



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«Магнитогорский государственный технический университет им. Г.И. Носова»



УТВЕРЖДАЮ  
Директор ИЭиАС  
С.И. Лукьянов

26.02.2020 г.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)**

***СРЕДСТВА И МЕТОДЫ ВИЗУАЛИЗАЦИИ ИНФОРМАЦИИ***

Направление подготовки (специальность)

09.04.01 Информатика и вычислительная техника

Направленность (профиль/специализация) программы

Программное обеспечение средств вычислительной техники и автоматизированных систем

Уровень высшего образования - магистратура

Форма обучения  
заочная

Институт/ факультет	Институт энергетики и автоматизированных систем
Кафедра	Вычислительной техники и программирования
Курс	1

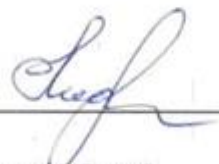
Магнитогорск  
2019 год

Рабочая программа составлена на основе ФГОС ВО по направлению подготовки 09.04.01 Информатика и вычислительная техника (уровень магистратуры) (приказ Минобрнауки России от 19.09.2017 г. № 918)

Рабочая программа рассмотрена и одобрена на заседании кафедры  
Вычислительной техники и программирования

19.02.2020 г. протокол № 5

Зав. кафедрой



О.С. Логунова

Рабочая программа одобрена методической комиссией ИЭиАС

26.02.2020 г. протокол № 5

Председатель



С.И. Лукьянов

Рабочая программа составлена:

доцент кафедры ВТиП, канд. техн. наук



Ю.Б. Кухта

зав. кафедрой ВТиП, д-р техн. наук



О.С. Логунова

Рецензент:

Начальник отдела инновационных разработок

ЗАО «КонсОМ-СКС», канд. техн. наук



А.Н. Панов

## Лист актуализации рабочей программы

---

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2020 - 2021 учебном году на заседании кафедры Вычислительной техники и программирования

Протокол от 07 октября 2020 г. № 2  
Зав. кафедрой \_\_\_\_\_ О.С. Логунова

---

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2021 - 2022 учебном году на заседании кафедры Вычислительной техники и программирования

Протокол от \_\_\_\_\_ 20\_\_ г. № \_\_\_\_  
Зав. кафедрой \_\_\_\_\_ О.С. Логунова

---

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2022 - 2023 учебном году на заседании кафедры Вычислительной техники и программирования

Протокол от \_\_\_\_\_ 20\_\_ г. № \_\_\_\_  
Зав. кафедрой \_\_\_\_\_ О.С. Логунова

---

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2023 - 2024 учебном году на заседании кафедры Вычислительной техники и программирования

Протокол от \_\_\_\_\_ 20\_\_ г. № \_\_\_\_  
Зав. кафедрой \_\_\_\_\_ О.С. Логунова

### **1 Цели освоения дисциплины (модуля)**

Целями освоения дисциплины (модуля) «Средства и методы визуализации информации» являются: формирование у выпускника комплекса компетенций, направленных на применение владение новых технологий при отображении информации различного вида в графической форме.

### **2 Место дисциплины (модуля) в структуре образовательной программы**

Дисциплина Средства и методы визуализации информации входит в часть учебного плана формируемую участниками образовательных отношений образовательной программы.

Для изучения дисциплины необходимы знания (умения, владения), сформированные в результате изучения дисциплин/ практик:

Для изучения дисциплины необходимы знания (умения, владения), сформированные в результате изучения дисциплин уровня бакалавриата по направлению 09.03.01 Информатика и вычислительная техника.

Знания (умения, владения), полученные при изучении данной дисциплины будут необходимы для изучения дисциплин/практик:

Методология и методы научного исследования

Основы научной коммуникации

Представление результатов научных исследований

Современные проблемы информатики и вычислительной техники

Выполнение и защита выпускной квалификационной работы

### **3 Компетенции обучающегося, формируемые в результате освоения дисциплины (модуля) и планируемые результаты обучения**

В результате освоения дисциплины (модуля) «Средства и методы визуализации информации» обучающийся должен обладать следующими компетенциями:

Код индикатора	Индикатор достижения компетенции
ПК-4	Обладает способностью к разработке компонентов системы управления базами данных, отладке разрабатываемой системы управления базами данных, документированию разработанной системы управления базами данных в целом и ее компонентов и сопровождению созданной системы управления базами данных
ПК-4.1	Определяет необходимость разработки компонентов системы управления базами данных
ПК-4.2	Оценивает качество разработки компонентов системы управления базами данных



2.1	Корреляция: коэффициент Пирсона, Кенделла, Спирмена	1				10	Изучение литературы	Беседа-обсуждение	
2.2	Способы графического отображения корреляции: корреляционное поле, матрица корреляция и ее графическое отображение;					10	Изучение электронных источников	Беседа-обсуждение	
Итого по разделу						20			
3. Отображение эмпирических зависимостей									
3.1	Регрессия: парная и множественная	1				15	Повторение основ построения эмпирических уравнений	Опрос	
3.2	Способы визуализации эмпирических зависимостей: диаграмма рассеяния, диаграмма остатков, квартет Эсконба, квантильный график распределения остатков, 3D график распределения остатков, изображение временных рядов, сезонная модель временного ряда, матрица рассеяния					15	Изучение литературы	Беседа-обсуждение	
3.3	Экзамен					8,7	Подготовка к экзамену	Экзамен	
Итого по разделу						38,7			
Итого за семестр		2	6/4И			88,7		экзамен	
Итого по дисциплине		2	6/4И			88,7		экзамен	

## **5 Образовательные технологии**

1. Традиционные образовательные технологии, ориентированные на организацию образовательного процесса и предполагающую прямую трансляцию знаний от преподавателя к аспиранту.

Формы учебных занятий с использованием традиционных технологий:

Информационная лекция – последовательное изложение материала в дисциплинарной логике, осуществляемое преимущественно вербальными средствами (монолог преподавателя).

Лабораторная работа – организация учебной работы с реальными материальными и информационными объектами, экспериментальная работа с аналоговыми моделями реальных объектов.

2. Технологии проблемного обучения – организация образовательного процесса, которая предполагает постановку проблемных вопросов, создание учебных проблемных ситуаций для стимулирования активной познавательной деятельности аспирантов.

3. Интерактивные технологии – организация образовательного процесса, которая предполагает активное и нелинейное взаимодействие всех участников, достижение на этой основе лично значимого для них образовательного результата.

Формы учебных занятий с использованием специализированных интерактивных технологий:

Лекция «обратной связи» – лекция–провокация (изложение материала с заранее запланированными ошибками), лекция–беседа, лекция–дискуссия, лекция–конференция.

4. Информационно-коммуникационные образовательные технологии – организация образовательного процесса, основанная на применении программных сред и технических средств работы со знаниями в различных предметных областях.

## **6 Учебно-методическое обеспечение самостоятельной работы обучающихся**

Представлено в приложении 1.

## **7 Оценочные средства для проведения промежуточной аттестации**

Представлены в приложении 2.

## **8 Учебно-методическое и информационное обеспечение дисциплины (модуля)**

### **а) Основная литература:**

1. Бабёнышев, С. В. Бабёнышев, С. В. Математические методы и информационные технологии в научных исследованиях : учебное пособие / С. В. Бабёнышев, Е. Н. Матеров. - Железногорск : ФГБОУ ВО Сибирская пожарно-спасательная академия ГПС МЧС России, 2018. - 215 с. - Текст : электронный. - URL: <https://new.znanium.com/catalog/product/1082157> (дата обращения: 31.03.2020)

2. 1Обработка экспериментальных данных на ЭВМ : учебник / О.С. Логунова, П.Ю. Романов, Е.А. Ильина [и др.]. – М. : ИНФРА-М, 2018. – 326 с. – (Высшее образование: Бакалавриат). – [www.dx.doi.org/10.12737/textbook\\_5aafb5a99fb14.44742313](http://www.dx.doi.org/10.12737/textbook_5aafb5a99fb14.44742313).

### **б) Дополнительная литература:**

1. Логунова, О. С. Экспертные оценки и системы в металлургии черных металлов [Электронный ресурс] : учебное пособие / О.С. Логунова, Е.А. Ильина, И.И. Мацко. - Магнитогорск : МГТУ, 2013. – 1 электрон. опт. диск (CD-ROM). – Режим доступа: <https://magtu.informsystema.ru/uploader/fileUpload?name=1059.pdf&show=dcatalogues/1/1119418/1059.pdf&view=true>. – Макрообъект.

2. Логунова, О. С. Эконометрика средствами Statistica 6.1. Временные и

динамические ряды [Электронный ресурс] : учебное пособие / О.С. Логунова, Е.А. Ильина, В.В. Королева ; МГТУ. – Магнитогорск, 2009. – 135 с. : ил., диагр., табл. – Режим доступа: <https://magtu.informsystema.ru/uploader/fileUpload?name=228.pdf&show=dcatalogues/1/1056118/228.pdf&view=true>. – Макрообъект.

**в) Методические указания:**

1. Логунова, О.С. Тестовые задания по дисциплине «Обработка экспериментальных данных на ЭВМ [Текст] / О.С. Логунова, Е.А. Ильина. – Магнитогорск : «МГТУ им. Г.И. Носова», 2007. – 12 с.

**г) Программное обеспечение и Интернет-ресурсы:**

**Программное обеспечение**

Наименование ПО	№ договора	Срок действия лицензии
MS Windows 7 Professional(для классов)	Д-1227-18 от 08.10.2018	11.10.2021
MS Office 2007 Professional	№ 135 от 17.09.2007	бессрочно
CorelDraw 2017 Academic Edition	Д-504-18 от 25.04.2018	бессрочно
Maple 14 Classroom License	К-113-11 от 11.04.2011	бессрочно
MathWorks MathLab v.2014 Classroom License	К-89-14 от 08.12.2014	бессрочно
MathCAD v.15 Education University Edition	Д-1662-13 от 22.11.2013	бессрочно
MS Office Project Prof 2019(для классов)	Д-1227-18 от 08.10.2018	11.10.2021
MS Office Visio Prof 2019(для классов)	Д-1227-18 от 08.10.2018	11.10.2021
Texmaker	свободно распространяемое	бессрочно
MS Visual Studio 2010 Professional(для класса)	Д-1227-18 от 08.10.2018	11.10.2021
STATISTICA в.6	К-139-08 от 22.12.2008	бессрочно

**Профессиональные базы данных и информационные справочные системы**

Название курса	Ссылка
Национальная информационно-аналитическая система – Российский индекс научного цитирования (РИНЦ)	URL: <a href="https://elibrary.ru/project_risc.asp">https://elibrary.ru/project_risc.asp</a>

Поисковая система Академия Google (Google Scholar)	URL: <a href="https://scholar.google.ru/">https://scholar.google.ru/</a>
Федеральное государственное бюджетное учреждение «Федеральный институт промышленной собственности»	URL: <a href="http://www1.fips.ru/">http://www1.fips.ru/</a>

### **9 Материально-техническое обеспечение дисциплины (модуля)**

Материально-техническое обеспечение дисциплины включает:

1. Лекционная аудитория Мультимедийные средства хранения, передачи и представления информации
2. Компьютерный класс. Персональные компьютеры с виртуальной машиной для установки серверного ПО, выходом в Интернет и с доступом в электронную информационно-образовательную среду университета.
3. Аудитории для самостоятельной работы: компьютерные классы; читальные залы библиотеки. Все классы УИТ и АСУ с персональными компьютерами, выходом в Интернет и с доступом в электронную информационно-образовательную среду университета.
4. Аудиторий для групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации. Ауд. 282 и классы УИТ и АСУ.
5. Помещения для самостоятельной работы обучающихся, оснащенных компьютерной техникой с возможностью подключения к сети «Интернет» и наличием доступа в электронную информационно-образовательную среду организации. Классы УИТ и АСУ.
6. Помещения для хранения и профилактического обслуживания учебного оборудования. Центр информационных технологий – ауд. 372.

Учебно-методическое обеспечение самостоятельной работы обучающихся

**ЛАБОРАТОРНАЯ РАБОТА  
СОЗДАНИЕ ТАБЛИЦ ИСХОДНЫХ ДАННЫХ ДЛЯ ЗАДАЧИ  
В СИСТЕМЕ STATISTICA**

*Цель работы:* освоить навыки по организации исходных данных в пакете Statistica.

Знакомство с системой Statistica следует начать с ввода данных. Данные в этой системе организуются в виде таблиц. Табличная структура данных позволяет естественно отобразить большинство реальных данных.

Электронная таблица состоит из строк и столбцов. Столбцы таблицы Statistica называются *Variables – Переменные*, а строки *Cases – Наблюдения*.

**Задание 1.1**

Создать файл, содержащий сведения о цене рекламных щитов. Исходные данные приведены в табл. 1.1.

Таблица 1.1

Статистические данные о цене рекламы			
Длина, мм	Ширина, мм	Площадь, мм	Цена, руб.
47	35	1645	1446000
47	73	3431	2768000
47	111	5217	3974000
47	149	7003	5147000
47	209	9823	6229000
47	225	10575	7537000
47	263	12361	8828000
47	301	14147	10260000

*Указания к выполнению задания*

Для запуска системы Statistica в группе *Программы* кнопки *Пуск* выберите подгруппу *Statistica*. При этом на экране появится список доступных моделей системы (см. рис.1.1).

Выберите модуль *Basic Statistics* (Базовая статистика). По умолчанию открывается файл, с которым производились последние действия и окно для выбора действий в открытой таблице. Выберите кнопку *Cancel* и закройте существующий файл. Выполним создание новой таблицы данных.

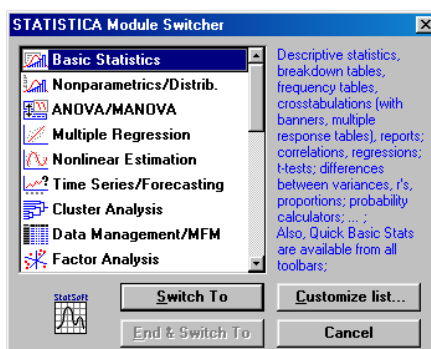


Рис. 1.1. Список модулей системы Statistica

Выполните команду: *File → New Data*. В окне *New Data* в поле *Save in* **определите рабочую папку** и в поле *File name* – имя рабочего файла (см. рис. 1.2). Для дальнейшей работы создайте файл с именем *reklama1.sta*.

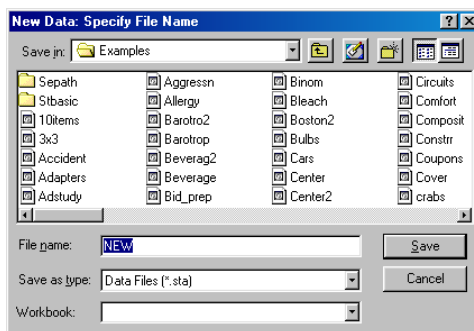


Рис. 1.2. Вид диалогового окна при создании файла

По умолчанию создается таблица, которая содержит 10 строк и 10 столбцов. Для создания таблицы необходимо переименовать столбцы данных и убрать лишние столбцы и строки. Для этого на панели инструментов выберите кнопку *Var* и для нее подпункт *Delete*. В диалоговом окне определите значения полей: *From variable* (Начиная с переменной) – *Var5*, *To variable* (Заканчивая переменной) – *Var10* (см. рис. 1.3). Данное действие определяет какие из столбцов следует удалить из таблицы.

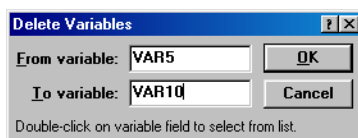


Рис. 1.3. Вид окна для удаления лишних столбцов

Необходимое число наблюдений 8. Поэтому необходимо удалить 9 и 10 строку. На панели инструментов выберите кнопку *Cases* и для нее подпункт *Delete*. Определите параметры окна так, как указано на рис. 1.4.

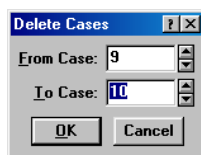


Рис. 1.4. Вид окна для удаления лишних строк

Для определения имен переменных необходимо вызвать Специальное окно (см. рис. 1.5), дважды щелкнув мышью по названию столбца *Var1*. Для всех переменных определите следующие значения параметра *name*: для первого столбца – ДЛИНА; для второго – ШИРИНА; для третьего – ПЛОЩАДЬ; для четвертого – ЦЕНА. Для всех столбцов определите ширину столбцов в 10 символов, количество знаков после запятой – 0, тип данных – Number. Для третьего столбца в поле для ввода формул введите выражение:  $=v1*v2$ . С помощью этой формулы вычисляется площадь стенда по длине и ширине.

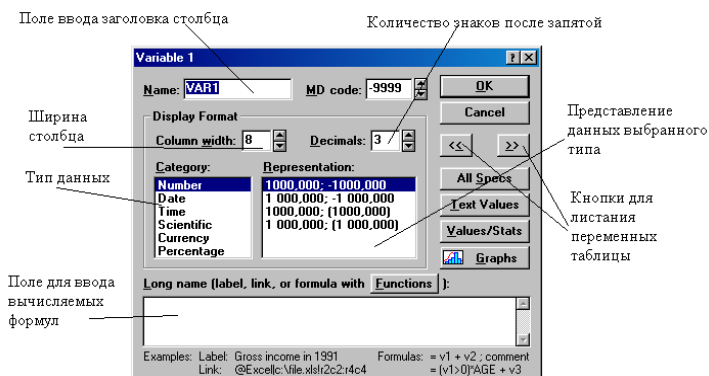


Рис. 1.5. Вид окна определения параметров для переменных

Введите данные в первый и второй столбец таблицы. Для пересчета данных в третьем столбце на панели инструментов кнопку с картинкой  $x=?$  (*Recalculate variables*) или клавишу F9. Заполните

данные четвертого столбца. При правильном выполнении задания Вы получите окно с таблицей, приведенной на рис. 1.6.

	1	2	3	4
	ДЛИНА	ШИРИНА	ПЛОЩАДЬ	ЦЕНА
1	47	35	1645	1446000
2	47	73	3431	2768000
3	47	111	5217	3974000
4	47	149	7003	5147000
5	47	209	9823	6290000
6	47	225	10575	7537000
7	47	263	12361	8828000
8	47	301	14147	10260000

Рис. 1.6. Вид окна с таблицей исходных данных по рекламе

Сохраните полученный файл.

### Задание 1.2

Создайте таблицу, содержащую сведения о прожиточном уровне пенсионеров в месяц по территории Центрального района за 1995 г. (см. табл. 1.2).

### Замечание

Значение первого столбца с названием стран должны располагаться в заголовках строк.

Таблица 1.2

Исходные данные к задаче 1.2

Район	Средний размер назначенных ежемесячных пенсий, тыс. руб., у	Прожиточный минимум в среднем на одного пенсионера в месяц, тыс. руб., х
Брянская обл.	240	178
Владимирская обл.	226	202
Ивановская обл.	221	197
Калужская обл.	226	201
Костромская обл.	220	189
г. Москва	250	302
Московская обл.	237	215
Орловская обл.	232	166
Рязанская обл.	215	199
Смоленская обл.	220	180
Тверская обл.	222	181
Тульская обл.	231	186
Ярославская обл.	229	250

### Задание 1.3

Создайте таблицу, содержащую сведения по данным, представленным в табл. 1.3. Изучается зависимость индекса человеческого развития от переменных:  $x_1$  – ВВП 1997 г., % к 1990 г.;  $x_2$  – расходы на конечное потребление в текущих ценах, % к ВВП;  $x_3$  – расходы домашних хозяйств, % к ВВП;  $x_4$  – валовое накопление, % к ВВП;  $x_5$  – суточная калорийность питания населения, ккал на душу населения;  $x_6$  – ожидаемая продолжительность жизни при рождении 1997 г. число лет.

Таблица 1.3

Исходные данные к задаче 1.3

Страна	у	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$
Австрия	0,904	115,0	75,5	56,1	25,2	3343	77,0
Австралия	0,922	123,0	78,5	61,8	21,8	3001	78,2
Белоруссия	0,763	74,0	78,4	59,1	25,7	3101	68,0
Бельгия	0,923	111,0	77,7	63,3	17,8	3543	77,2
Великобритания	0,918	113,0	84,4	64,1	15,9	3237	77,2
Германия	0,906	110,0	75,9	57,0	22,4	3330	77,2

Страна	у	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>
Дания	0,905	119,0	76,0	50,7	20,6	3808	75,7
Индия	0,545	146,0	67,5	57,1	25,2	2415	62,6
Испания	0,894	113,0	78,2	62,0	20,7	3295	78,0
Италия	0,900	108,0	78,1	61,8	17,5	3504	78,2
Канада	0,932	113,0	78,6	58,6	19,7	3056	79,0
Казахстан	0,740	71,0	84,0	71,7	18,5	3007	67,6
Китай	0,701	210,0	59,2	48,0	42,4	2844	69,8
Латвия	0,744	94,0	90,2	63,9	23,0	2861	68,4
Нидерланды	0,921	118,0	72,8	59,1	20,2	3259	77,9
Норвегия	0,927	130,0	67,7	47,5	25,2	3350	78,1
Польша	0,802	127,0	82,6	65,3	22,4	3344	72,5
Россия	0,747	61,0	74,4	53,2	22,7	2704	66,6
США	0,927	117,0	83,3	67,9	18,1	3642	76,7
Украина	0,721	46,0	83,7	61,7	20,1	2753	68,8
Финляндия	0,913	107,0	73,8	52,9	17,3	2916	76,8
Франция	0,918	110,0	79,2	59,9	16,8	3551	78,1
Чехия	0,833	99,2	71,5	51,5	29,9	3177	73,9
Швейцария	0,914	101,0	75,3	61,2	20,3	3280	78,6
Швеция	0,923	105,0	79,0	53,1	14,1	3160	78,5

## ОПИСАТЕЛЬНЫЕ СТАТИСТИКИ. ПОСТРОЕНИЕ ПРОСТЕЙШИХ СТАТИСТИЧЕСКИХ ГРАФИКОВ

*Цель работы:* освоить навыки вычисления описательных статистик и построения простейших статистических графиков.

К числу основных описательных статистик относятся: среднее, выборочная дисперсия, стандартное отклонение, медиана, мода, максимальное и минимальное значения, размах, квантили. Вычисление описательных характеристик в *Statistica* выполняется чрезвычайно просто.

Покажем расчет описательных характеристик для таблицы с данными о рекламе. Откройте файл *reclama1.sta*.

**Шаг 1.** Выделите столбец ЦЕНА. Для выделения столбца выполните одинарный щелчок клавишей мыши на заголовке столбца. Вызовите контекстное меню для этого столбца. Результат работы представлен на рис. 2.1.

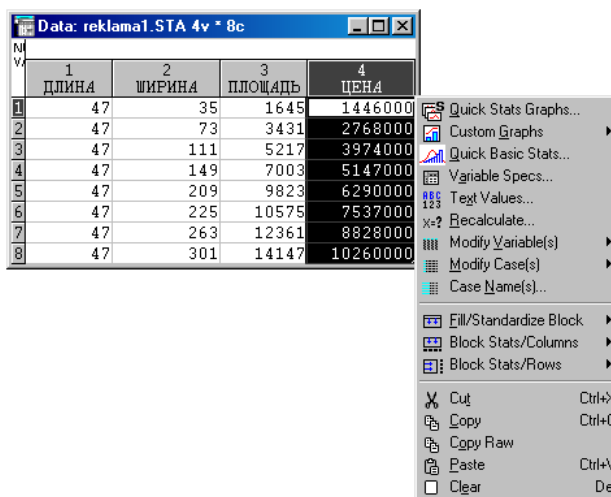


Рис. 2.1. Вид рабочей таблицы после вызова контекстного меню

**Шаг 2.** Выберите кнопку *Quick Basic Stats* (Быстрые базовые статистики). Быстрые основные статистики позволяют быстро рассчитать основные статистики для одной переменной или для целого списка переменных. Все виды анализа могут быть проведены для сгруппированных данных. После выбора кнопки для быстрых статистик появится ниспадающее меню.

**Шаг 3.** В выпадающем меню выбрать верхнюю строку *Descriptives of ЦЕНА* (Описание переменной ЦЕНА). Результаты работы приведены на рис. 2.2.

Таблица результатов отличается от таблицы с исходными данными. Этот специальный тип таблиц в *Statistica* называется **scrollsheets**.

The first screenshot (a) shows the 'BASIC STATS' section of the Descriptive Statistics dialog. The second screenshot (б) shows the 'Descriptives' section.

BASIC STATS	Valid N	Mean	Confid. -95,000%	Confid. 95,000	Sum	Minimum	Maximum	Range
ЦЕНА	8	5781250.	3248904.	8313596.	4625E4	1446000.	1026E4	8814000.

Descriptives	Range	Variance	Std. Dev.	Standard Error	Skewness	Std. Err. Skewness	Kurtosis	Std. Err. Kurtosis
ЦЕНА	8814000.	917512E7	3029047.	1070930.	.059425	.752101	-1.05283	1.480880

Рис. 2.2. Таблица с описательными статистиками столбца ЦЕНА

Прокручивая электронную таблицу результатов, вы увидите слева направо следующие описательные статистики переменной **ЦЕНА**:

- Valid N – истинное число наблюдений переменной ЦЕНА (число наблюдений без пропусков);
- Mean – выборочное Среднее;

- Confid - 95% – нижняя граница 95% доверительного интервала для среднего;
- Confid + 95% – верхняя граница 95% доверительного интервала для среднего;
- Sum – сумма (сумма значений переменной ЦЕНА);
- Minimum – минимум (минимальное значение переменной ЦЕНА);
- Maximum – максимум (максимальное значение переменной ЦЕНА);
- Range – размах (то есть разность между максимумом и минимумом);
- Variance – выборочная дисперсия;
- Std.Dev. – стандартное отклонение;
- Std.Err. – стандартная ошибка;
- Skewness – выборочный коэффициент асимметрии;
- Std.Err. Skewness – стандартная ошибка коэффициента асимметрии;
- Kurtosis – выборочный коэффициент эксцесса;
- Std.Err. Kurtosis – стандартная ошибка эксцесса.

Корреляция, или, точнее, коэффициент корреляции, является мерой зависимости двух величин. Коэффициент корреляции – это безразмерная величина.

Значение коэффициента корреляции лежит между -1 и +1. Если наблюдается тенденция возрастания одной величины при росте другой, то говорят о положительной коррелированности величин, если наблюдается тенденция увеличения одной величины при уменьшении другой, то говорят об отрицательной коррелированности величин.

### Задание 2.1

Выполнить построение матрицы корреляции для данных представленных в задании 1.3.

#### Выполнение задания

**Шаг 1.** Откройте файл с данными задания 1.3. Выделите столбцы для X1 – X6.

**Шаг 2.** Для выделенных столбцов вызовите контекстное меню и выберите пункт *Quick Basic Stats*.

**Шаг 3.** В ниспадающем меню выберите пункт *Correlations of X1 – X6* (см. рис. 2.3).

1	2	3	4	5	6	7	8
СТРАНА	Y	X1	X2	X3	X4	X5	X6
1 Австрия	,904				25,2	3343	77,0
2 Австрали	,922				21,8	3001	78,2
3 Белорусс	,763				25,7	3101	68,0
4 Бельгия	,923				17,8	3543	77,2
5 Великобр	,918				15,9	3237	77,2
6 Германия	,906				22,4	3330	77,2
7 Дания	,905				20,6	3808	75,7
8 Индия	,545				25,2	2415	62,6
9 Испания	,894				20,7	3295	78,0
10 Италия	,900				17,5	3504	78,2
11 Канада	,932				19,7	3056	79,0
12 Казахстан	,740	71,0	84,0	71,7	18,5	3007	67,6
13 Китай	,701	210,0	59,2	48,0	42,4	2844	69,8
14 Лаввия	,744	94,0	90,2	63,9	23,0	2861	68,4
15 Нидерлан	,921	118,0	72,8	59,1	20,2	3259	77,9
16 Норвегия	,927	130,0	67,7	47,5	25,2	3350	78,1
17 Польша	,802	127,0	82,6	65,3	22,4	3344	72,5
18 Россия	,747	61,0	74,4	53,2	22,7	2704	66,6
19 США	,927	117,0	83,3	67,9	18,1	3642	76,7
20 Украина	,721	46,0	83,7	61,7	20,1	2753	68,8
21 Финлянди	,913	107,0	73,8	52,9	17,3	2916	76,8
22 Франция	,918	110,0	79,2	59,9	16,8	3551	78,1
23 Чехия	,833	99,2	71,5	51,5	29,9	3177	73,9
24 Швейцари	,914	101,0	75,3	61,2	20,3	3280	78,6
25 Швеция	,923	105,0	79,0	53,1	14,1	3160	78,5

Рис. 2.3. Этап работы для построения матрицы корреляции

**Шаг 4.** Выбор пункта *Correlations of X1 – X6* приводит к построению матрицы корреляции (см. рис. 2.4).

Continue...	X1	X2	X3	X4	X5	X6
X1	1,00	-,63	-,37	,54	,08	,16
X2	-,63	1,00	,76	-,67	,19	,05
X3	-,37	,76	1,00	-,50	,11	-,05
X4	,54	-,67	-,50	1,00	-,33	-,41
X5	,08	,19	,11	-,33	1,00	,70
X6	,16	,05	-,05	-,41	,70	1,00

Рис. 2.4. Вид окна с таблицей корреляции для факторов X1 – X6

Следует иметь в виду, что недостаточно выполнить расчет коэффициентов корреляции. Необходимо проводить дальнейший корреляционный анализ.

Графическое представление исходных данных всегда считалось лучшим средством визуализации данных. Приведем несколько примеров для построения графиков различных типов.

### Задание 2.2

Построить диаграмму рассеяния для изучения зависимости стоимости рекламы от ее площади.

#### Выполнение задания

**Шаг 1.** Откройте файл *reklama1.sta*.

**Шаг 2.** Установите курсор в столбец *Площадь* и вызовите контекстное меню. В этом меню выберите пункт *Quick Stats Graphs* (Быстрые Stats-графики).

**Шаг 3.** В выпадающем меню выберите пункт *Scatterplot by* (Диаграмма рассеяния) и далее пункт *Regular* (Регулярный).

**Шаг 4.** В диалоговом окне *Select one variable* (Выбрать одну переменную) выберите переменную *ЦЕНА*. Активизируйте кнопку *Ok*. В результате получите график представленный на рис. 2.5.

**Шаг 5.** При необходимости можно сохранить график как отдельный файл. Горячие клавиши для сохранения графика *Ctrl+S*.

Выполним настройку диаграммы. Каждая область диаграммы является активной и для нее можно вызвать диалоговое окно с параметрами настройки.

Выполните двойной щелчок кнопкой мыши на точках диаграммы рассеяния. При этом на экран будет выведено диалоговое окно с параметрами меток графика (см. рис. 2.6).

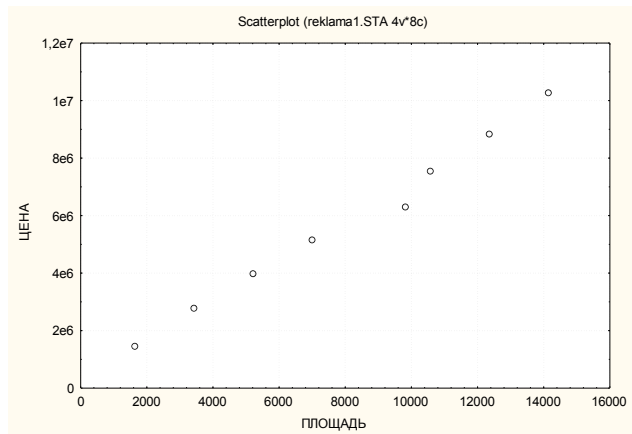


Рис. 2.5. Диаграмма рассеяния для данных из файла *reklama1.sta*

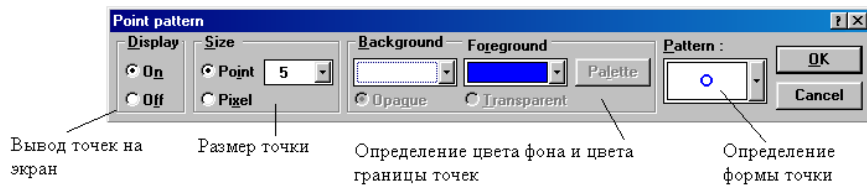


Рис. 2.6. Вид окна параметров для настройки меток значений на диаграмме

Установите параметры окна следующим образом: *Display – On; Size Point – 7; Background – синий; Foreground – синий*; форму метки оставьте без изменения.

Выполните двойной щелчок на сетке диаграммы рассеяния и установите параметры как показано на рис. 2.7.

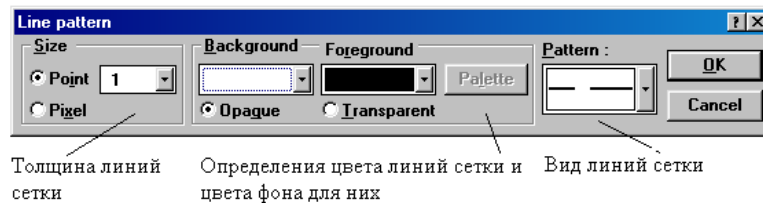


Рис. 2.7. Вид окна параметров сетки диаграммы рассеяния

Двойной щелчок на подписи к осям координат приводит к открытию диалогового окна для форматирования всех подписей диаграммы (см. рис. 2.8).

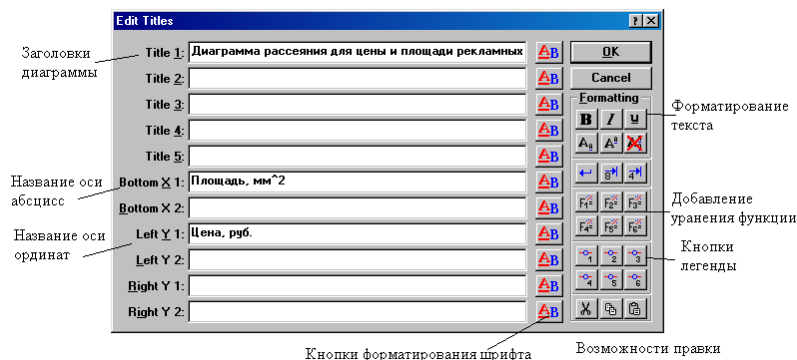


Рис. 2.8. Вид окна для форматирования подписей к диаграмме

Установите параметры как показано на рис. 8.

Настройка осей координат происходит при вызове диалогового окна *Scale Options* (см. рис. 2.9).

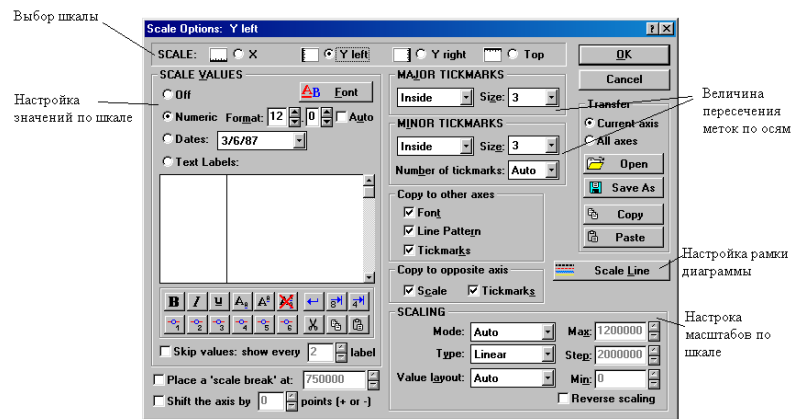


Рис. 2.9. Диалоговое окно настройки осей на диаграмме рассеяния  
Выполните на свое усмотрение настройку диаграммы, используя возможности окна параметров.

В области диаграммы можно вызвать контекстное меню для изменения общих настроек диаграммы (*Change General Layout*) и настроек ряда (*Change Plot Layout*).

Выберите в контекстном меню пункт *Change General Layout* и установите в списке *Graph Type* параметр *Step Plot*. Перейдите в диалоговое окно настройки ряда и с помощью кнопки *Plot Layout*. В диалоговом окне *Plot 1* выберите кнопку *Data Labels* и активизируйте переключатель в рамке *Coordinates – X*. Подтвердите принятие всех установок. При правильном выполнении всех шагов получите диаграмму, приведенную на рис. 2.10.

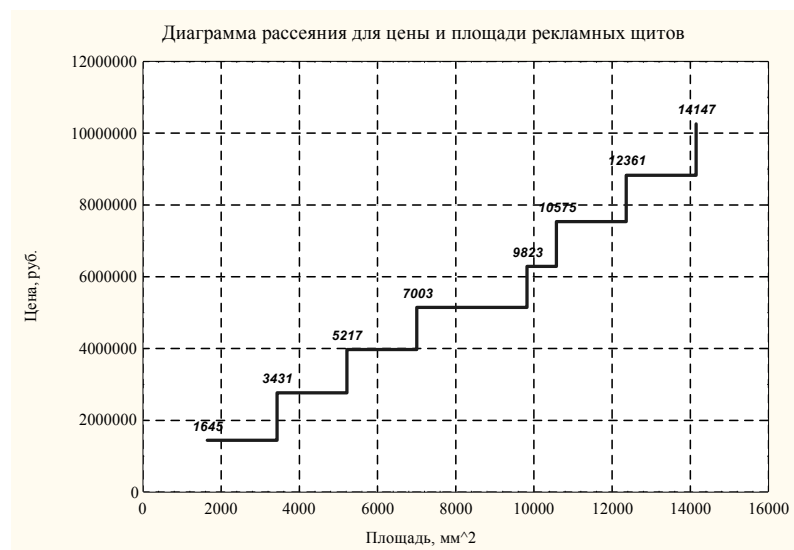


Рис. 2.10. Преобразованная диаграмма рассеяния  
Верните диаграмму к исходному виду.

### Задание 2.3

Выполнить построение столбчатых гистограмм для данных задания 1.2.

#### Выполнение задания

**Шаг 1.** Вызовите контекстное меню для столбца с данными *y* и выберите пункт Быстрые статистические графики.

**Шаг 2.** В ниспадающем меню откройте группу *Histogram of Пенсия*.

**Шаг 3.** Выберите пункт *Normal Fit* (Диаграмма с нормальным распределением).

**Шаг 4.** Приведите диаграмму к виду, приведенному на рис. 2.11.

Вторая строка заголовка характеризует параметры нормального распределения для данной величины и представляет собой уравнение плотности распределения с данным средним значением и среднеквадратичным отклонением. Для данного столбца среднее значение составляет 228,3846 и среднеквадратичное отклонение или стандартная ошибка – 9,67418.



Рис. 2.11. Вид столбчатой диаграммы для столбца, содержащего данные о среднем размере пенсии

#### Задание 2.4

Постройте для данных задания 1.2 диаграмму, приведенную на рис. 2.12.

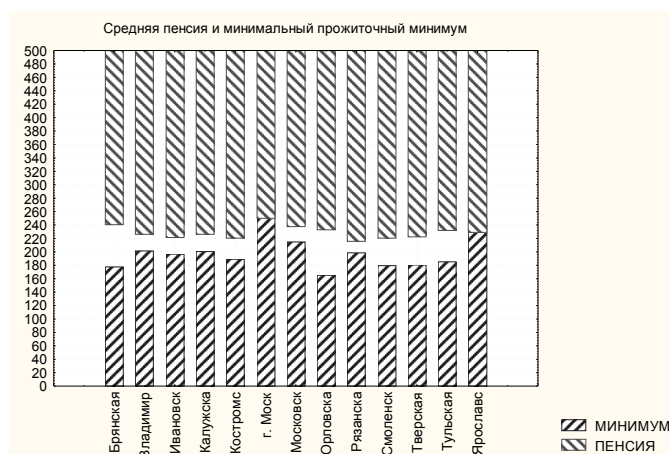


Рис. 2.12. Диаграмма для сравнения исходных данных задания 1.2

При изучении множественных связей в *Statistica* допускается построение графиков, характеризующих зависимость функции от двух переменных. Для этого в *Custom Graphs* (Выборочные графики) предусмотрены *3D XYZ Graphs...* Выполним построение поверхности рассеяния для данных задания 1.3.

В качестве функции отклика будем использовать *Уровень интеллектуального развития человека*, в качестве независимых факторов: *расходы на домашнее хозяйство и суточную калорийность питания*.

**Шаг 1.** В области одного из столбцов вызовите контекстное меню, выберите пункт *Custom Graphs* и в нем подпункт *3D XYZ Graphs...*

**Шаг 2.** В диалоговом окне *Spreadsheets: Custom 3D Scatterplots and Surfaces* (см. рис. 2.13) в списке *Graph Type* выберите пункт *Surface Plot* и в списке *Data Type* – *Spline*. В полях *Variable* определите названия переменных по осям координат: X – X3; Y – X5 и Z1 – Y.

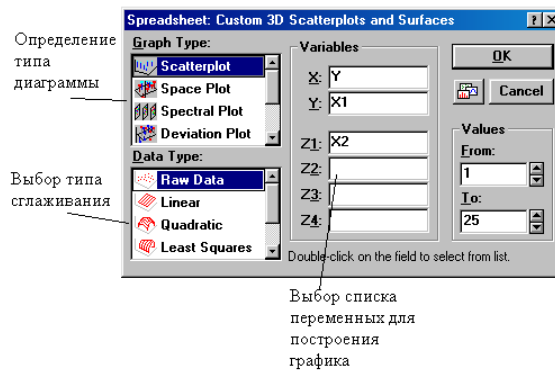


Рис. 2.13. Вид окна для определения параметров 3D Graph

**Шаг 3.** Приведите диаграмму к виду, представленному на рис. 2.14.

**Шаг 4.** Выберите наиболее наглядный вид для поверхности, используя вращение диаграммы. Для этого в контекстном меню для области построения выберите пункт *Rotate Graph/Perspective*.

Одним из мощных средств разведочного анализа данных являются многомерные пиктографики. Главная идея пиктографиков состоит в представлении отдельных единиц наблюдения в виде определенных графических объектов; при этом значения переменных ставятся в соответствие определенным характеристикам или параметрам этих объектов (обычно одно наблюдение = одному объекту). Это соответствие таково, что общий вид объектов меняется как функция конфигурации значений. Таким образом, наблюдатель может идентифицировать уникальные для каждой конфигурации значений наглядные характеристики объектов. Исследование таких пиктограмм может помочь обнаружить определенные группы простых зависимостей и взаимосвязей между переменными.

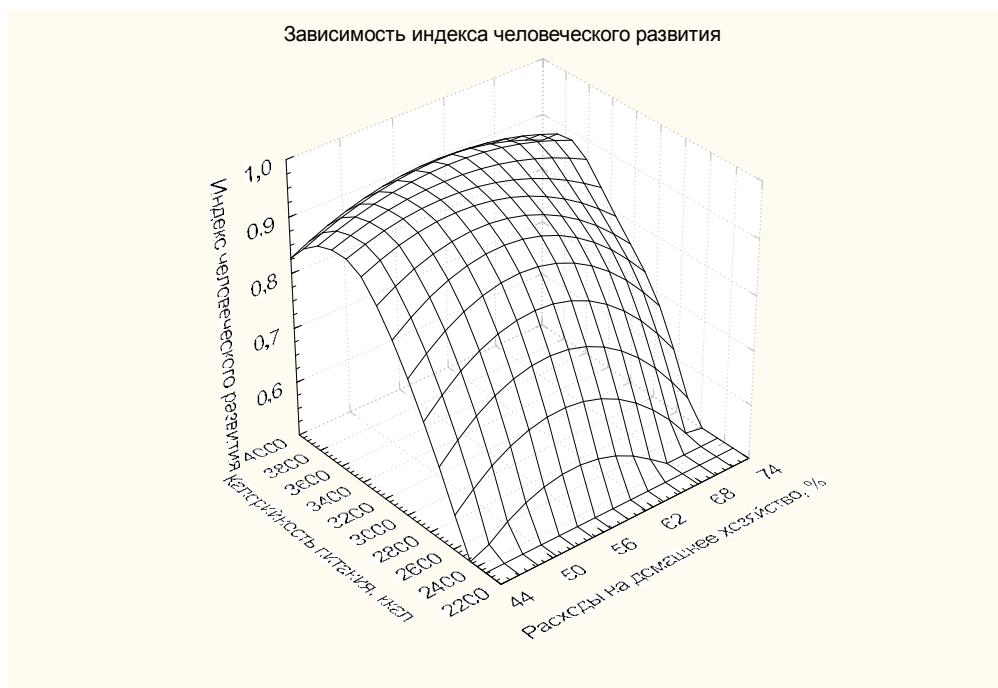


Рис. 2.14. Вид

диаграммы *Surface Plot*

«Лица Чернова» – это один из наиболее искусно разработанных типов пиктографики. Для каждого наблюдения рисуется отдельное «лицо», где относительные значения выбранных переменных представлены как формы и размеры отдельных черт лица (например, длина носа, угол между бровями, ширина лица).

**Задание 2.5**

Выполните построение для данных задания 1.3 пиктографической диаграммы лиц Чернова.

**Выполнение задания**

**Шаг 1.** Для построения графика в области любого столбца вызовите контекстное меню и выберите пункт *Custom Graph* и подпункт *Icon Plots*.

**Шаг 2.** В открывшемся диалоговом окне установите параметры как показано на рис. 2.15.

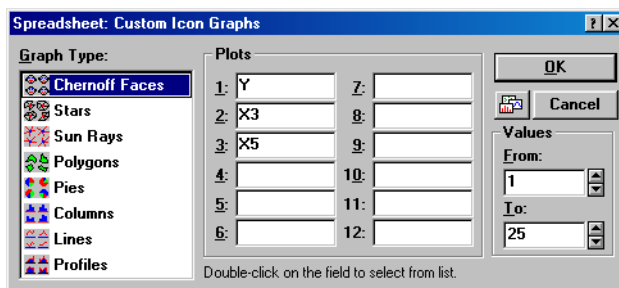


Рис. 2.15. Вид окна с установленными параметрами для построения лиц Чернова

**Шаг 3.** Подтвердите построение графика и приведите его к виду, представленному на рис. 2.16.

Изменение параметров диаграммы возможно при вызове окон с параметрами диаграммы двойным щелчком клавиши мыши по выбранной области.



Рис. 2.16. Диаграмма с «лицами Чернова»

**Задание 2.6**

Выполните пояснения на русском языке всех элементов «лиц Чернова». Например: FACE WIDTH – ширина лица.

Определите для функции отклика значение параметра – размер уха.

**Задание 2.7**

Постройте диаграмму с «лицами Чернова» по всем факторам и функции отклика.

**Задание 2.8**

Выполните изменение вида диаграммы «лица Чернова», построив ее же вид с лучами, со звездами и многоугольниками.

## РЕГРЕССИОННЫЙ АНАЛИЗ ДАННЫХ

*Цель работы:* освоить навыки построения регрессионных моделей и оценки параметров этих моделей.

Прогнозирование и управление выбранным процессом чаще всего производят на основе эмпирических зависимостей, поэтому одной из стандартных практических задач является определение таких зависимостей по известному набору данных. Например, для определения стоимости рекламного щита с размерами 45×47 мм необходимо найти зависимость цены рекламы ( $Y$ ) от размеров объявления ( $X$  – ширины объявления).

Переменная  $X$  носит название независимой переменной, или предиктора, переменная  $Y$  называется зависимой переменной, или откликом. Значение переменной  $X$  в  $i$ -том опыте будем обозначать через  $X(i)$ , соответствующее значение величины  $Y$  обозначим через  $Y(i)$ ,  $0 < i \leq n$ . В статистике подобные задачи решаются в рамках регрессионного анализа.

Пусть наблюдаемые величины связаны между собой регрессионной зависимостью вида:

$$Y(i) = B_i \cdot X(i) + B_0 + e(i), 0 < i \leq n, \text{ где}$$

$B_i, B_0$  – неизвестные константы или параметры модели,  $e(i)$  – ненаблюдаемые случайные величины (наблюдаются только  $X(i), Y(i), 0 < i \leq n$ ) со средним 0 (как говорят является несмещенными) и неизвестной дисперсией, не меняющейся от опыта к опыту.

Иногда случайные величины  $e(i)$  предполагается, что они не коррелированы в разных опытах. Кроме того, часто предполагается, что ошибки имеют нормальное распределение. В этом случае некоррелированность влечет независимость.

Можно рассматривать и более общие линейные модели, например, с несколькими независимыми переменными:

$$Y(i) = B_1 \cdot X_1(i) + B_2 \cdot X_2(i) + \dots + B_K \cdot X_K(i) + B_0 + e(i), 0 < i \leq n, \text{ где}$$

$B_0, B_1, B_2, \dots, B_K$  – неизвестные константы.

Общая задача регрессионного анализа состоит в том, чтобы по наблюдениям  $(X(1), Y(1)), \dots, (X(n), Y(n))$ :

- оценить параметры модели  $B_i, B_0$  наилучшим образом;
- построить доверительные интервалы для  $B_1, B_0$ .
- проверить гипотезу о значимости регрессии;
- оценить степень адекватности модели и т.д.

К основным понятиям регрессионного анализа относятся:

1) *предсказанные значения (Predictable values)*: значения  $Y$ -ов вычисленные по уравнению с оцененными параметрами. Значения  $Y$ -ов, предсказанных в точках  $X(i)$ , будем обозначать  $PrY(i)$ ,  $0 < i \leq n$ ;

2) *остатки (Residuals)*: разности между наблюдаемыми значениями и предсказанными:  $Y(i) - PrY(i)$ ,  $0 < i \leq n$ ;

3) *сумма квадратов  $Y$ -ов, скорректированная на среднее, – SS*:

$$SS = (Y(1) - \bar{Y})^2 + (Y(2) - \bar{Y})^2 + \dots + (Y(n) - \bar{Y})^2,$$

где  $\bar{Y}$  – среднее  $Y$ -ов –  $\bar{Y} = \frac{Y(1)+Y(2)+\dots+Y(n)}{n}$ ;

4) сумма квадратов  $\text{Pr}Y(i)$ , скорректированная на среднее  $SSPr$ :

$$SSPr = (\text{Pr}Y(1) - \bar{Y})^2 + (\text{Pr}Y(2) - \bar{Y})^2 + \dots + (\text{Pr}Y(n) - \bar{Y})^2;$$

5) сумма квадратов остатков  $SSRes$ :

$$SSRes = (Y(1) - \text{Pr}Y(1))^2 + (Y(2) - \text{Pr}Y(2))^2 + \dots + (Y(n) - \text{Pr}Y(n))^2;$$

б) коэффициент детерминации  $R^2$ :  $R^2 = \frac{SSPr}{SS}$ .

### Задание 3.1

Исследовать, как цена рекламы зависит от ширины объявления. (reklama1.sta).

#### Выполнение задания

**Шаг 1.** Откройте файл *reklama1.sta*.

**Шаг 2.** Из переключателей модулей Statistica откройте модуль *Multiple regression* (Множественная регрессия). Высветите название модуля и далее нажмите кнопку *Switch to* (Переключиться в) либо просто дважды щелкните мышью по названию модуля: *Multiple regression*.

**Шаг 3.** На экране появится диалоговое окно, приведенное на рис. 3.1.

Нажмите кнопку *Open Data* (Открыть данные) и откройте файл *reklama1.sta*. Далее выберите переменные для анализа. Выбор переменных осуществляется с помощью кнопки *Variables* (Переменные), находящейся в левом верхнем углу панели.

После того как кнопка будет нажата, на экране появится диалоговое окно *Select dependent and independent variable list* (Выбрать списки зависимых и независимых переменных).

Высветив имя переменной в левой части окна, выберите зависимую переменную. Высветив имя переменной в правой части окна, выберите независимую переменную. То же можно сделать набрав номера переменных в строках: *Dependent variable list* (Список зависимых переменных) и *Independent variable list* (Список независимых переменных).

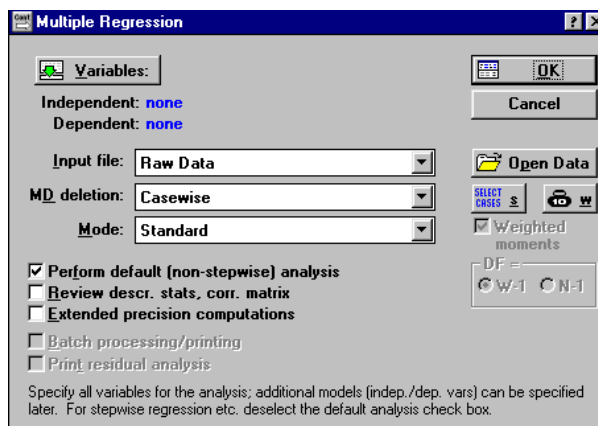


Рис. 3.1. Стартовая панель модуля Множественная регрессия

В данном примере независимой переменной является Ширина, зависимой – Цена. Высветив имена этих переменных, как показано на рис. 3.2, нажмите кнопку *Ok* в правом верхнем углу диалогового окна *Select dependent and independent variable list* (Выбрать списки зависимых и

независимых переменных). На экране откроется стартовая панель модуля *Multiple regression* (Множественная регрессия). Нажмите кнопку *Ok* в правом углу стартовой панели.

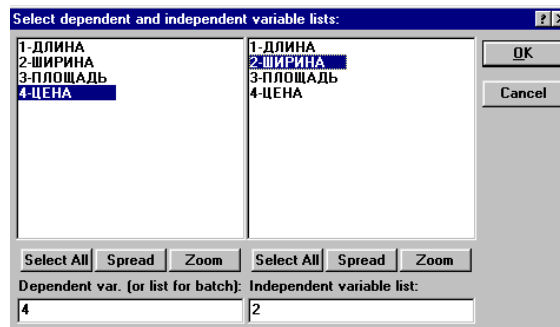


Рис. 3.2. Вид окна выбора переменных для анализа

**Шаг 4.** На экране появится диалоговое окно *Model Definition* (Построение модели). В данном окне выберите стандартный метод оценивания в опции *Method* (Метод): *Standard* (Стандартный) (см. рис. 3.3). Далее нажмите кнопку *Ok*.

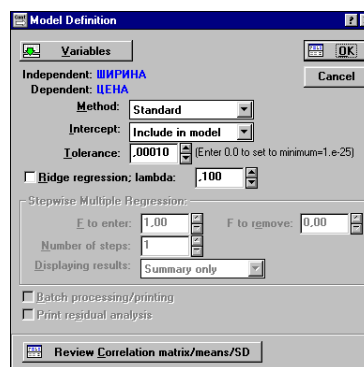


Рис. 3.3. Вид окна построения регрессионной модели в модуле

Программа произведет оценивание параметров модели стандартным методом, и через секунду на экране появится следующее диалоговое окно результатов.

**Шаг 5.** В диалоговом окне *Multiple Regression Results* (Результаты Множественной регрессии) просмотрите результаты оценивания (рис. 4). Результаты можно просмотреть в численном и в графическом виде. Окно результатов анализа имеет следующую структуру: верх окна – информационный. Он состоит из двух частей: в первой части содержится основная информация о результатах оценивания, во второй высвечиваются значимые регрессионные коэффициенты. Внизу окна *Multiple Regression Results* (Результаты Множественной регрессии) находятся функциональные кнопки, позволяющие всесторонне просмотреть результаты анализа.

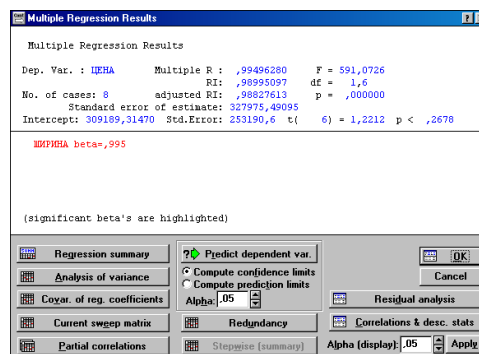


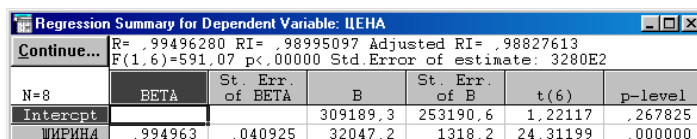
Рис. 3.4. Вид окна оценивания параметров регрессии на примере по оценке стоимости рекламы. Рассмотрим информационную часть окна. В ней содержатся краткие сведения о результатах анализа. А именно:

- *Der. Var.* (Имя зависимой переменной) – ЦЕНА.
- *No. of Cases* (Число наблюдений, по которым построена регрессия) – 8.
- *Multiple R* (Коэффициент множественной корреляции) – 0,99496280.

- $R\text{-square} - RI$  (Квадрат коэффициента множественной корреляции) – 0,98995097, обычно называемый коэффициентом детерминации. Показывает долю общего разброса (относительно выборочного среднего зависимой переменной), которая объясняется построенной регрессией. Это очень хорошее значение, показывающее, что построенная регрессия объясняет более 98% разброса значений переменной ЦЕНА относительно среднего.
- $Adjusted R\text{-square}$ : adjusted RI (Скорректированный коэффициент детерминации), определяемый как:  $Adjusted R\text{-square} = 1 - (1 - R\text{-square}) \cdot \frac{n}{n - p}$ , где  $n$  – число наблюдений в модели,  $p$  – число параметров модели (число независимых переменных плюс 1, так как в модель включен свободный член) – 0,98827613.
- $Std. Error of estimate$  (Стандартная ошибка оценки) – 327975,49095.
- $Std. Error$  (Стандартная ошибка оценки свободного члена) – 253190,6.
- Стандартная ошибка коэффициента  $B_0$  в уравнении регрессии.
- $t(df)$  and  $p\text{-value}$  (Значение  $t$ -критерия и уровень  $p$ ).  $T$ -критерий используется для проверки гипотезы о равенстве 0 свободного члена регрессии.
- $F$  – значения  $F$ -критерия – 591,0726.
- $df$  – число степеней свободы  $F$ -критерия – 1,6.
- $p$  – уровень значимости.

Рассмотрим вторую часть информационного окна. В этой части система говорит о значимых регрессионных коэффициентах, высвечивая строку: **ШИРИНА**  $beta = 0.995$  и на пояснение значимые  $beta$  высвечены – *significant beta's are highlighted*. Отметим, что в данном случае  $beta$  есть стандартизованный коэффициент  $B_1$ , то есть коэффициент при независимой переменной ШИРИНА.

Перейдем в функциональную часть окна результатов. Прежде всего нажмите кнопку *Regression summary* (Итоговый результат регрессии). На экране появится электронная таблица вывода – *spreadsheets*, в которой представлены итоговые результаты оценивания регрессионной модели (рис. 3.5).



Regression Summary for Dependent Variable: ЦЕНА						
Continue...						
R= .99496280 RI= .98995097 Adjusted RI= .98827613						
F(1,6)=591.07 p<.00000 Std Error of estimate: 3280E2						
N=8	BETA	St. Err. of BETA	B	St. Err. of B	t(6)	p-level
Intercept			309189.3	253190.6	1.22117	.267825
ШИРИНА	.994963	.040925	32047.2	1318.2	24.31199	.000000

Рис. 3.5. Итоговая таблица регрессии

Это стандартная таблица вывода регрессионного анализа. В первом столбце таблицы даны значения коэффициентов  $beta$  – стандартизованные коэффициенты регрессионного уравнения, во втором – стандартные ошибки  $beta$ , в третьем – точечные оценки параметров модели: свободный член  $B_0=309189$ , коэффициент  $B_1$  (при независимой переменной ШИРИНА)=32047.

Далее стандартные ошибки для  $B_0$ ,  $B_1$ , значения статистик  $t$ -критерия и т.д.

Из таблицы видно, что оцененная модель имеет вид:

$$\text{ЦЕНА} = 32047 \cdot \text{ШИРИНА} + 309189.$$

**Шаг 6.** Оценка адекватности модели. Важным элементом анализа является оценка адекватности модели.

Анализ адекватности основывается на анализе остатков.

Остатки представляют собой разности между наблюдаемыми значениями и модельными, то есть значениями, подсчитанными по модели с оцененными параметрами.

В Statistica в модуле *Множественная регрессия* имеется специальное диалоговое окно, в котором проводится всесторонний анализ остатков.

Для проведения анализа остатков выберите кнопку *Residual Analysis – Анализ остатков* в окне *Multiple Regression Results*.

Появится диалоговое окно представленное на рис. 3.6.

Зачем проверять остатков? Идея проста: если остатки существенно коррелированы (зависимы), то модель неадекватна (нарушено важное предположение о независимости ошибок в регрессионной модели).

Рассмотрим более подробно статистику Дарбина – Уотсона, которая является стандартом для проверки некоторых видов зависимости остатков.

Статистика Дарбина – Уотсона используется для проверки гипотезы о том, что остатки построенной регрессионной модели *некоррелированы*, против альтернативы: остатки связаны: *авторегрессионной* зависимостью вида:

$$\varepsilon_i = \rho\varepsilon_{i-1} + \delta.$$

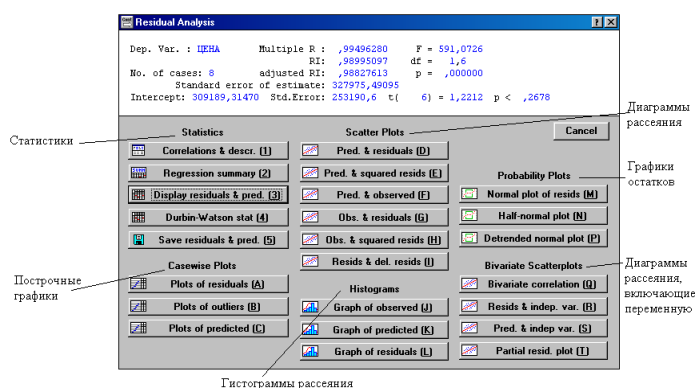


Рис. 3.6. Вид диалогового окна для Анализа остатков

Формально статистика Дарбина – Уотсона вычисляется следующим образом: 
$$d = \frac{\sum_{i=2}^n (\varepsilon_i - \varepsilon_{i-1})^2}{\sum_{i=1}^n \varepsilon_i^2}.$$

Иными словами, сумма квадратов первых разностей остатков *нормируется* суммой квадратов остатков. Вычисленное значение критерия Дарбина – Уотсона сравнивается с табличным значением для верхнего и нижнего уровней критерия, при заданном количестве предикторов и заданном уровне значимости.

Если  $d < DL_k$ , то гипотеза о независимости остатков отвергается в пользу альтернативы; если  $d > DU_k$ , то гипотеза о независимости остатков *не отвергается*; случай  $DL_k < d < DU_k$  является сомнительным. Здесь  $d$  – значение критерия Дарбина – Уотсона,  $DL_k$  – нижняя граница критерия и  $DU_k$  – верхняя граница критерия.

После того как доказана адекватность модели, полученные результаты можно уверенно использовать для дальнейших действий.

Для расчета критерия Дарбина – Уотсона выберите кнопку *\_Darbin – Watson Stat (4)*. В результате будет получена таблица (см. рис. 3.7), в которой выводятся значения статистики Дарбина – Уотсона и серийного коэффициента корреляции для остатков.

Для рассматриваемого примера, количество наблюдений  $k=8$  и количество предикторов – 1. Табличные значения критерия при 5% уровне значимости составят:  $DL_k = 0,76$  и  $DU_k=1,36$ .

	Durbin-Watson d	Serial Corr.
Estimate	1,877112	-,012394

Рис. 3.7. Вид таблицы со значениями критерия Дарбина – Уотсона  
 Проверяем заданные неравенства:  $1,877112 > 1,36$ , следовательно, гипотезу о независимости остатков отвергать нельзя. Полученные вывод подтверждается значением коэффициента корреляции. Его значение составляет  $-0,12394$ , которое является довольно низким и можно считать, что не отличается от нуля и, следовательно, можно делать вывод о независимости остатков.

**Шаг 7.** Существует довольно много способов визуального сравнения остатков. Например, сравнение наблюдаемых и предсказываемых значений (см. рис. 3.8).

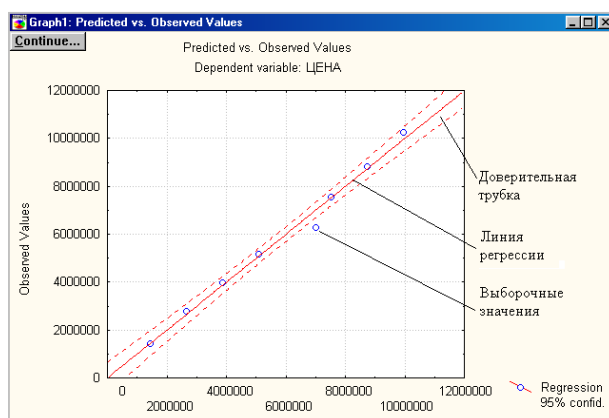


Рис. 3.8. Вид окна результатами графического сравнения наблюдаемых и предсказываемых значений функции отклика

Для получения данного графика достаточно выбрать кнопку *Pred. & Observed*.

**Шаг 8.** По построенному уравнению можно выполнить предсказание значений функции отклика. Для этого в окне *Multiple Regression Results* выберите кнопку *Predict Depended var.* (Предсказать зависимую переменную). В диалоговом окне (см. рис. 3.9) определите значение зависимой переменной равное 23 и подтвердите действие кнопкой *Ok*.

Рис. 3.9. Вид окна для определения значения независимой переменной  
 В результате получится окно с результатами предсказания и оценки этого предсказания (см. рис. 3.10).

variable	B-Weight	Value	B-Weight * Value
ШИРИНА	32047.21	23.00000	737086.
Intercept			309189.
Predicted			1046275.
-95,0%CL			491646.
+95,0%CL			1600904.

Рис. 3.10. Вид окна с результатами предсказания

### Задание 3.2

Опишите назначение всех кнопок для окна *Multiple Regression Results* и *Residual Analysis*.

### Задание 3.3

Выполнить построение множественной регрессии для задания 3.1. Исследовать, построенное уравнение на независимость остатков и подтвердить это графически. Сохраните результаты работы.

Нигде и никогда не бывает данных в чистом виде. Реальные данные всегда засорены, в них имеются выбросы, аномальные наблюдения, ошибки измерений, пропуски и т.д. Эти точки могут существенно повлиять на результаты обработки, в частности, на построение зависимостей.

В Statistica есть прекрасное средство, позволяющее визуально на графике удалять графики и группы точек. Это средство называется *Brush* (Кисть).

### Задание 3.4

Определить на диаграмме рассеяния аномальную точку и выполнить ее удаление на диаграмме рассеяния для зависимости цены рекламного щита от его ширины.

**Шаг 1.** Постройте диаграмму рассеяния *Regression? 95% conf.* Для данных из файла *reklama1.sta*, используя в качестве отклика цену на рекламные щиты и в качестве фактора – ширину щита. В результате получится график, приведенный на рис. 3.11.

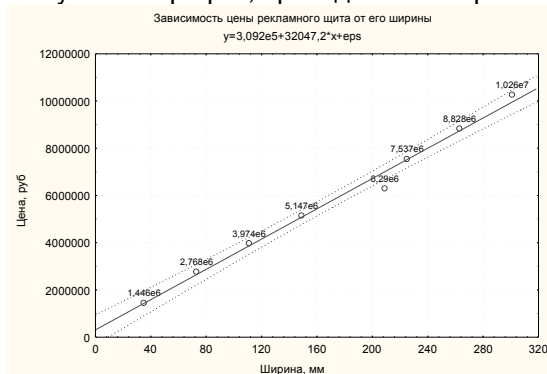



Рис. 3.11. Вид диаграммы рассеяния для зависимости цены рекламных щитов от их ширины

**Шаг 2.** На панели инструментов выберите кнопку  - *Brushing Tool* (Кисть). На экране появиться дополнительная панель инструментов *Brushing*. На этой панели активизируйте переключатели: *Turn Off* и *Point*.

**Шаг 3.** На графике выделите точку с шириной 209 мм, и на панели инструментов нажмите кнопку *Update*. В результате, выделенная точка будет удалена с графика и уравнение регрессии будет пересчитано по данным новой выборки.

**ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ПРОВЕДЕНИЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ**

**а) Планируемые результаты обучения и оценочные средства для проведения промежуточной аттестации:**

Код индикатора	Индикатор достижения компетенции	Оценочные средства
<p>ПК-4: Обладает способностью к разработке компонентов системы управления базами данных, отладке разрабатываемой системы управления базами данных, документированию разработанной системы управления базами данных в целом и ее компонентов и сопровождению созданной системы управления базами данных</p>		
ПК-4.1	<p>Определяет необходимость разработки компонентов системы управления базами</p>	<p><i>Перечень теоретических вопросов</i></p> <p>1. Опишите программные средства хранения эмпирических данных для визуализации описательных статистик, включая базы данных.</p>
ПК-4.2	<p>Оценивает качество разработки компонентов системы управления базами данных</p>	<p>2. Перечислите этапы создания портейшей базы исходных данных в пакете Statistica.</p> <p>3. Опишите действия, которые необходимо выполнить для импортирования таблиц из электронных таблиц (Excel).</p> <p>4. Определите понятия описательных статистик и средства их визуализации.</p> <p>5. Опишите назначение коэффициента корреляции и средства их визуализации.</p> <p>6. Перечислите виды диаграмм, используемые в данной работе для представления данных.</p> <p><i>Практические задания</i></p> <p>1. Опишите назначение нижеприведенных диаграмм:</p>  <p><b>Рисунок 2.5</b> — Матрица корреляций соответствующая расстоянию между городами РФ. На графике также показано автоматическое разбиение элементов матрицы на кластеры</p>

