3
19,3	17,8	16,2	15,7	22,8	21,9	12,5

10,1	21,1	18,3	14,7	14,5	18,5	18,4
13,9	19,1	18,5	20,2	23,8	16,7	20,4
19,5	17,2	19,6	17,8	21,3	17,5	19,4
17,8	13,5	17,8	11,8	18,6	19,1	

Ход работы

1. Создайте новый файл данных в программе Statistica (1v*55с).
2. В переменную Var1 введите выборку из 55 наблюдений и сохраните ее в отчет: Add to report (Добавить в отчет) – New report (Новый отчет). Сохраните отчет: Отчет_лаб3.str.
3. Откройте диалоговое окно Frequency tables (Таблицы частот) и выберите вкладку Normality (Критерии нормальности), установите галки для критериев – Kolmogorov-Smirnov test (Колмогоров-Смирнов), Lillefors test (Лиллиефорс) и Shapiro-Wilk's W test (Шапиро-Уилкс) и нажмите кнопку Tests for normality (Критерии нормальности).
4. Полученную таблицу результатов тестирования выборки на нормальность по трем критериям сохраните в Отчет_лаб3.str и сделайте вывод: подчиняется ли выборка нормальному распределению (сравнить полученное значение p – вероятность отклонения гипотезы о значимости статистики критерия с выбранным уровнем значимости α (0,05)) (рис. 1.6.).

Tests of Normality (Spreadsheet2)						
Variable	N	max D	K-S p	Lilliefors p	W	p
Var1	55	0,091107	p > .20	p > .20	0,986504	0,791935

Рис. 1.6. Результаты тестирования выборки на нормальность по критериям Колмогорова-Смирнова, Лиллиефорса и Шапиро-Уилкса

Пример: Проверить гипотезу о нормальном, логнормальном и гамма-распределениях для данной выборки из 55 наблюдений с помощью критерия χ^2 (хи-квадрат).

Ход работы

1. Работа продолжается с файлом Отчет_лаб3.str и с выборкой из предыдущего примера.
2. Выполните команду: Statistics (Анализ) – Distribution Fitting (Подбор распределений). В стартовой панели Distribution Fitting (Подбор распределений) в списке Continuous Distributions (Непрерывные распределения) выберите Normal (Нормальное распределение) и нажмите Ok (рис. 1.7.).

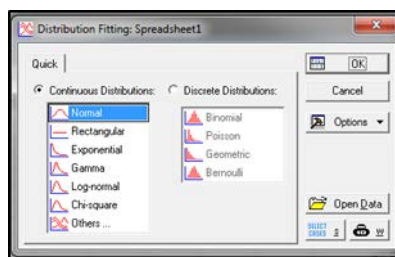


Рис. 1.7. Стартовая панель Подбора распределений

3. В диалоговом окне Fitting Continuous Distributions (Подбор непрерывного распределения) нажмите кнопку Variable (Переменная) и выберите столбец, содержащий исходную выборку. Перейдите на вкладку Parameters (Параметры): после выбора переменной здесь вычисляются оценки математического ожидания и дисперсии и по умолчанию число интервалов группировки устанавливается 18 (значение можно менять). Нажмите Ок.
4. В построенной таблице для расчета статистики χ^2 важны столбцы: observed frequency (наблюдаемые частоты) и expected frequency (ожидаемые частоты), а также столбец разности – observed expected; в заголовке таблицы важны значения статистики χ^2 (Chi-Square): 1,83940, количество степеней свободы $df = 4$, значение вероятности $p = 0,76527$.
5. Вернитесь в диалоговое окно Fitting Continuous Distributions (Подбор непрерывного распределения) и постройте гистограмму частот, показывающую результат подгонки эмпирического распределения к теоретическому. Заголовок гистограммы совпадает с заголовком таблицы значений статистики χ^2 . Полученную гистограмму сохраните в Отчет_лаб3.str (рис. 1.8.).

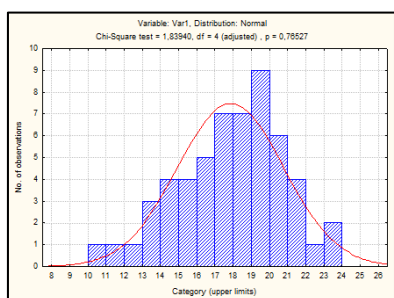


Рис. 1.8. Гистограмма частот подгонки эмпирического распределения к теоретическому с результатами оценки по критерию χ^2

6. В отчете сделайте вывод о соответствии проверяемых данных нормальному закону распределения, сравнив полученное значение p – вероят-

ности отклонения гипотезы о значимости статистики критерия χ^2 с выбранным уровнем значимости α (0,05).

7. Аналогично по критерию χ^2 (хи-квадрат) проверьте подчиняются ли исходные данные логнормальному и гамма-распределению. Постройте в обоих случаях гистограммы частот, показывающие результаты подгонки эмпирического распределения к теоретическим, сохраните их в Отчет_лаб3.str и сделайте вывод о соответствии выборки данным законам распределения.

8. В отчете сформулируйте общий вывод по использованию критерия χ^2 : какое из трех распределений (нормальное, логнормальное или гамма), больше соответствует исходной выборке (сравните между собой полученные значения p – вероятности отклонения гипотезы о значимости статистики критерия χ^2 полученные для каждого вида распределения).

Практические задания к лабораторной работе 3

Для задачи своего варианта выполните следующие задания. Выполнение заданий сохраните в отчет.

1. Определите оценки среднего, дисперсии, медианы, коэффициентов асимметрии и эксцесса.
2. Постройте 90% доверительные интервалы для среднего.
3. Произведите группировку данных, используя 5 и 8 интервалов.
4. По сгруппированным данным построьте гистограмму и сделайте заключение о близости полученного распределения к нормальному. Проверьте себя с помощью критериев Колмогорова-Смирного, Лиллиефорса и Шапиро-Уилкса.
5. По критерию χ^2 (хи-квадрат) проверьте гипотезы о виде распределения наилучшим образом подходящее к исходной выборке.
6. Ответьте на вопросы задачи.

Варианты заданий

1. Приведен вес в килограммах 100 пациентов, желающих пройти курс лечения, чтобы снизить вес.

103	90	95	106	101	79	98	91	79	87
120	93	88	111	82	84	86	81	86	98
79	83	91	108	105	117	107	97	94	101
106	93	82	121	107	84	87	99	88	111
86	82	79	83	106	106	82	91	85	114
70	79	89	78	112	90	103	82	79	84
98	86	96	90	96	103	83	89	96	99
100	97	87	77	117	87	88	110	104	82

82	61	110	82	95	92	110	108	103	117
94	99	104	102	103	85	95	89	77	93

Используя данные постройте огиву. Определите процент пациентов, имеющих вес более чем 100 кг.

2. В течение 50 дней фиксировалось время для набора титульного листа газеты. Данные представлены в таблице.

20,8	28,0	21,9	20,0	20,7	20,9	25,0	22,0	28,0	20,1
25,3	20,7	25,0	21,2	23,8	23,3	20,9	29,0	23,5	19,5
23,7	20,3	23,6	19,0	25,1	25,0	19,5	24,1	24,2	21,8
21,3	21,5	23,1	19,9	24,2	24,1	19,8	23,9	28,0	23,9
19,7	24,2	23,8	20,7	23,8	24,3	21,1	20,9	21,6	27

Используя данные, постройте огиву. Определите процент случаев, в которых страница набирается не более чем за 24 минуты.

3. Менеджер компании фиксирует время (в мин.), которое идет на переналадку и текущий ремонт оборудования в шахте в течение рабочей смены. Результаты 35 последних наблюдений приведены в таблице.

60	72	126	110	91	115	112
80	66	101	75	93	129	105
113	121	93	87	119	111	97
102	116	114	107	113	119	100
110	99	139	108	128	84	99

Известно, что среднее время простоя оборудования составляет 108 минут. Постройте огиву и определите во скольких случаях оборудование простаивало более 108 мин., а во скольких менее?

4. Производительность труда бригады шахтеров (в тоннах угля за смену) представлена в таблице.

356	331	299	391	364	317	386
360	281	360	402	411	390	362
311	357	300	375	427	370	383
322	380	353	371	400	379	380
369	393	377	389	430	340	368

Используя данные, постройте огиву. Определите, во скольких случаях производительность была ниже 380 т за смену, а во скольких – выше?

5. Имеются следующие данные о величине процентной ставки коммерческих банков.

14,7	19,0	24,5	20,8	12,3	24,6	17,0	14,2	19,7	18,8
18,1	20,5	20,7	20,7	20,4	14,7	25,1	22,7	19,0	19,6
19,0	18,9	20,0	20,0	13,8	25,6	13,0	19,0	18,7	21,1
13,3	20,7	19,9	19,9	21,9	16,0	16,9	15,3	21,4	20,4
12,8	20,0	18,0	18,0	15,1	23,8	18,5	14,4	21,0	19,0

Используя данные, постройте огиву. Определите, какой интервал имеет наибольшую частоту и какую процентную ставку имеют 50% банков.

6. В результате выборочного исследования семей одного из городов РФ получены данные о месячном среднедушевом доходе (в руб.).

4001	2120	3450	5080	4150	1042	851	2534
4405	2837	3412	2212	2902	3935	3267	3012
1388	1701	2346	1190	1083	917	5890	1450
2125	6084	1257	2835	2634	784	1293	2604
1050	2187	2538	4747	811	4165	1358	1025
2851	3450	879	2156	2713	854	745	2127

Используя данные, постройте огиву и сделайте экономические выводы.

7. Получены данные, характеризующие урожайность зерновых культур сельскохозяйственных предприятий (ц/га).

31,2	22,1	23,5	23,5	31,4	29,9	24,5	27,5	23,6	22,0
14,3	34,1	27,2	24,5	32,0	32,0	20,0	24,2	22,5	22,5
13,8	23,9	17,5	31,6	32,5	25,7	25,5	22,0	23,1	29,0
26,3	14,8	21,8	30,7	26,5	20,5	29,6	21,5	18,2	27,5
23,4	25,1	15,5	33,0	26,4	31,2	28	22,5	17,5	22,0

Используя данные, постройте огиву и сделайте соответствующие выводы.

8. В результате выборочного исследования получена продолжительность работы ламп.

13,4	14,7	15,2	15,1	8,8	14,0	17,9	15,1	16,5	14,2
16,3	14,6	11,7	15,1	17,6	14,1	18,8	11,6	16,6	18,0
12,4	17,2	14,2	16,3	13,7	15,5	14,0	13,9	11,3	10,7
16,9	15,8	16,1	12,3	14,9	14,7	16,2	8,4	14,7	15,4
10,1	15,8	18,3	17,5	12,7	17,7	14,7	16,2	17,1	17,7

Используя данные, постройте огиву и сделайте соответствующие выводы.

9. Положительные отклонения (в мм) от номинального размера у партии деталей приведены в таблице.

177	121	168	120	123	128	212	210	117	208
122	179	119	210	129	119	117	211	117	118
112	117	114	116	121	119	117	123	121	127
126	115	119	120	116	118	211	221	110	120
210	115	119	210	211	221	213	119	114	211

Используя данные, постройте огиву и определите, какой интервал отклонений имеет наибольшую частоту и какие отклонения имеют 50% деталей.

10. Ошибки при стрельбе в спортивной команде по наземной цели имеют следующие результаты (в рад.).

0,926	1,375	0,785	-0,963	1,022	-0,472	1,279	3,521	0,571	-1,851
0,194	1,192	1,394	-0,555	0,046	0,321	2,945	1,974	0,258	-0,941
1,192	1,394	-0,555	0,046	0,321	2,945	1,974	0,258	-0,941	1,192
-0,523	0,525	0,595	0,888	-0,934	1,579	0,161	0,412	0,906	0,007
0,769	0,971	0,712	1,09	-0,631	-1,501	-0,488	-0,162	-0,136	1,033

Постройте огиву и определите какие ошибки допущены в 50% случаях.

РАЗДЕЛ 2. МЕТОДЫ КОРРЕЛЯЦИОННОГО И ДИСПЕРСИОННОГО АНАЛИЗА В ПРОГРАММЕ STATISTICA

Лабораторная работа 4

Тема: Корреляционный анализ

Цель: получить навыки по применению методов корреляционного анализа для вычисления парных корреляций, множественного коэффициента корреляции и коэффициента детерминации, а также частных корреляций с помощью инструментария программы Statistica.

Корреляционный анализ представляет собой статистический метод выявления взаимосвязей между несколькими признаками. Основные требования: используемые для анализа наблюдения должны быть случайно выбранными из нормально распределенной генеральной совокупности; число наблюдений должно превышать кол-во анализируемых признаков в 3-4 раза.

Пример. Имеются данные о результатах наблюдений за деятельностью 23 предприятий одной из отраслей промышленности: рентабельность в % (Рентабел), удельный вес рабочих в составе ППП, ед. (Доля_раб), фондоотдача, ед. (ФО), среднегодовая стоимость основных производственных фондов, млн. руб. (Осн_фонды), непроизводственные расходы, тыс. руб. (Непр_расх). Требуется исследовать зависимость рентабельности от других показателей методами корреляционного анализа (данные выборки подчиняются нормальному закону распределения).

	1	2	3	4	5
	Рентабел	Доля_раб	ФО	Осн_фонды	Непр_расх
1	13,9	0,59	1,96	45,44	24,43
2	9,68	0,71	1,02	41,08	22,79
3	10,03	0,55	1,85	98,56	28,49
4	9,13	0,73	0,88	42,39	23,71
5	5,37	0,6	0,62	37,39	22,79
6	9,86	0,81	1,08	65,34	18,43
7	12,62	0,82	1,6	47,55	17,49
8	5,02	0,8	1,53	32,46	17,46
9	21,18	0,79	1,4	94,25	16,78
10	25,17	0,77	2,21	41,76	15,71
11	19,4	0,79	1,33	74,59	15,79
12	21	0,81	1,45	61,28	16,78
13	6,57	0,73	0,71	57,16	19,48
14	14,19	0,75	2,1	92,46	17,79
15	15,81	0,76	1,45	55,19	18,49
16	5,23	0,81	1,51	57,16	14,78
17	7,99	0,76	1,42	64,19	21,79
18	17,5	0,75	1,8	61,49	17,46
19	17,16	0,76	2,61	45,18	19,75
20	14,54	0,79	1,73	85,46	16,47
21	6,24	0,69	1,56	29,46	25,47
22	12,08	0,73	2,1	34,72	17,46
23	9,49	0,72	1,17	94,76	24,46

Ход работы

Построение корреляционной матрицы

1. Создайте новый файл данных в программе Statistica (5v*23с) и сохраните полученную таблицу в отчет: Отчет_лаб4.str.
2. Вычислите парные коэффициенты корреляции: Statistics (Анализ) – Basic Statistics/Tables (Основные статистики и таблицы) – Correlation matrices (Корреляционная матрица). В открывшемся диалоговом окне корреляционного анализа выберите One variable list (Один список переменных), для построения квадратной матрицы коэффициентов корреляции для всех переменных. Сохраните полученную матрицу парных коэффициентов корреляции в Отчет_лаб4.str (рис. 2.1.).

Correlations (Spreadsheet1)					
Marked correlations are significant at p < .05000					
N=23 (Casewise deletion of missing data)					
Variable	Рентабел	Доля раб	ФО	Осн_фонды	Непр_расх
Рентабел	1.00	0.29	0.49	0.21	-0.46
Доля раб	0.29	1.00	0.05	0.03	-0.82
ФО	0.49	0.05	1.00	0.02	-0.20
Осн_фонды	0.21	0.03	0.02	1.00	0.02
Непр_расх	-0.46	-0.82	-0.20	0.02	1.00

Рис. 2.1. Матрица парных коэффициентов корреляции

3. В отчете сделайте вывод о наличии мультиколлинеарности (существует связь между самими независимыми переменными (коэффициент корреляции $r > 0.7$) – непрямые расходы и доля рабочих) и выберите переменную, которую необходимо исключить из дальнейшего анализа (которая не имеет сильной зависимости от рентабельности).
4. Постройте новую квадратную матрицу коэффициентов корреляции без исключенной переменной, сохраните полученную матрицу в Отчет_лаб4.str и оцените связь между рентабельностью и остальными переменными: прямая или обратная связь и теснота связи (по шкале Чеддока: $0.1 < r_{xy} < 0.3$: слабая; $0.3 < r_{xy} < 0.5$: умеренная; $0.5 < r_{xy} < 0.7$: заметная; $0.7 < r_{xy} < 0.9$: высокая; $0.9 < r_{xy} < 1$: весьма высокая). Сделайте вывод в отчете какие переменные мало влияют на рентабельность и исключите их из дальнейшего анализа.
5. Поработайте с диалоговым окном построения корреляционной матрицы и постройте квадратные корреляционные матрицы с вычислением p – вероятности отклонения гипотезы о значимости коэффициента корреляции (Display r, p-levels, and N's), с подробной таблицей результатов (Display detailed table of results) и постройте прямоугольную корреляционную матрицу – Two list (rect. Matrix) (зависимая переменная Рентабел, остальные не-

зависимые). Сохраните полученные матрицы в Отчет_лаб4.str с соответствующими комментариями.

Графическое отображение корреляционной зависимости

6. С помощью команд диалогового окна корреляционного анализа постройте графическое изображение корреляционных зависимостей между рентабельностью (y) и наиболее тесно связанными с ней признаками: фондоотдачей (x1) и непроизводственными расходами (x2) (для которых $r_{x1y} = 0.49$ и $r_{x2y} = -0.46$ соответственно): диаграмму 2М-рассеяние, матричный график, корреляционное поле 3М-рассеяние, поверхность и 3D-гистограммы (рис. 2.2.). Сохраните полученные графики в Отчет_лаб4.str с соответствующими комментариями.



Рис. 2.2. Диалоговое окно корреляционного анализа

Расчет множественного и частных коэффициентов корреляции

7. Вызовите модуль Множественной регрессии: Statistics (Анализ) -> Multiple regression (Множественная регрессия). В диалоговом окне Multiple regression (Множественная регрессия) выберите переменные для анализа: зависимая – рентабельность, независимые – фондоотдача и непроизводственные расходы; в поле Input file (Файл ввода) укажите в качестве исходных данных таблицу наблюдений (Raw Data), чтобы продолжить анализ нажмите Ок.

8. В открывшемся диалоговом окне Multiple regression results (Результаты множественной регрессии) рассмотрите полученные результаты и выделите коэффициент множественной корреляции (R), коэффициент детерминации (R2) и результаты F-критерия Фишера (рис. 2.3.).



Рис. 2.3. Окно результатов множественной регрессии

9. Скопируйте результаты множественной регрессии в Отчет_лаб4.str и сделайте следующие выводы:

- F-критерий Фишера показывает, что вероятность отклонения гипотезы о значимости множественного коэффициента корреляции $p = 0,014599$ не превышает уровень значимости $\alpha = 0,05 \Rightarrow$ множественный коэффициент корреляции $R = 0,64622546$ является значимым на уровне $\alpha = 0,05$.

- множественный коэффициент корреляции $R = 0,64622546 \Rightarrow$ коэффициент детерминации $R^2 = 0,4176073 \Rightarrow$ лишь 41,7% дисперсии показателя рентабельность объясняется изменением показателей фондоотдачи и непроизводственных расходов, что говорит о недостаточности числе факторов, введенных в модель.

10. В диалоговом окне Multiple regression results (Результаты множественной регрессии) нажмите кнопку Partial correlations (Частные корреляции) и рассмотрите полученную таблицу (рис. 2.4.).

Variable	Beta in	Partial Cor.	Semipart Cor.	Tolerance	R-square	t(20)	p-level
ФО	0.416197	0.458511	0.407904	0.960546	0.039454	2.30736	0.031858
Непр_расх	-0.374018	-0.420639	-0.366565	0.960546	0.039454	-2.07352	0.051255

Рис. 2.4. Таблица результатов расчета частных коэффициентов корреляции

11. Скопируйте таблицу частных корреляций в Отчет_лаб4.str и сделайте выводы:

- частный коэффициент корреляции $r_{x1y|x2} = 0,46$ между рентабельностью (y) и фондоотдачей (x1) без учета непроизводственных расходов (x2): является значимым на уровне $\alpha = 0,05$, т.к. вероятность отклонения гипотезы о значимости коэффициента по t-критерию $p = 0,03 < \alpha = 0,05$ и показывает, что после ввода в модель показателя непроизводственных расходов влияние фондоотдачи на рентабельность снижается с 0,49 до 0,46 (значение парного коэффициента корреляции);

- аналогично рассмотрите частный коэффициент корреляции $r_{x2y|x1} = 0,42$ (по модулю) между рентабельностью (y) и непроизводственными расходами (x2) без учета фондоотдачи (x1) и сделайте соответствующий вывод.

Практические задания к лабораторной работе 4

Для заданных выборок из своего варианта проведите корреляционный анализ и оформите выполнение задания в отдельном отчете системы Statistica. Варианты заданий в приложении.

Лабораторная работа 5

Однофакторный дисперсионный анализ

Цель: освоить использование метода однофакторного дисперсионного анализа для проверки гипотезы о неравенстве средних между исследуемыми признаками с помощью средств программы Statistica.

Задачей *дисперсионного анализа* является изучение влияния одного или нескольких факторов на рассматриваемый признак. Однофакторный дисперсионный анализ используется в тех случаях, когда есть в распоряжении три или более независимые выборки, полученные из одной генеральной совокупности путем изменения какого-либо одного независимого фактора, для которого по каким-либо причинам нет количественных измерений.

Пример. Удобрения для комнатных растений фасуются в пакеты с одинаковым весом. Из партии пакетов расфасованных в течение суток, случайным образом отобрали 30 пакетов. Они были распределены по трем различным условиям хранения. После хранения в течение одной недели определялось содержание влаги в продукте, хранящемся в каждом пакете. Данные о содержании влаги:

Условия хранения	Содержание влаги, %
1	10,1 7,3 5,6 6,2 8,4 8,1 8,0 7,6 5,3 7,2
2	11,7 12,2 11,8 7,8 8,9 9,9 12,4 11,0 10,3 13,8 10,5 9,8 9,1
3	10,2 12,0 8,8 8,7 10,5 11,0 9,1

На уровне значимости $\alpha = 0,05$ проверить гипотезу о том, что условия хранения продукта не оказывают влияния на содержание влаги.

Ход работы

Основными условиями применения однофакторного дисперсионного анализа являются: а) каждая выборка распределена по нормальному закону; б) групповые дисперсии однородны (т.е. между ними нет статистически значимой разницы).

Проверка условий применения однофакторного дисперсионного анализа

1. Создайте новый файл данных в программе Statistica (2v*30с) и сохраните полученную таблицу в отчет: Отчет_лаб5.str (рис. 2.5.).
2. Откройте диалоговое окно для выполнения однофакторного дисперсионного анализа: Statistics (Анализ) – Basic Statistics/Tables (Основные

статистики и таблицы) – Breakdown & one-way Anova (Однофакторный дисперсионный анализ). В открывшемся диалоговом окне Statistics by Groups (Внутригрупповые статистики и корреляции) установите переменные для анализа (Усл_хр – группирующая (grouping), Содерж_вл – зависимая (dependent)) и нажмите Ок (рис. 2.5.).

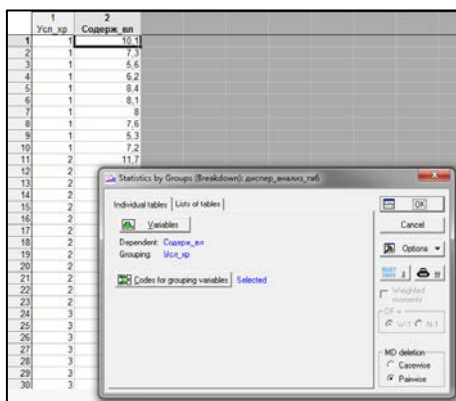


Рис. 2.5. Таблица данных и диалоговое окно однофакторного дисперсионного анализа

3. В открывшемся окне результатов Statistics by Groups – Results (Внутригрупповые статистики и корреляции – Результаты) нажмите кнопку Categorized normal prob. Plots (Категоризованные нормальные вероятностные графики). Рассмотрите полученные графики по группам: точки, соответствующие нормально распределенным данным, укладываются на прямые или плотно группируются относительно прямых. Сохраните полученные графики в Отчет_лаб5.str и сделайте вывод подчиняются ли выборки (по группам) нормальному распределению.

4. Для проверки условия о равенстве дисперсий – там же в окне результатов Statistics by Groups – Results (Внутригрупповые статистики и корреляции – Результаты) можно использовать критерий Левена (Levene test...) или критерий Брауна-Форсайта (Brown-Forsythe test...). Выполните один из критериев, сохраните полученные результаты в Отчет_лаб5.str и сделайте вывод о равенстве дисперсий (сравните значение p - вероятности отклонения гипотезы о неравенстве дисперсий, с заданным уровнем значимости α).

5. В отчете Отчет_лаб5.str сделайте общий вывод о возможности применения однофакторного дисперсионного анализа.

Применение однофакторного дисперсионного анализа

6. Вернитесь в окно результатов Statistics by Groups – Results (Внутригрупповые статистики и корреляции – Результаты) и выполните дисперсионный анализ – нажмите кнопку Analysis of variance (Дисперсионный анализ). Полученную таблицу сохраните в отчет Отчет_лаб5.str и сделайте вывод о равенстве средних (сравните значение p - вероятность отклонения гипотезы о равенстве средних с выбранным уровнем значимости α).
7. В отчете Отчет_лаб5.str сделайте общий вывод о значимости влияния фактора (условий хранения) на признак (содержание влаги).

Метод линейных контрастов

Если в результате однофакторного анализа гипотеза о равенстве средних принимается (фактор имеет значимое влияние на группы признаков), то требуется определить какие именно группы между собой имеют значимые различия. Для этого используется **метод линейных контрастов**.

8. Вернитесь в окно результатов Statistics by Groups – Results (Внутригрупповые статистики и корреляции – Результаты) и выполните сравнение средних по методу Шеффе (Scheffe test...). Рассмотрите полученную таблицу результатов – здесь указаны вероятности отклонения гипотез о равенстве средних для всех групп фактора (Условия хранения). Например, вероятность отклонения гипотезы о равенстве средних 2 и 3 группы велика ($p=0.642827$), следовательно, гипотеза о различии средних отклоняется и отличие среднего значения ($M = 10,708$) 2 группы от среднего значения 3 группы ($M = 10,043$) не значимо (рис. 2.6.).

Scheffe Test; Variable: Содерж_вл (диспер_анализ_таб)				
Marked differences are significant at $p < .05000$				
Усл_хр	{1}	{2}	{3}	
1	M=7,3800	M=10,708	M=10,043	
2		0,000068	0,004890	
3		0,004890	0,642827	

Рис. 2.6. Таблица результатов сравнения средних по методу Шеффе

9. Сохраните полученную таблицу результатов в отчет и сделайте вывод средние каких групп имеют значимые отличия.
10. В отчете Отчет_лаб5.str сделайте общий вывод о том, между какими группами существуют различия в зависимости от условий хранения.

Практические задания к лабораторной работе 5

Проведите однофакторный дисперсионный анализ для задачи своего варианта. Если результаты анализа говорят о значимом влиянии фактора (гипо-

теза о неравенстве средних принимается) проведите попарное сравнение средних, используя метод линейных контрастов, иначе найдите несмещенные оценки средних в группах и дисперсии ошибок наблюдений.

Варианты заданий

1. На химическом заводе разработаны два варианта технологического процесса. Дневная производительность завода приводится в таблице. Можно ли считать, что производительность завода изменилась при переходе на новые варианты технологического процесса. Принять $\alpha=0,05$.

День работы	Существующая схема	Вариант 1	Вариант 2
1	46	74	52
2	48	82	63
3	73	64	64
4	52	72	48
5	72	84	70
6	44	68	78
7	66	76	68
8	46	88	70
9	60	70	54
10	48	60	75

2. С помощью методов однофакторного дисперсионного анализа исследуйте данные о товарообороте трех магазинов в течение 6 месяцев (в млн. руб.). Принять $\alpha = 0,1$. Сделайте выводы по результатам дисперсионного анализа.

1 мага- зин	2 мага- зин	3 мага- зин
4	6	8
2	5	9
3	4	10
4	7	7
5	6	8
3	8	6

3. В трех магазинах, продающих товары одного вида, данные товарооборота за 8 месяцев работы (в тыс. руб.) составили следующую сводку. Требуется проверить данные методами однофакторного дисперсионного анализа. Принять $\alpha = 0,1$.

Магазин	Месяц							
	1	2	3	4	5	6	7	8
1	19	23	26	18	20	20	18	35
2	20	20	32	27	40	24	22	18
3	16	15	18	26	19	17	19	18

4. Проверить влияет ли на стоимость однокомнатных квартир (в тыс. у.е.) район города. Принять $\alpha = 0,05$.

Коньково	39, 30, 36, 33, 40, 30, 40, 43, 30, 40, 43, 29, 37, 30, 32, 41, 43, 31
Юго-Запад	43, 37, 43, 33, 42, 34, 42, 33, 37
Проспект Вернадского	33, 28, 40, 40, 32, 28

5. Исследуется влияние трех различных видов рекламы на объемы продаж (тыс. руб.) магазинов некоторого филиала. Требуется проверить гипотезу о влиянии фактора (рекламы) с $\alpha = 0,05$.

Вид рекламы	Номер магазина						
	1	2	3	4	5	6	7
1	3000	2230	2657	3245	810	788	2993
2	895	756	2060	2892	805	500	1234
3	651	619	2138	1322	775	520	910

6. Методами однофакторного дисперсионного анализа проверьте зависит ли производительность труда на предприятии от работающей бригады. Принять $\alpha = 0,05$.

№ бригады	Производительность
Бригада 1	170, 205, 158, 170, 166, 155, 195, 168, 179, 166, 169, 178, 173
Бригада 2	169, 171, 180, 165, 180, 173, 190, 170, 181, 168, 180, 159, 156, 150, 161

7. Методами однофакторного дисперсионного анализа проверьте зависит ли величина товарооборота однотипных торговых организаций от микрорайона города в которых они расположены. Принять $\alpha = 0,05$.

Микр-н	Товарооборот
1 микр-н	170, 153, 145, 155, 168, 174, 180, 149, 130, 139, 180, 157
2 микр-н	144, 177, 135, 164, 150, 179, 158, 178, 133, 172, 153, 145

8. Имеются данные о проверке учеников вторых классов двух школ города, где применялись разные методики преподавания, на скорость прочтения контрольного текста (слов в мин.). Проверьте одинаковые ли результаты дают различные методики преподавания.

Школа	Товарооборот
Школа №1	51, 62, 60, 64, 47, 63, 50, 53, 64, 70, 55, 68, 67, 67, 67, 65, 69
Школа №2	54, 56, 58, 60, 67, 65, 66, 62, 52, 59, 53, 63, 54, 49, 54, 65, 52

9. На опытном поле были выбраны 27 одинаковых по площади участков. Весной в 7 из них внесли «старое» удобрение, в 7 – новое удобрение 1, а в оставшиеся 7 – новое удобрение 2. В конце года была определена урожайность культуры по каждому участку. Влияет ли удобрение на среднюю урожайность культуры.

Старое удобрение	Новое удобрение 1	Новое удобрение 2
1920	2250	2252
2020	2410	2400
2060	2260	2261
1960	2200	2205
1960	2360	2368
2140	2320	2321
1980	2240	2249
1940	2300	2296
1790	2090	2100

10. Имеются данные статистического наблюдения, характеризующие величину ежедневных надоев молока на одной из ферм, применившей в ра-

ционе коров новую пищевую добавку, предназначенную для повышения их продуктивности. Оказывает ли влияние новая пищевая добавка на суточный удой коров.

Общая суточная величина надоев молока без применения пищевой добавки	Общая суточная величина надоев молока с применением пищевой добавки
380	374
365	389
388	376
382	387
377	378
375	385
370	372
378	390
385	387
380	392

РАЗДЕЛ 3. РЕГРЕССИОННЫЙ АНАЛИЗ В ПРОГРАММЕ STATISTICA

Лабораторная работа 6

Тема: Простая линейная регрессия

Цель: освоить приемы использования модуля множественной регрессии программы Statistica для построения и анализа модели простой линейной регрессии.

Регрессионный анализ – состоит в определении связи между некоторой характеристикой Y наблюдаемого явления и величинами $X_1, X_2, \dots, X_n$, которые обуславливают (объясняют) изменения Y . Переменная Y называется зависимой переменной, объясняющие переменные $X_1, X_2, \dots, X_n$ – независимыми переменными.

Простая линейная регрессия – это простейшая регрессионная модель, описывающая линейную зависимость переменной Y от одного фактора X :

$$Y = \beta_0 + \beta_1 * X$$

Пример: Имеются данные о курсах покупки и продажи акций Иркутск-энерго за период – один месяц. Определить, как цена покупок связана с ценой продаж.

	1	2
	Покупка	Продажа
1	0,096	0,1103
2	0,098	0,111
3	0,0986	0,1117
4	0,0992	0,1121
5	0,0988	0,1115
6	0,0992	0,1114
7	0,0979	0,1121
8	0,0981	0,1149
9	0,0972	0,1118
10	0,0974	0,1118
11	0,0981	0,1119
12	0,099	0,1127
13	0,0986	0,1127
14	0,0987	0,1129
15	0,0996	0,1137
16	0,1	0,1145
17	0,1003	0,1156
18	0,1016	0,1163
19	0,1018	0,1161
20	0,1039	0,1176
21	0,1068	0,1236

Ход работы:

1. Создайте новый файл данных в программе Statistica (2v*21c) и сохраните полученную таблицу в отчет: Отчет_лаб6.str.

Исследование существования зависимости между X и Y

2. Постройте диаграмму рассеяния командой Graphs (График) – Scatter-plots (Диаграмма рассеяния) (при построении установите Fit (Подгонка прямой): Off (Выкл)), сохраните ее в Отчет_лаб6.str и сделайте вывод о возможности линейной зависимости между X и Y (рис. 3.1.).

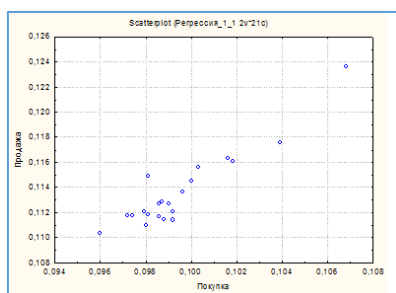


Рис. 3.1. Диаграмма рассеяния

3. Найдите коэффициент корреляции между признаками X и Y, сохраните данные в Отчет_лаб6.str, оцените тесноту существующей связи ($-1 \leq r \leq 1$). И сделайте вывод о возможности линейной зависимости между X и Y: в случае, если между величинами существует достаточно тесная статистическая зависимость, то ее можно аппроксимировать уравнением линейной регрессии Y на X: $Y = \beta_0 + \beta_1 * X$

Применение метода регрессионного анализа

4. Постройте линейную модель зависимости с помощью инструментов регрессионного анализа: Statistics (Анализ) -> Multiple regression (Множественная регрессия). В диалоговом окне Multiple regression (Множественная регрессия) выберите переменные для анализа: зависимая – Продажа, независимая – Покупка и нажмите Ок.
5. На экране появится окно результатов множественной регрессии – нажмите кнопку Summary: Regression results (Результаты регрессии) – получите итоговую таблицу результатов регрессии (рис. 3.2, 3.3.).

Multiple Regression Results					
Dependent: Продажа	Multiple R =	.93400145	F =	129,8546	
	R ² =	.87235870	df =	1,19	
No. of cases: 21	adjusted R ² =	.86564074	p =	.000000	
	Standard error of estimate:	.001107331			
Intercept:	-.002363958	Std.Error:	.0101818	t(19) =	-.2322 p = .8189
Покупка beta=.934					

Рис. 3.2. Окно результатов регрессии

Regression Summary for Dependent Variable: Продажа (Perpe						
R= .93400145 R ² = .87235870 Adjusted R ² = .86564074						
F(1,19)=129,85 p<.000000 Std.Error of estimate: .00111						
N=21	Beta	Std.Err. of Beta	B	Std.Err. of B	t(19)	p-level
Intercept			-0,002364	0,010182	-0,23218	0,818884
Покупка	0,934001	0,081963	1,166145	0,102335	11,39538	0,000000

Рис. 3.3. Итоговая таблица результатов регрессии

6. Сохраните полученные результаты в Отчет_лаб6.str и выделите наиболее важные данные: коэффициенты уравнения линейной регрессии: - 0.002364 (β_0) и 1.1166145 (β_1), множественный коэффициент корреляции: $r = 0.93400145$, коэффициент детерминации: $r^2 = 0.8723587$, значения t-критерия: $t(19) = -0.2322$, $p = 0.8189$ и $t(19) = 11.395$, $p = 0.0000$, значения F-критерия: $F = 129.8546$ и уровень $p = 0,000000$.
7. Сформулируйте в отчете выводы по выделенным данным:
 - запишите полученное уравнение линейной регрессии;
 - результаты t-критерия Стьюдента используйте для проверки значимости коэффициентов β_0 и β_1 (p – вероятность отклонения гипотезы о значимом отличии коэффициентов от нуля сравните с заданным уровнем значимости α (0,05);
 - значения F-критерия Фишера используйте для проверки значимости построенной модели линейной регрессии (p – вероятность отклонения гипотезы о значимости модели уравнения сравните с заданным уровнем значимости α);

- по значению коэффициента детерминации r^2 оцените качество построенной модели: если $r^2 \geq 0.5 \Rightarrow$ модель оценивается как качественная и показывает на сколько процентов построенная модель регрессии объясняет разброса значений переменной Y относительно выборочного среднего.

Проверка адекватности модели

Линейная регрессионная модель называется **адекватной**, если предсказанные по ней значения переменной Y согласуются с результатами наблюдений. Анализ адекватности основывается на анализе остатков (остатки представляют собой разности между наблюдаемыми значениями и модельными, т.е. значениями, полученными по модели). Грубая оценка адекватности модели может быть проведена по графику остатков.

8. В окне Multiple regression results (Результаты множественной регрессии) нажмите кнопку Residual Analysis (Анализ остатков). В окне Residual Analysis (Анализ остатков) выполните построение двух графиков: График 1 – отображающий зависимость наблюдаемых значений и остатков (Observed vs. Residuals) и График 2 – зависимость предсказанных значений и остатков (Predicted vs. Residuals).

9. Сохраните полученные графики в Отчет_лаб6.str и сделайте выводы по графикам – если остатки (точки на плоскости) хаотично разбросаны относительно прямой регрессии остатков, в их поведении нет закономерностей $\Rightarrow$ остатки не коррелированы между собой и также нет резко выделяющихся остатков. Отсюда можно заключить, что модель достаточно адекватно описывает данные.

Прогнозирование

Если регрессионная модель значима и адекватна, то она может быть использована для определения прогноза $Y(X_0)$ при заданном значении независимой переменной $X=X_0$. Расчет доверительных интервалов позволяет определить область, внутри которой следует ожидать варьирование значения прогнозируемой величины.

10. Вычислите прогнозное значение Y в точке $X_0 = (X_{\max} + X_{\min})/2$. В окне Multiple regression results (Результаты множественной регрессии) нажмите кнопку Predict dependent variable (Предсказать зависимую переменную). Предварительно необходимо задать уровень значимости α и вид вычисляемого доверительного интервала: Доверительные границы для средних.

11. Сохраните полученные значения прогноза и доверительные интервалы (95%-ные) в Отчет_лаб6.str.

Практические задания к лабораторной работе 6

Постройте модель простой линейной регрессии для задачи своего варианта. Если модель значима и адекватно описывает данные выполните прогноз. Варианты заданий в приложении.

Лабораторная работа 7

Множественная линейная регрессия

Цель: освоить приемы использования модуля множественной регрессии программы Statistica для построения и анализа модели простой линейной регрессии.

Множественная линейная регрессия – это регрессионная модель, описывающая линейную зависимость переменной Y от целого ряда независимых факторов $X_1, X_2, \dots, X_n$:

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \dots + \beta_n * X_n$$

Пример: Руководство авиакомпании по результатам анализа деятельности 15 своих представительств получило следующие данные за март месяц, представленные в таблице. Y (зависимая переменная) – общий доход от проданных билетов, млн. руб.; X_1 – средства на развитие компаний в регионе, млн. руб.; X_2 – число конкурирующих компаний; X_3 – процент пассажиров, летавших бесплатно. Определите от чего зависит общий доход от проданных билетов.

	1 Y	2 X1	3 X2	4 X3
1	79.3	2.5	10	3
2	200.1	5.5	8	6
3	163.2	6	12	9
4	200.1	7.9	7	16
5	146	5.2	8	15
6	177.7	7.6	12	9
7	30.9	2	12	8
8	291.9	9	5	10
9	160	4	8	4
10	339.4	9.6	5	16
11	159.6	5.5	11	7
12	88.3	3	12	8
13	237.5	6	6	10
14	107.2	5	10	4
15	155	3.5	10	4

Ход работы

1. Создайте новый файл данных в программе Statistica (4v*15c) и сохраните полученную таблицу в отчет: Отчет_лаб7.str.

Применение метода регрессионного анализа

2. Постройте линейную модель зависимости с помощью инструментов регрессионного анализа: Statistics (Анализ) -> Multiple regression (Множественная регрессия). В диалоговом окне Multiple regression (Множественная регрессия) выберите переменные для анализа: зависимая – Y , независимые – X_1, X_2, X_3 и нажмите Ок. На экране появится окно результатов множественной регрессии – нажмите кнопку Summary: Regression results (Результаты регрессии) – получите итоговую таблицу результатов регрессии.

3. Сохраните таблицу результатов регрессии в Отчет_лаб7.str и выделите желтым цветом наиболее важные данные: коэффициенты уравнения линейной регрессии ($\beta_0, \beta_1, \beta_2, \beta_3$) и значения вероятности отклонения гипотезы о значимости коэффициентов (p) полученные по t-критерию для каждого коэффициента.

4. Выполните оценку значимости построенной модели с помощью дисперсионного анализа: в диалоговом окне Multiple regression (Множественная регрессия) – ANOVA (Дисперсионный анализ). Сохраните полученную таблицу дисперсионного анализа в Отчет_лаб7.str и выделите желтым цветом статистику F-критерия (F и p) (рис. 3.4.).

Analysis of Variance; DV: Y (Регрессия_1_2)					
Effect	Sums of Squares	df	Mean Squares	F	p-level
Regress.	80720,39	3	26906,80	34,82106	0,000007
Residual	8499,88	11	772,72		
Total	89220,26				

Рис. 3.4. Таблица результатов дисперсионного анализа

5. Сформулируйте в отчете Отчет_лаб7.str выводы по выделенным данным:

- запишите полученное уравнение линейной регрессии;
- по таблице результатов регрессии: оцените какие коэффициенты в уравнении линейной регрессии значимы ($\beta_0, \beta_1, \beta_2$), а какие не значимы (β_3);
- по таблице результатов дисперсионного анализа: оцените значимость модели.

Отбор наиболее существенных факторов для модели

6. В полученном уравнении линейной регрессии не значим коэффициент β_3 , следовательно, фактор X_3 не имеет важного значения для предсказания Y , следовательно, из модели его можно исключить. Выполните множественный регрессионный анализ для зависимой переменной – Y , и независимых – X_1, X_2 и сформулируйте в Отчет_лаб7.str следующие выводы:

- пересчитайте уравнение множественной регрессии, используя два фактора X_1 и X_2 ;
- проверьте значимость коэффициентов ($\beta_0, \beta_1, \beta_2$) (по таблице регрессионного анализа);
- оцените значимость построенной модели (по таблице дисперсионного анализа);
- оцените качество построенной модели (по коэффициенту детерминации r^2);

- охарактеризуйте влияние факторов X_1 (средства на развитие компаний в регионе, млн. руб) и X_2 (число конкурирующих компаний) на зависимую переменную Y (общий доход от проданных билетов) (что происходит с Y при увеличении или уменьшении факторов X_1 и X_2).

Проверка адекватности модели

Анализ адекватности модели множественной регрессии также основывается на анализе остатков (остатки представляют собой разности между наблюдаемыми значениями и модельными, т.е. значениями, полученными по модели), следующим образом:

- дисперсия остатков постоянна;
- остатки не коррелированы;
- остатки имеют нормальное распределение ($N()$).

7. Проверка условия постоянства дисперсии остатков проводится по графику остатков: если все остатки укладываются в симметричную относительно нулевой линии полосу, то можно считать, что дисперсия ошибок постоянна. Multiple regression results (Результаты множественной регрессии) – Residual Analysis (Анализ остатков) – Plot of residuals (График остатков) (рис. 3.5.)

										Raw Residual (Перспекция_1_2)								
Raw Residuals										Dependent variable: Y								
Case	-3s			0					+3s	Observed Value	Predicted Value	Residual	Standard Residual	Standard Pred. v	Std Err Pred Val	Mahalanobis Distance	Deleted Residual	Cook's Distance
1	.	.	.	*	.	.	.	.	.	79.3000	90.5178	-11.2178	-1.04146	-0.39753	12.91724	2.000232	-14.1915	0.017666
2	.	.	.	.	*	.	.	.	.	200.1000	182.7455	17.3545	0.18116	0.61500	8.30152	0.278301	18.9988	0.013077
3	.	.	.	.	*	.	.	.	.	163.2000	143.8134	19.3866	-0.33495	0.68702	14.44502	2.735197	26.2705	0.075702
4	.	.	*	.	.	.	.	.	.	200.1000	249.0087	-48.9087	1.05958	-1.73321	10.87774	1.147002	-57.4447	0.005264
5	.	.	.	*	.	.	.	.	.	146.0000	176.0291	-30.0291	0.09212	-1.06416	8.74939	0.412563	-33.2230	0.044419
6	.	.	.	.	*	.	.	.	.	177.7000	179.6345	-1.9345	0.13992	-0.06855	19.31221	5.623892	-3.6388	0.002596
7	.	.	.	*	.	.	.	.	.	30.9000	54.2606	-23.3606	-1.52211	-0.82784	13.72636	2.379251	-30.6013	0.092752
8	.	.	.	.	*	.	.	.	.	291.9000	298.6988	-6.7988	1.71830	-0.24093	15.00633	3.025845	-9.4797	0.010638
9	.	.	.	.	*	.	.	.	.	160.0000	149.1632	10.8368	-0.26403	0.38403	11.76768	1.501325	13.1181	0.012527
10	.	.	.	.	.	*	.	.	.	339.4000	312.1317	27.2682	1.89637	0.96632	16.05110	3.596329	40.3107	0.220082
11	.	.	.	.	*	.	.	.	.	159.6000	145.1508	14.4492	-0.31722	0.51204	10.31924	0.938863	16.6797	0.015574
12	.	.	.	.	*	.	.	.	.	88.3000	76.6488	11.6512	-1.22532	-1.4289	11.88165	1.548712	14.1620	0.014885
13	.	.	.	.	*	.	.	.	.	237.5000	219.0027	18.4973	0.66180	0.65550	12.65852	1.883894	23.1572	0.045172
14	.	.	*	.	.	.	.	.	.	107.2000	146.4863	-39.2863	-0.29949	-1.39228	7.79312	0.134440	-42.5322	0.057756
15	.	.	.	.	*	.	.	.	.	155.0000	112.9060	42.0940	-0.74467	-1.49171	9.91241	0.794153	48.0192	0.119103
Minimum	.	.	.	*	.	.	.	.	.	30.9000	54.2606	-48.9087	-1.52211	-1.73321	7.79312	0.134440	-57.4447	0.002596
Maximum	.	.	.	.	*	.	.	.	.	339.4000	312.1317	27.2682	1.89637	1.49171	19.31221	5.623892	48.0192	0.220082
Mean	.	.	.	.	*	.	.	.	.	169.0800	169.0800	-0.0000	0.00000	-0.00000	12.24797	1.866667	0.6403	0.063148

Рис. 3.5. График остатков

8. Проверка гипотезы (H_0): все корреляции между остатками равны нулю ($r_k = 0$), происходит с помощью критерия Дарбина-Уотсона: H_0 принимается на уровне значимости 2α , если $d > d_2$ или $4-d > d_2$, H_0 отклоняется на уровне значимости 2α , если $d < d_1$ или $4-d < d_1$, где d_1 и d_2 – критические точки определяются по таблице Дарбина-Уотсона. Вычисление статистики Дарбина-Уотсона: Multiple regression results (Результаты множественной регрессии) – Residual Analysis (Анализ остатков) – Durbin-Watson stat (рис. 3.6.).

Durbin-Watson d (Перегрессия_1_2) and serial correlation of residuals		
Durbin-Watson d	Serial	
Estimate	Corr.	
	1.896936	-0.058644

Рис. 3.6. Результаты статистики Дарбина-Уотсона

9. Проверка гипотезы о нормальном распределении происходит с помощью построения нормального вероятностного графика: если точки расположены близко к прямой, то остатки имеют нормальное распределение: Multiple regression results (Результаты множественной регрессии) – Residual Analysis (Анализ остатков) – Normal Probability Plot of Residuals:

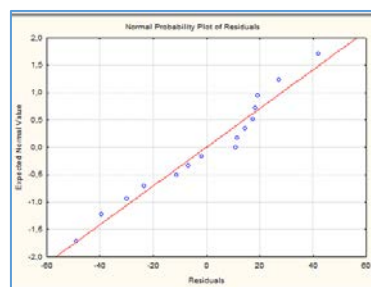


Рис. 3.7. Нормальный вероятностный график остатков

10. Сохраните полученные результаты в Отчет_лаб7.str и сделайте соответствующие выводы об адекватности модели.

Прогнозирование

11. Определите какой доход может иметь компания, вложившая в развитие 2,5 млн. руб., если число конкурирующих компаний в регионе будет равно 10.

Практические задания к лабораторной работе 7

Постройте модель множественной линейной регрессии для задачи своего варианта. Если модель значима и адекватно описывает данные выполните прогноз. Варианты заданий в приложении.

Лабораторная работа 8

Пошаговая регрессия

Цель: закрепить использование инструментов модуля множественной регрессии в программе Statistica при работе с методом пошаговой регрессии.

Пошаговая регрессия – позволяет выполнить отбор наиболее существенных факторов, достаточно точно описывающих зависимую переменную Y, т.е. решает задачу по выбору «наилучшей» регрессии.

Пример: Выполнить пошаговую регрессию для набора следующих данных.

	1 X1	2 X2	3 X3	4 X4	5 Y
1	7	26	6	60	78.5
2	1	29	15	52	74.3
3	11	56	8	20	104.3
4	11	31	8	47	87.6
5	7	52	6	33	95.9
6	11	55	9	22	109.2
7	3	71	17	6	102.7
8	1	31	22	44	72.5
9	2	54	18	22	93.1
10	21	47	4	26	115.9
11	1	40	23	34	83.8
12	11	66	9	12	113.3
13	10	68	8	12	109.4

Ход работы:

1. Создайте новый файл данных в программе Statistica (5v*13с) и сохраните полученную таблицу в отчет: Отчет_лаб8.str.
2. Для исследования связей между выборками выполните корреляционный анализ: Statistics (Анализ) - Basic Statistics/Tables (Основные статистики и таблицы) -> Correlation matrices (Корреляционная матрица) (рис. 3.8.).

Marked correlations are significant at p < .05000 N=13 (Casewise deletion of missing data)					
Variable	X1	X2	X3	X4	Y
X1	1.00	0.23	-0.82	-0.25	0.73
X2	0.23	1.00	-0.14	-0.97	0.82
X3	-0.82	-0.14	1.00	0.03	-0.53
X4	-0.25	-0.97	0.03	1.00	-0.82
Y	0.73	0.82	-0.53	-0.82	1.00

Рис. 3.8. Корреляционная матрица

3. Сохраните полученную матрицу в Отчет_лаб8.str и сделайте выводы по полученным коэффициентам корреляции:
 - оцените связь независимых переменных X_1 , X_2 , X_3 , X_4 с зависимой переменной Y ;
 - оцените наличие мультиколлинеарности: существование значимых связей между независимыми переменными X_1 , X_2 , X_3 , X_4 .

Пошаговая регрессия с исключением

В методе пошаговой регрессии с исключением из модели пошагово исключаются переменные, которые вносят наименьший вклад в регрессию на данном шаге.

4. Запустите модуль регрессионного анализа: Statistics (Анализ) -> Multiple regression (Множественная регрессия). В диалоговом окне Multiple regression (Множественная регрессия) выберите переменные для анализа: зависимая – Y , независимые – X_1 , X_2 , X_3 , X_4 , установите галку Advanced options (stepwise or ridge regression) (Пошаговая или гребневая регрессия) и нажмите Ок.

5. В открывшемся диалоговом окне *Model definition* (Определение модели) на вкладке *Stepwise* (Пошаговый) установите следующие параметры: *Method* (Процедура): *Backward stepwise* (Пошаговая с исключением), *F to enter* (F-статистика включения): 4, *F to remove* (F-статистика удаления): 3,9 и *Display result* (Отображение результатов): *At each step* (На каждом шаге) и нажмите *Ок*.

6. В открывшемся диалоговом окне *Multiple regression result* (Результаты множественной регрессии) добавляется возможность выполнять пошаговую регрессию с помощью кнопки *Next* (Следующий). Выполняя пошаговую регрессию, для каждого шага сохраняйте таблицу результатов регрессии (*Summary regression results*) в *Отчет_лаб8.str*.

7. В отчете *Отчет_лаб8.str* опишите полученную модель:

- запишите уравнение множественной линейной регрессии в виде:

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \dots + \beta_n * X_n;$$

- по таблице результатов регрессии: оцените значимость коэффициентов в уравнении;

- по таблице результатов дисперсионного анализа: оцените значимость модели;

- по коэффициенту детерминации: оцените качество модели.

Пошаговая регрессия с включением

В методе пошаговой регрессии с включением в модель пошагово включаются переменные, которые вносят наибольший вклад в регрессию на данном шаге.

8. Запустите модуль регрессионного анализа: *Statistics* (Анализ) -> *Multiple regression* (Множественная регрессия). В диалоговом окне *Multiple regression* (Множественная регрессия) выберите переменные для анализа: зависимая – *Y*, независимые – *X₁*, *X₂*, *X₃*, *X₄*, установите галку *Advanced options (stepwise or ridge regression)* (Пошаговая или гребневая регрессия) и нажмите *Ок*.

9. В открывшемся диалоговом окне *Model definition* (Определение модели) на вкладке *Stepwise* (Пошаговый) установите следующие параметры: *Method* (Процедура): *Forward stepwise* (Пошаговая с включением), *F to enter* (F-статистика включения): 4, *F to remove* (F-статистика удаления): 3,9 и *Display result* (Отображение результатов): *At each step* (На каждом шаге) и нажмите *Ок*.

10. В открывшемся диалоговом окне *Multiple regression result* (Результаты множественной регрессии) добавляется возможность выполнять поша-

говую регрессию с помощью кнопки *Next* (Следующий). Выполняя пошаговую регрессию, для каждого шага сохраняйте таблицу результатов регрессии (*Summary regression results*) в Отчет_лаб8.str.

11. В отчете Отчет_лаб8.str опишите полученную модель:

- запишите уравнение множественной линейной регрессии в виде:

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \dots + \beta_n * X_n;$$

- по таблице результатов регрессии: оцените значимость коэффициентов в уравнении;

- по таблице результатов дисперсионного анализа: оцените значимость модели;

- по коэффициенту детерминации: оцените качество модели.

12. Сформулируйте общие выводы в отчете Отчет_лаб8.str:

- сравните результаты обеих процедур пошаговой регрессии;

- сравните результаты пошаговой регрессии с результатами корреляционного анализа.

Практические задания к лабораторной работе 8

Используя методы пошаговой регрессии с исключением и включением постройте модель множественной линейной регрессии для задачи своего варианта. Варианты заданий в приложении.

СПИСОК ЛИТЕРАТУРЫ

1. Вуколов Э.А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов Statistica и Excel: учебное пособие. – 2-е изд., испр. и доп. – М.: ФОРУМ, 2008. – 464 с.
2. Практикум по курсу «Статистика» (в системе Statistica). Салин В.Н., Чурилова Э.Ю. М. «Издательский дом «Социальные отношения», Издательство «Перспектива». 2012. - 188 с.
3. Стукач О.В. Программный комплекс Statistica в решении задач управления качеством: учебное пособие. Томск: Изд-во Томского политехнического университета, 2011. – 163 с.

Приложение

Варианты заданий по корреляционному и регрессионному анализу

Рассматриваются следующие показатели для 50 предприятий:

Y_1 – производительность труда;

Y_2 – индекс снижения себестоимости продукции;

Y_3 – рентабельность;

X_4 – трудоемкость единицы продукции;

X_5 – удельный вес рабочих;

X_6 – удельный вес покупных изделий;

X_7 – коэффициент сменности оборудования;

X_8 – премии и вознаграждения на одного работника;

X_9 – удельный вес потерь от брака;

X_{10} – фондоотдача;

X_{11} – среднегодовая численность работников;

X_{11} – среднегодовая стоимость основных производственных фондов;

X_{11} – среднегодовой фонд заработной платы;

X_{11} – фондовооруженность труда;

X_{11} – непроизводственные расходы.

Варианты заданий

Номер варианта	Результативный признак, Y_i	Факторные признаки, X_i
1	1	6, 8, 11, 12, 15
2	1	8, 11, 12, 13, 15
3	1	8, 9, 13, 14, 15
4	3	8, 9, 10, 11, 15
5	3	8, 9, 10, 12, 15
6	2	4, 5, 6, 8, 9
7	2	4, 5, 6, 7, 9
8	2	4, 5, 6, 8, 9
9	2	4, 5, 8, 9, 15
10	2	4, 5, 7, 9, 15

Таблица исходных данных

№ предприятия	Y ₁	Y ₂	Y ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀	X ₁₁	X ₁₂	X ₁₃	X ₁₄	X ₁₅
1	9,26	204,2	13,26	0,23	0,78	0,40	1,37	1,23	0,23	1,45	26006	167,69	47750	6,40	17,72
2	9,38	209,6	10,16	0,24	0,75	0,26	1,49	1,04	0,39	1,30	23935	186,10	50391	7,80	18,39
3	12,11	222,6	13,72	0,19	0,68	0,40	1,44	1,80	0,43	1,37	22589	220,45	43149	9,76	26,46
4	10,81	236,7	12,85	0,17	0,70	0,50	1,42	0,43	0,18	1,65	21220	169,30	41089	7,90	22,37
5	9,35	62,0	10,63	0,23	0,62	0,40	1,35	0,88	0,15	1,91	7394	39,53	14257	5,35	28,13
6	9,87	53,1	9,12	0,43	0,76	0,19	1,39	0,57	0,34	1,68	11586	40,41	22661	9,90	17,55
7	8,17	172,1	25,83	0,31	0,73	0,25	1,16	1,72	0,38	1,94	26609	102,96	52509	4,50	21,92
8	9,12	56,5	23,39	0,26	0,71	0,44	1,27	1,70	0,09	1,89	7801	37,02	14903	4,88	19,52
9	5,88	52,6	14,68	0,49	0,69	0,17	1,16	0,84	0,14	1,94	11587	45,74	25587	3,46	23,99
10	6,3	46,6	10,05	0,36	0,73	0,39	1,25	0,60	0,21	2,06	9475	40,07	16821	3,60	21,76
11	6,22	53,2	13,99	0,37	0,68	0,33	1,13	0,82	0,42	1,96	10811	45,44	19459	3,56	25,68
12	5,49	30,1	9,68	0,43	0,74	0,25	1,10	0,84	0,05	1,02	6371	41,08	12973	3,65	18,13
13	6,50	146,4	10,03	0,35	0,66	0,32	1,15	0,67	0,29	1,85	26761	136,14	50907	4,28	25,74
14	6,61	18,1	9,13	0,38	0,72	0,02	1,23	1,04	0,48	0,88	4210	42,39	6920	8,85	21,21
15	4,32	13,6	5,37	0,42	0,68	0,06	1,39	0,66	0,41	0,62	3557	37,39	5736	8,52	22,97
16	7,37	89,8	9,86	0,30	0,77	0,15	1,38	0,86	0,62	1,09	14148	101,78	26705	7,19	16,38
17	7,02	62,5	12,62	0,32	0,78	0,08	1,35	0,79	0,56	1,6	9872	47,55	20068	4,82	13,21
18	8,25	46,3	5,02	0,25	0,78	0,20	1,42	0,34	1,76	1,53	5975	32,61	11487	5,46	14,48
19	8,15	103,5	21,18	0,31	0,81	0,20	1,42	0,34	1,76	1,53	16662	103,25	32029	6,20	13,38
20	8,72	73,3	25,17	0,26	0,79	0,30	1,41	1,46	0,45	2,22	9166	38,95	18946	4,25	13,69
21	6,64	76,6	19,40	0,37	0,77	0,24	1,35	1,27	0,50	1,32	15118	81,32	28025	5,38	16,66
22	8,10	73,01	21,0	0,29	0,78	0,10	1,48	1,58	0,77	1,48	11429	67,26	20968	5,88	15,06
23	5,52	32,3	6,57	0,34	0,72	0,11	1,24	0,68	1,20	0,68	6462	59,92	11049	9,27	20,09
24	9,37	199,6	14,19	0,23	0,79	0,47	1,40	0,86	0,21	2,30	24628	107,34	45893	4,36	15,98
25	13,17	598,1	15,81	0,17	0,77	0,53	1,45	1,98	0,25	1,37	49727	512,60	99400	10,31	18,27
26	6,67	71,2	5,23	0,29	0,80	0,34	1,40	0,33	0,15	1,51	11470	53,81	20719	4,69	14,42
27	5,68	90,8	7,99	0,41	0,71	0,20	1,28	0,45	0,66	1,43	19448	80,83	36813	4,16	22,76
28	5,22	82,1	17,50	0,41	0,79	0,24	1,33	0,74	0,74	1,82	18963	59,42	33956	3,13	15,41

29	10,02	76,2	17,16	0,22	0,76	0,54	1,22	0,03	0,32	2,62	9185	36,96	17016	4,02	19,35
30	8,16	119,5	14,54	0,29	0,78	0,40	1,28	0,99	0,89	1,75	17478	91,43	34873	5,23	16,83
31	3,78	21,9	6,24	0,51	0,62	0,20	1,47	0,24	0,23	1,54	6256	17,16	11237	2,74	30,53
32	6,48	48,4	12,08	0,36	0,75	0,64	1,27	0,57	0,32	2,25	8810	27,29	17306	3,10	17,98
33	10,44	173,5	9,49	0,23	0,71	0,42	1,51	1,22	0,54	1,07	17659	184,33	39250	10,44	22,09
34	7,65	74,1	9,28	0,26	0,74	0,27	1,46	0,68	0,75	1,44	10324	58,42	19074	5,65	18,29
35	8,77	68,6	11,42	0,27	0,65	0,37	1,27	1,0	0,16	1,40	8901	59,40	18452	6,67	26,05
36	7,00	60,8	10,31	0,29	0,66	0,38	1,43	0,81	0,24	1,31	32625	391,27	7888	11,99	17,26
37	11,06	355,6	8,65	0,01	0,84	0,35	1,50	1,27	0,59	1,12	8402	49,63	17500	5,91	26,20
38	9,02	264,8	10,94	0,02	0,74	0,42	1,35	1,14	0,56	1,16	31160	258,62	58947	8,30	18,83
39	13,28	526,6	9,87	0,18	0,75	0,32	1,41	1,89	0,63	0,88	46461	75,66	94697	1,63	19,70
40	9,27	118,6	6,14	0,25	0,75	0,33	1,47	0,67	1,10	1,07	13833	123,68	29626	8,94	16,87
41	6,70	37,1	12,93	0,31	0,79	0,29	1,35	0,96	0,39	1,24	6391	37,21	11688	5,82	14,63
42	6,69	57,7	9,78	0,38	0,72	0,30	1,40	0,67	0,73	1,49	11115	53,37	21955	4,80	22,17
43	9,42	51,6	13,22	0,24	0,70	0,56	1,20	0,98	0,28	2,03	6555	32,87	12243	5,01	22,62
44	7,24	64,7	17,29	0,31	0,66	0,42	1,15	1,16	0,10	1,84	11085	45,63	20193	4,12	26,44
45	5,39	48,3	7,11	0,42	0,69	0,26	1,09	0,54	0,68	1,22	9484	48,41	20122	5,10	22,26
46	5,61	15,0	22,49	0,51	0,71	0,16	1,26	1,23	0,87	1,72	3967	13,58	7612	3,49	19,13
47	5,59	87,5	12,14	0,31	0,73	0,45	1,36	0,78	0,49	1,75	15283	63,99	27404	4,19	18,28
48	6,57	108,4	15,25	0,37	0,65	0,31	1,15	1,16	0,16	1,46	20874	104,55	39648	5,01	28,23
49	6,54	267,3	31,34	0,16	0,82	0,08	1,87	4,44	0,85	1,60	19418	222,11	43799	11,44	12,39
50	4,23	34,2	11,56	0,18	0,80	0,68	1,17	1,06	0,13	1,47	3351	25,76	6235	7,67	11,64

**МЕТОДИЧЕСКИЕ УКАЗАНИЯ
К ЛАБОРАТОРНЫМ РАБОТАМ
ПО ДИСЦИПЛИНЕ
«СТАТИСТИЧЕСКИЙ ПАКЕТ ПРОГРАММ STATISTICA»
ДЛЯ СТУДЕНТОВ НАПРАВЛЕНИЙ ОБУЧЕНИЯ
09.03.03 "ПРИКЛАДНАЯ ИНФОРМАТИКА",
ПРОФИЛЬ "ПРИКЛАДНАЯ ИНФОРМАТИКА В ЭКОНОМИКЕ"
01.03.02 "ПРИКЛАДНАЯ МАТЕМАТИКА И ИНФОРМАТИКА",
ПРОФИЛЬ – ОБЩИЙ
(очная форма обучения)**

**Составитель:
*В.В. Юданова***

Технический редактор *Л.В. Николаева*

Подписано в печать 22.03.2016. Формат 60х84/16.
Бумага тип. №2. Гарнитура «Таймс». Печать офсетная.
Печ. л. 2,6. Тираж 50 экз. Заказ 306.
Издательство ТИ (ф) СВФУ, 678960, г. Нерюнгри, ул. Кравченко, 16.

Отпечатано в ТИ (ф) ФГАОУ ВПО «СВФУ»
г. Нерюнгри