

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ
ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение
высшего образования «Казанский (Приволжский) федеральный
университет»

Набережночелнинский институт (филиал)

**Кафедра Бизнес-информатики и математических методов в
экономике**

Статистика
**Выполнение основных статистических расчетов с
использованием пакета SPSS**

Учебно-методическое пособие

Набережные Челны
2019 г.

УДК 519.252
ББК 65.051

Печатается по решению учебно-методической комиссии экономического отделения Набережночелнинского института (филиала) федерального государственного автономного образовательного учреждения высшего образования «Казанский (Приволжский) федеральный университет», от «26» февраля 2019г. (протокол №6)

Рецензенты:

Доктор физ.-мат. наук, профессор А.Г. Исавнин

Доктор экономических наук, профессор А.Н. Макаров

Розенцвайг А.К. Исавнин А.Г. Статистика. Выполнение основных статистических расчетов с использованием пакета SPSS: учебно-методическое пособие / Розенцвайг А.К. Исавнин А.Г. – Набережные Челны: Изд-во Набережночелнинского института КФУ, 2019. – 24 с.

Учебно-методическое пособие содержит последовательное изложение базовых понятий теории выполнения основных статистических расчетов с использованием пакета SPSS. Пошаговая инструкция запуска и выполнения индивидуальных задач. Построение графиков и диаграмм. Освоение коэффициента корреляции Пирсона. Освоение анализа и данных.

Учебно-методическое пособие предназначено для использования в учебном процессе студентами технических направлений в экономике и экономических направлений дневной, заочной и дистанционной форм обучения.

© Розенцвайг А.К. Исавнин А.Г., 2019

© НЧИ КФУ, 2019

© Кафедра Бизнес-информатики и математических методов в экономике, 2019 г.

SPSS Statistics (аббревиатура англ. «*Statistical Package for the Social Sciences*» — «статистический пакет для социальных наук») — компьютерная программа для статистической обработки данных, один из лидеров рынка в области коммерческих статистических продуктов, предназначенных для проведения прикладных исследований в социальных науках.

Норман Най, Хедли Халл и Дейл Бент разработали первую версию системы в 1968 году, затем этот пакет развивался в рамках Чикагского университета. Первое пользовательское руководство вышло в 1970 году в издательстве McGraw-Hill, а с 1975 года проект выделился в отдельную компанию *SPSS Inc.* Первая версия пакета под Microsoft Windows вышла в 1992 году. На данный момент существуют версии под MacOS X и Linux.

В 2009 году компания SPSS произвела ребрендинг своего статистического пакета, который теперь стал называться PASW Statistics (Predictive Analytics SoftWare). 29 июля 2009 года компания SPSS объявила о том, что она приобретается фирмой IBM.

SPSS является программным продуктом, предназначенным для выполнения всех этапов статистического анализа: от просмотра данных, создания таблиц и вычисления дескриптивных статистик до применения сложных статистических методов. SPSS предлагает совокупность методов и средств, позволяющих обработать данные, и представить результаты в виде, легко воспринимаемом другими. Графика, встроенная в статистические процедуры, равно необходима как для отображения данных, так и для представления результатов.

1. Подготовка данных

На первом этапе проводится процесс подготовки данных. За основу мы возьмем вымышленный опрос - так называемый "воскресный вопрос", который студенты, изучающие политологию в университете, задавали избирателям.

После заполнения анкет, их следует подготовить для ввода данных в компьютер и обработки с помощью программы SPSS. Следует создать кодировочную таблицу. Кодировочная таблица устанавливает соответствие между отдельными вопросам анкеты и переменными, используемыми при компьютерной обработке данных.

Для нашей анкеты мы можем составить следующую кодировочную таблицу. Она приводится в самой анкете.

Предположим, что 20 анкет были заполнены следующим образом:

N	fragebnr	sex	age	party
1	W-001	женский	45	ХДС/ХСС
2	W-002	мужской	22	СДПГ
3	W-003	мужской	19	СДПГ
4	W-004	женский	42	ХДС/ХСС
5	W-005	мужской	34	Зеленые/Союз 90
6	W-006	женский	72	СДПГ
7	W-007	мужской	38	СДПГ
8	W-008	женский	56	СДПГ
9	W-009	мужской	61	ХДС/ХСС

10	W-010	женский	77	ХДС/ХСС
11	W-011	женский	23	Зеленые/Союз 90
12	W-012	мужской	67	Республиканцы
13	W-013	мужской	79	Прочие
14	W-014	женский	26	СДПГ
15	W-015	мужской	59	ХДС/ХСС
16	O-001	женский	34	Зеленые/Союз 90
17	O-002	мужской	18	Республиканцы
18	O-003	женский	44	ХДС/ХСС
19	O-004	мужской	68	ХДС/ХСС

Приведенная выше таблица называется матрицей данных. Данные, предназначенные для обработки в SPSS для Windows, должны быть представлены в виде такой матрицы. Матрица данных состоит из определенного числа строк и столбцов. Строки и столбцы образуют прямоугольную таблицу. При этом каждая строка соответствует одной анкете, а каждый столбец - одной переменной. Каждая строка включает четыре столбца для переменных *fragebnr* (номер вопроса), *sex* (пол), *age* (возраст) и *party* (партия, за которую собирается голосовать избиратель).

Первый шаг при проверке данных обычно состоит в поиске значений, выходящих за разумные пределы значений переменной, - необходимо выяснить, действительно ли это выбросы или это ошибки.

- Используйте процедуру Частоты для подсчета появления каждого отдельного значения (если переменная не имеет сотни значений, как, например, переменная номер страхового полиса). Так вы можете обнаружить опечатки и неожиданные значения. Ищите также пропущенные значения, которые представлены как валидные.

- Для количественных переменных используйте гистограммы в процедуре Частоты или Исследовать, а также ящичковые диаграммы и диаграммы "ствол-лист" в процедуре Исследовать. Обращайте внимание на выбросы, которые показывают диаграммы.

- Для данных большого объема определите максимальные и минимальные значения с помощью процедур *Дескриптивные* или *Средние*. Вы можете обнаружить коды значений, не входящие в заданный набор, или коды пропущенных величин (например, 999), которые трактуются как обычные данные.

- Используйте процедуру *Итоги по наблюдениям* для просмотра данных. Можно выбрать группирующую переменную и разбить данные на содержательно интерпретируемые группы. Для просмотра может оказаться полезной и предварительная сортировка данных по значению некоторой переменной. По умолчанию для просмотра выводятся только первые 100 объектов выборки; вы можете увеличить или уменьшить это число. Чтобы отобрать наблюдения для просмотра, используйте пункт *Отбор наблюдений* в меню *Данные*.

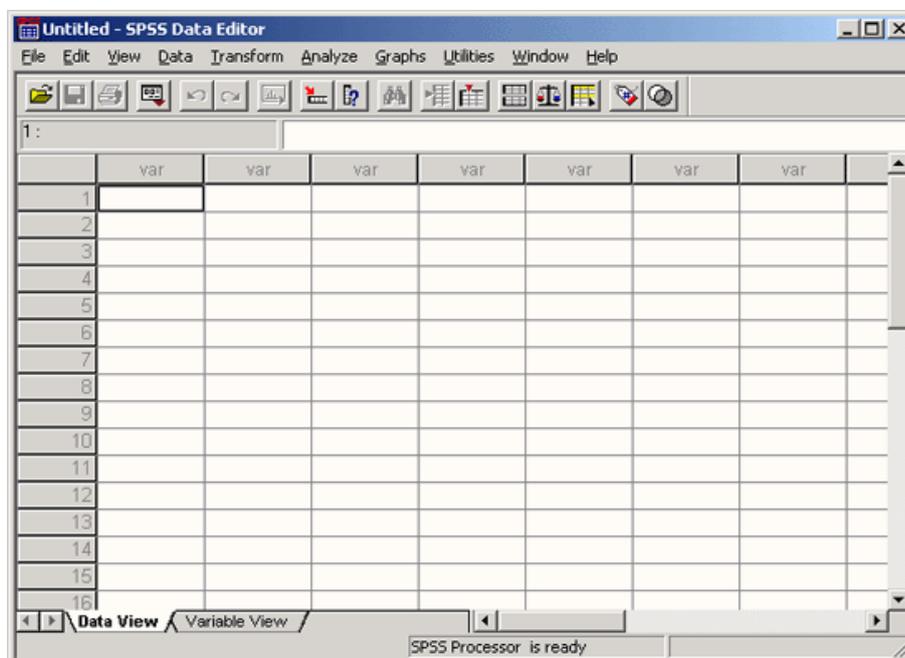
2. Запуск SPSS

Начнем с ввода данных для небольшого примера анализа. Запустите SPSS для Windows, дважды щелкнув левой кнопкой мыши на значке SPSS.

Откроется редактор данных SPSS .

Редактор данных — это одно из многих окон SPSS. Здесь можно вводить новые данные или загружать существующие из файлов данных с помощью команд меню File (Файл) Open... (Открыть...)

Так как при запуске SPSS ни один файл данных еще не загружен, в заголовке редактора данных стоит "Untitled" (Без имени). Над изображением таблицы в редакторе данных имеются строка меню и панель символов. С помощью редактора данных создается файл данных. Редактор данных это приложение, напоминающее электронную таблицу. Под электронной таблицей подразумевается рабочий лист, разделенный на строки и столбцы, который позволяет просто и эффективно вводить данные. При обработке данных наблюдений анкеты в одной переменной хранятся ответы на отдельный вопрос. Отдельные ячейки таблицы содержат значения переменных для каждого отдельного наблюдения; в каждой ячейке хранится одно значение переменной.



Определение переменных

Начнем с определения переменных. Переменную можно определить следующим образом. В редакторе данных дважды щелкните на ячейке с надписью var или щелкните на ярлычке Variable view (Просмотр переменных) на нижнем краю таблицы.

В обоих случаях вы перейдете в режим просмотра переменных, который обеспечивает редактор данных. Здесь мы можем последовательно, строка за строкой определить необходимые переменные.

2.1. Имя переменной

Чтобы задать имя переменной, поступите следующим образом:

- Введите в текстовом поле Name (Имя) выбранное имя переменной. В нашем примере мы сначала определим переменную fragebnr. Для этого введите в поле Name текст "fragebnr".

При выборе имени переменной следует соблюдать определенные правила:

- Имена переменных могут содержать буквы латинского алфавита и цифры. Кроме того, допускаются специальные символы _ (подчеркивание), . (точка), а также символы @ и #. Не разрешаются, например, пробелы, знаки других алфавитов и специальные символы, такие как !, ?, " и *.
- Имя переменной должно начинаться с буквы.
- Последний символ имени не может быть точкой или знаком подчеркивания (_).
- Длина имени переменной не должна превышать восьми символов.
- Имена переменных нечувствительны к регистру, то есть прописные и строчные буквы не различаются.
- Нажмите на клавишу <Tab>, чтобы подтвердить ввод и перейти к установке типа переменной

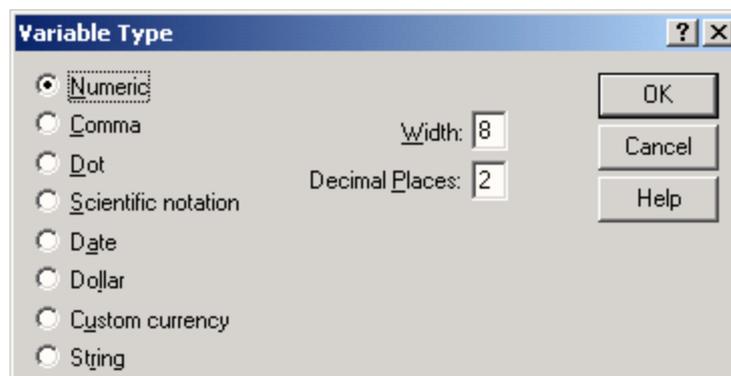
2.2. Тип переменной

Как видно из электронной таблицы, вновь созданные в SPSS переменные по умолчанию являются численными с максимальной длиной восемь знаков, причем дробная часть состоит из двух знаков (формат F8.2).

- Если требуется изменить тип переменной, щелкните в ячейке на кнопке с тремя точками:



Откроется диалоговое окно Define Variable Type (Определение типа переменной).



В SPSS существуют следующие типы переменных:

Numeric (Численный)	К допустимым значениям относятся цифры, перед которыми стоит знак плюс или минус и десятичный разделитель. Знак плюс перед числом, в отличие от минуса, не отображается. В текстовом поле Length (Длина) задается максимальное количество знаков, включая позицию для десятичного разделителя. В текстовом поле Decimals (Десятичные разряды) вводится количество отображаемых знаков дробной части.
Comma (Запятая)	К допустимым значениям относятся цифры, перед которыми стоит знак плюс или минус, точка, как десятичный разделитель и одна или несколько запятых в качестве разделителей групп разрядов. Если запятые опускаются при вводе, они вставляются

	автоматически. Длина такой переменной равна максимальному количеству знаков, включая десятичный разделитель и запятые между группами разрядов.
Dot (Точка)	К допустимым значениям относятся цифры, перед которыми стоит знак плюс или минус, запятая, как десятичный разделитель и одна или несколько точек в качестве разделителей групп разрядов. Если точки опускаются при вводе, они вставляются автоматически.
Scientific notation (Экспоненциальное представление)	При вводе данных разрешаются все допустимые численные значения, включая экспоненциальное представление, о котором свидетельствует содержащаяся в числе буква E или D, а также знак плюс или минус.
Date (Дата)	Допустимые значения — дата и/или время.
Dollar (Доллар)	К допустимым значениям относятся: знак доллара, точка, как десятичный разделитель и запятые, как разделители групп разрядов. Если знак доллара или запятые опускаются при вводе, они вставляются автоматически.
Special currency (Специальная валюта)	Пользователь может задавать собственные форматы валюты. В поле Length в этом случае задается максимальное количество знаков, включая все знаки, заданные пользователем. Обозначение валюты при вводе не указывается; оно вставляется автоматически.
String (Строка)	Строка символов. К допустимым значениям относятся: буквы, цифры и специальные символы. Различаются короткие и длинные строковые переменные. Короткие строковые переменные могут содержать не более восьми знаков. В большинстве процедур SPSS применение длинных строковых переменных ограничивается или вообще не допускается.

При вводе и выводе данных надо учитывать следующие особенности:

- Численные форматы: В численных форматах десятичным разделителем может быть либо точка, либо запятая. Тип десятичного разделителя зависит от настроек диалогового окна Язык и стандарты (Regional Settings) на панели управления Windows. Точное значение переменной хранится внутри программы, а Редактор данных отображает на экране лишь заданное число десятичных разрядов. Значения, которые имеют больше десятичных разрядов, округляются. Для вычислений применяется точное значение.

- Строковые форматы: В длинных строковых переменных значения дополняются пробелами до максимальной длины. Например, в строковой переменной длины 10 значение "SPSS" хранится внутри программы как "SPSS ".

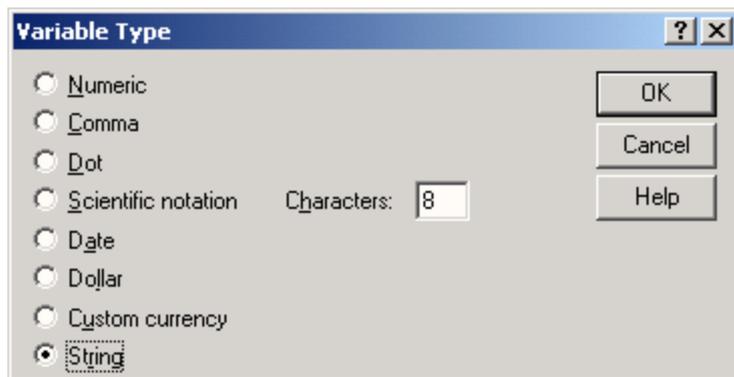
- Форматы даты и времени: В форматах даты в качестве разделителей между значениями дня, месяца и числа могут применяться косая черта, дефис, пробел, запятая или точка. Можно выбрать один из нескольких форматов даты (dd-mm-yyyy, dd-mmm-yy, mm/dd/yyyy и т.д.). Дата в формате dd-mmm-yy отображается с разделителем-дефисом и сокращением названия месяца из трех букв. Дата в

форматах dd/mm/yy и mm/dd/yy отображается с разделителем-косой чертой и номером месяца вместо названия.

- Всего доступно 27 различных форматов даты и времени, которые отображаются в разворачивающемся списке. В форматах времени в качестве разделителей между значениями часов, минут и секунд могут использоваться двоеточие, точка или пробел.

- Специальная валюта: Форматы отображения валюты CCA, CCB, CCC, CCD и CCE задаются с помощью вкладки Currency (Валюта), которая открывается командой меню Edit (Правка) Options... (Параметры...)

- Установите для переменной fragebnr тип String и длину пять символов и щелкните на кнопке ОК.



Переменная fragebnr получила строковый тип. С такими переменными нельзя выполнять никаких вычислительных операций, но можно проводить, например, подсчеты повторяемости. Кроме того, становится возможным ввод букв, например, "W" для старых федеральных земель и "O" — для новых. Мы выбрали длину пять символов, чтобы можно было кодировать до 999 анкет для обеих групп земель. В этом случае для анкет в старых федеральных землях можно будет задавать номера анкет от "W-001" до "W-999", а для новых федеральных земель — от "O-001" до "O-999".

2.3. Сохранение файла данных

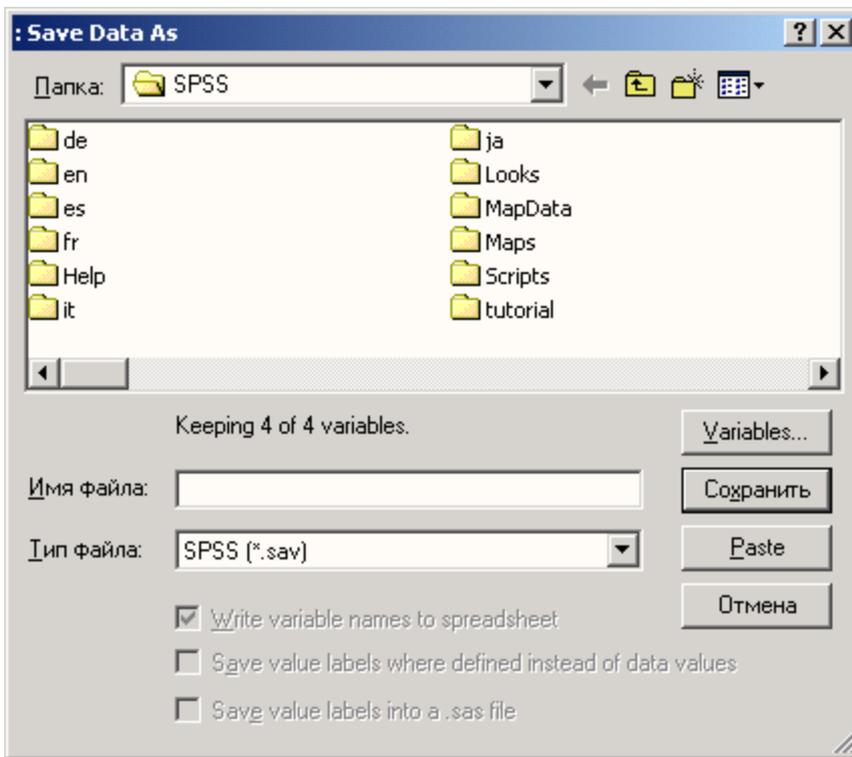
Сейчас мы сохраним созданный файл данных. Поступите следующим образом:

Выберите в меню команды File (Файл) Save as... (Сохранить как...). Откроется диалоговое окно Save Data as (Сохранить данные как).

По умолчанию SPSS сохраняет файл данных в текущем каталоге с расширением .sav.

Если вы следовали указаниям по инсталляции и задали рабочий каталог \SPSSBOOK, он будет предлагаться по умолчанию.

Задайте имя файла, соответствующее соглашению об именах в DOS. Для рассматриваемого примера мы предлагаем имя файла "btwahl.sav". Расширение .sav SPSS присваивает файлам данных по умолчанию, поэтому расширение .sav вводить необязательно.

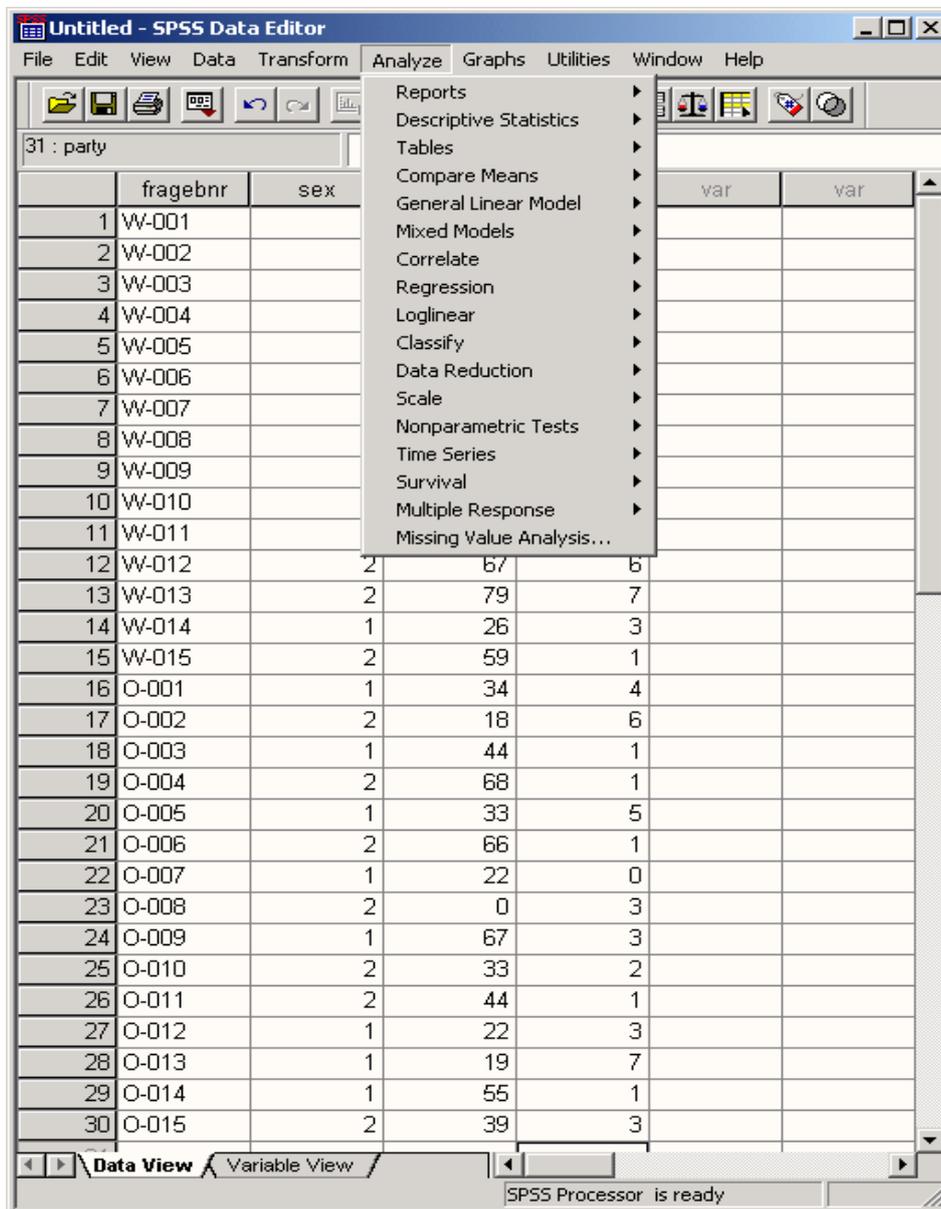


2.4. Выбор статистической процедуры

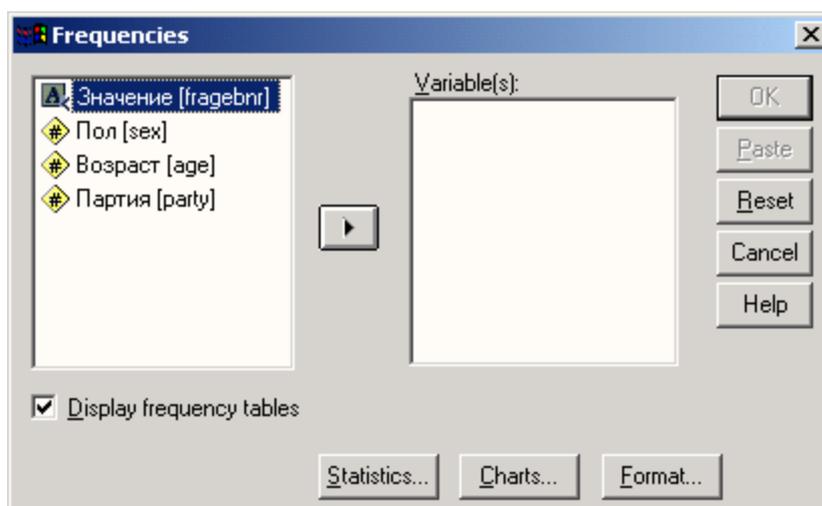
Меню статистики, которое открывается по команде меню Analyze (Анализ), содержит список статистических методов. После каждого пункта этого меню стоит стрелка. Она указывает на существование следующего уровня меню.

В качестве примера попробуем построить частотное распределение. Выполните следующие действия.

Выберите в меню команды Analyze (Анализ) - Descriptive statistics (Описательная статистика) - Frequencies... (Частоты...).



Появится диалоговое окно Frequencies.



Диалоговые окна статистических процедур содержат следующие компоненты:

- Список исходных переменных - список всех переменных в файле данных. В данный момент в списке исходных переменных присутствуют следующие переменные: age, fragebnr, party, sex. Перед именем каждой переменной стоит значок; по которому можно определить, является ли эта переменная численной или строковой.

- Список выбранных переменных - список, содержащий переменные файла данных, которые были выбраны для анализа. Список выбранных переменных также называют целевым списком или списком тестируемых переменных. Этот список имеет заголовок Variable(s) (Переменная(ые)). Так как мы еще не выбрали ни одной переменной, этот список пуст.

- Командные кнопки — кнопки, при щелчке на которые выполняются определенные действия. В этом диалоговом окне расположены кнопки ОК, Paste (Вставить), Reset (Сброс или Отклонить), Cancel (Отмена) и Help (Справка), а также кнопки, открывающие вспомогательные диалоговые окна: Statistics... (Статистика), Charts... (Диаграммы или Графики) и Format... (Формат). Кнопки вспомогательных диалоговых окон отличаются троеточием (...) после названия.

Пять стандартных командных кнопок в главном диалоговом окне имеют следующее назначение:

- ОК — кнопка ОК запускает соответствующую процедуру. Одновременно она закрывает диалоговое окно.

- Paste — эта кнопка переносит выбранный в диалоговом окне синтаксис команды в редактор синтаксиса. Здесь можно отредактировать синтаксис команды и дополнить его другими опциями, недоступными в данном диалоговом окне.

- Reset — эта кнопка отменяет перенос выделенной переменной в целевой список переменных.

- Cancel — эта кнопка отменяет все изменения, сделанные с момента последнего открытия диалогового окна, и закрывает его.

- Help — эта кнопка выводит контекстно-чувствительную справку. При щелчке на ней открывается окно справки, содержащее сведения о текущем диалоговом окне.

2.5. Выбор переменных

Сначала мы построим частотное распределение для переменной party. Выполните следующие действия:

- Выделите переменную party в списке исходных переменных.

- Щелкните на кнопке, которая находится рядом со списком выбранных переменных. Переменная party будет перенесена из списка исходных переменных в список выбранных переменных. Можно также дважды щелкнуть на нужной переменной, и она будет перенесена в список выбранных переменных.

- Подтвердите операцию, щелкнув на кнопке ОК. Результаты будут отображены в окне просмотра (Viewer).

Окно просмотра разделено на две части. В левой отображается структура вывода, а в правой — собственно выводимые данные. В разделе вывода отображаются как таблицы, так и графики.

Вернемся в редактор данных. Это можно сделать двумя различными способами:

- Выберите в меню команды Window (Окно) 1 Wahl.sav — SPSS Data Editor или щелкните на панели инструментов на символе редактора данных 

Statistics

Партия		
N	Valid	29
	Missing	2

Партия

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ХДС/ХСС	10	32,3	34,5	34,5
	СДП	2	6,5	6,9	41,4
	СЕПГ	9	29,0	31,0	72,4
	Зеленые/Союз90	3	9,7	10,3	82,8
	ПДС	1	3,2	3,4	86,2
	Республиканцы	2	6,5	6,9	93,1
	Прочие	2	6,5	6,9	100,0
Total		29	93,5	100,0	
Missing	Нет данных	1	3,2		
	System	1	3,2		
	Total	2	6,5		
Total		31	100,0		

Чтобы для построения частотного распределения выбрать все переменные, содержащиеся в файле данных, выполните следующие действия:

- Щелкните на имени первой переменной и задержите нажатой левую кнопку мыши. Перетащите мышью, пока не будут выделены все переменные.
- Затем, щелкнув на кнопке с треугольником, перенесите переменные в список выбранных переменных.

Для выполнения этой же задачи можно также щелкнуть на первой переменной, а затем, нажав клавишу <Shift>, — на последней переменной (метод "Shift-клик"). Чтобы выделить несколько переменных, которые находятся в разных местах списка, следует поступить следующим образом:

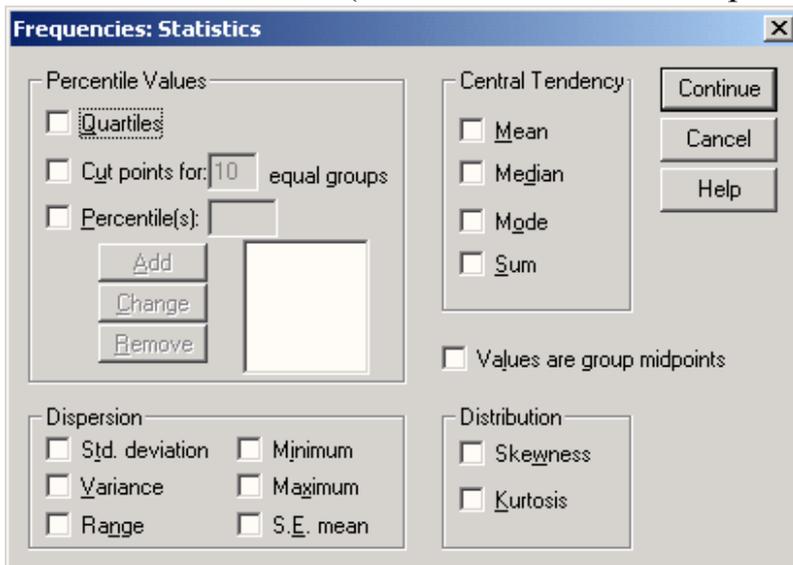
- Щелкните на первой переменной, а затем, при нажатой клавише <Ctrl>, — на следующей и т.д. (метод "Ctrl-клик").

2.6. Вспомогательные диалоговые окна

Сейчас мы попробуем определить наименьшее, наибольшее и среднее значения переменной age.

- Выберите в меню команды Analyze (Анализ) Descriptive statistics (Дескриптивные статистики) Frequencies... (частота распределения)
- В диалоговом окне Frequencies щелкните сначала на кнопке Reset(Сброс). Затем перенесите переменную age в конечный список переменных.

- Щелкните на кнопке Statistics... Откроется диалоговое окно Frequencies: Statistics (Частотное распределение: Статистика)



- Установите флажки Minimum (Наименьшее значение), Maximum (Наибольшее значение) и Average (Среднее значение).
- Щелкните на кнопке Next (Далее). Настройки будут сохранены и мы вернемся в главное диалоговое окно.
- Снимите флажок Display frequency tables (Показывать частотные таблицы).
- Запустите вычисление, щелкнув на кнопке ОК. Результаты будут показаны в окне просмотра:

Статистика

Возраст		
N	Имеется	29
	Отсутствует	1
Среднее		44,28
Наименьшее		18
Наибольшее		79

2.7. Построение и редактирование графиков

Представим в графическом виде значения переменной party (партия).

Выберите в меню Analyze (Анализ) Descriptive Statistics (Дескриптивные статистики) Frequencies... (Частоты)

При помощи кнопки Reset (Сброс) удалите все предыдущие установки.

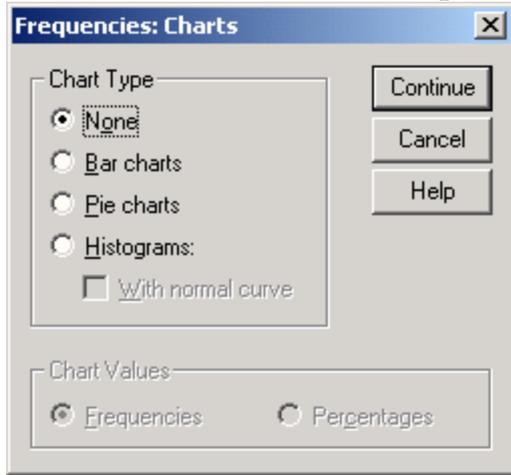
Щелкните дважды на переменной party (партия), чтобы поместить её в список отобранных переменных.

Щелкните на выключателе Charts... (Диаграммы). Откроется диалоговое окно Frequencies: Charts (Частоты: Диаграммы)

Щелкните на опции Bar Charts (Столбчатые диаграммы), в области Chart Values (Значения диаграммы) щелкните на опции Percentages (Проценты) и затем на Continue (Далее).

В главном диалоговом окне деактивируйте опцию Display frequency tables (Показать частотные таблицы).

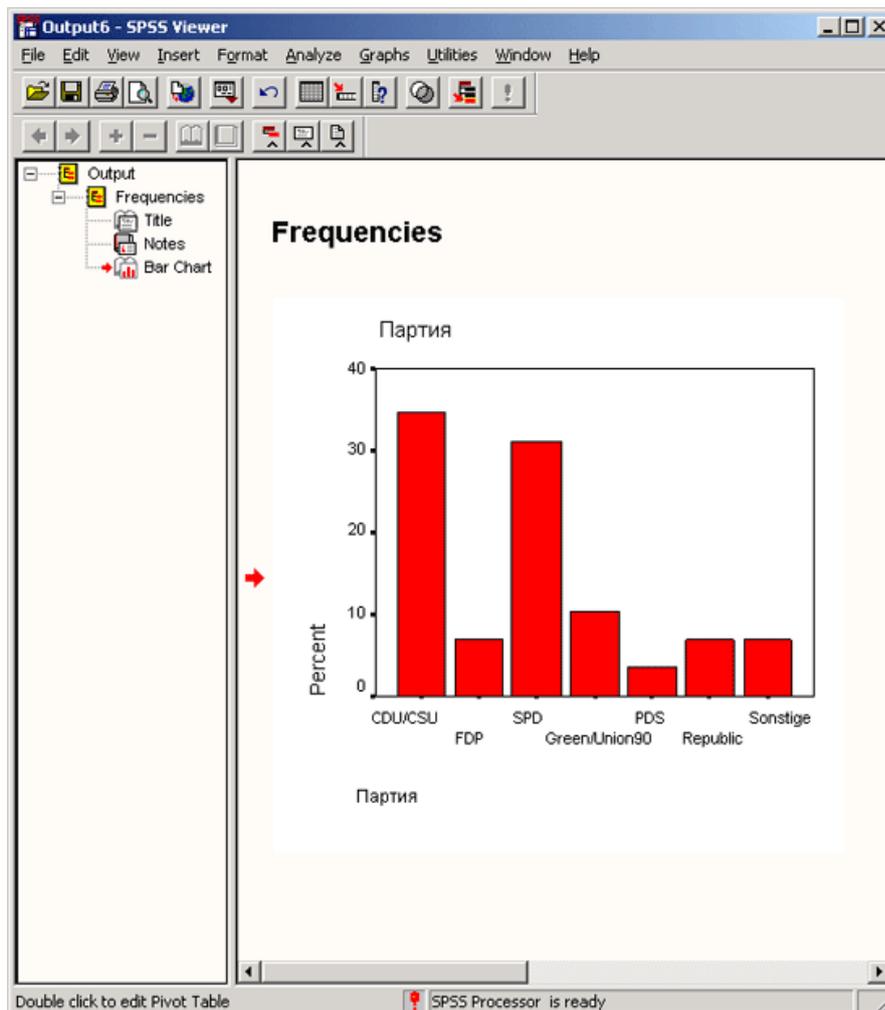
Щёлкните на ОК. В окне просмотра появится столбчатая диаграмма.

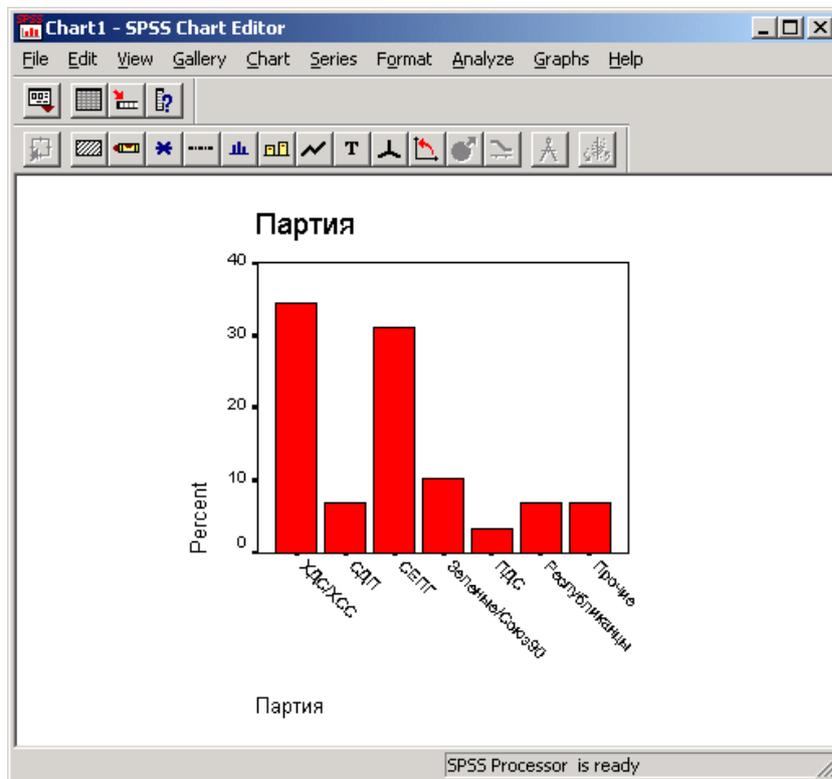


Предположим, у Вас появилось желание отредактировать построенный график в соответствии со своими требованиями.

Щёлкните дважды на какой-либо точке в пределах графика. После этого он будет помещён в редактор диаграмм

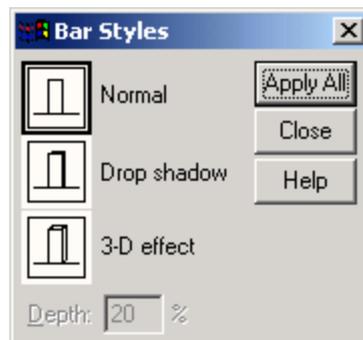
Панель меню изменилась. Теперь в меню присутствуют только опции, предназначенные для обработки графиков. Также претерпели изменения и панели инструментов. Изменим сначала метод представления столбцов. Столбцы, в соответствии с нашим желанием, должны быть представлены в трёхмерном виде.





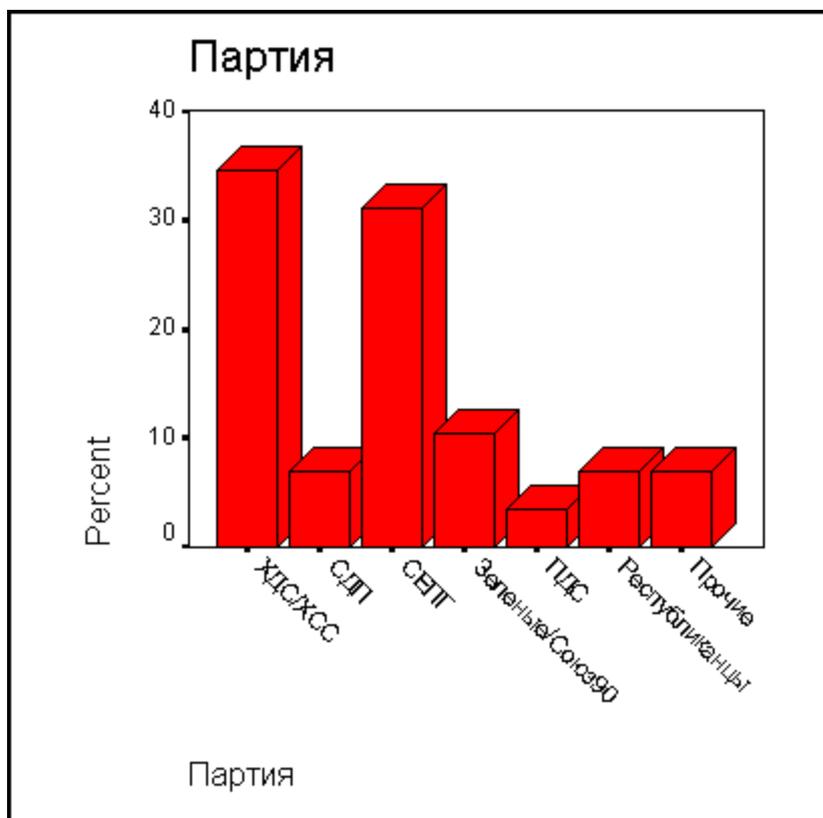
В меню редактора диаграмм выберите Format (Формат) Bar Style... (Вид столбца)

Откроется диалоговое окно Bar Styles (Виды столбцов)



Щёлкните на области 3-D effect (3-D эффект). В поле Depth (Глубина) введите число "40".

Щёлкните Apply All (Применить для всех) и затем на выключателе Close (Закреть). Теперь столбчатая диаграмма выглядит так, как изображено на рисунке.



3. Корреляционно - регрессионный анализ

3.1. Коэффициент корреляции Пирсона

Данный коэффициент вычисляется по следующей формуле:

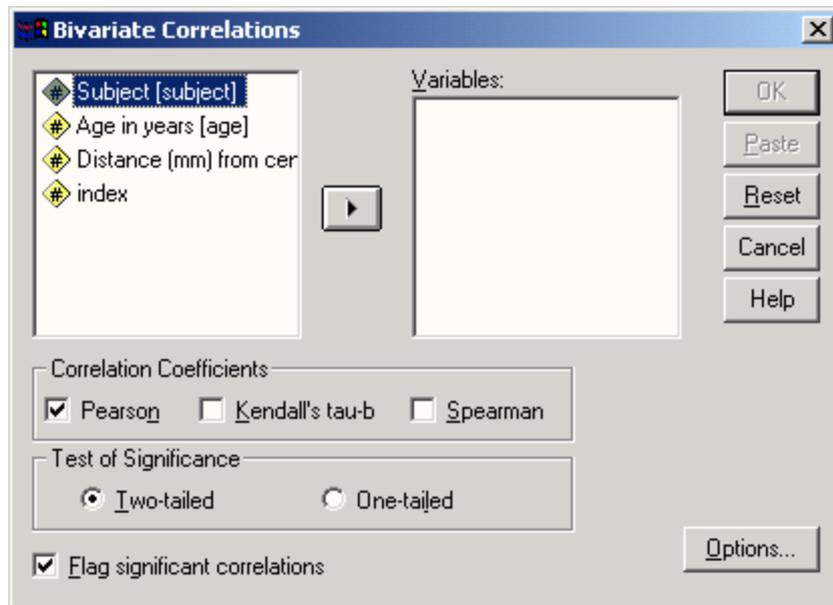
$$r = \frac{\sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})}{n s_x s_y}$$

где x_i и y_i - значения двух переменных, \bar{X} и \bar{Y} их средние значения, а s_x и s_y их стандартные отклонения; n - количество пар значений.

На основании данных исследования гипертонии нам нужно рассчитать коэффициент корреляции по Пирсону попарно для переменных chol0, chol1, chol6 и chol12 (то есть сформировать для этих переменных корреляционную матрицу).

- Откройте файл hyper.sav.
- Выберите в меню Analyze... (Анализ) Correlate...(Корреляция) Bivariate... (Парные) Появится диалоговое окно Bivariate Correlations (Парные корреляции)

- Переменные chol0, chol1, chol6 и chol12 перенесите по очереди в поле тестируемых переменных. Расчёт коэффициента корреляции по Пирсону является предварительной установкой, также как двусторонняя проверка значимости и маркировка значимых корреляций.
- Начните расчёт путём нажатия кнопки ОК.
- В окне просмотра появятся результаты Correlations (Корреляции)



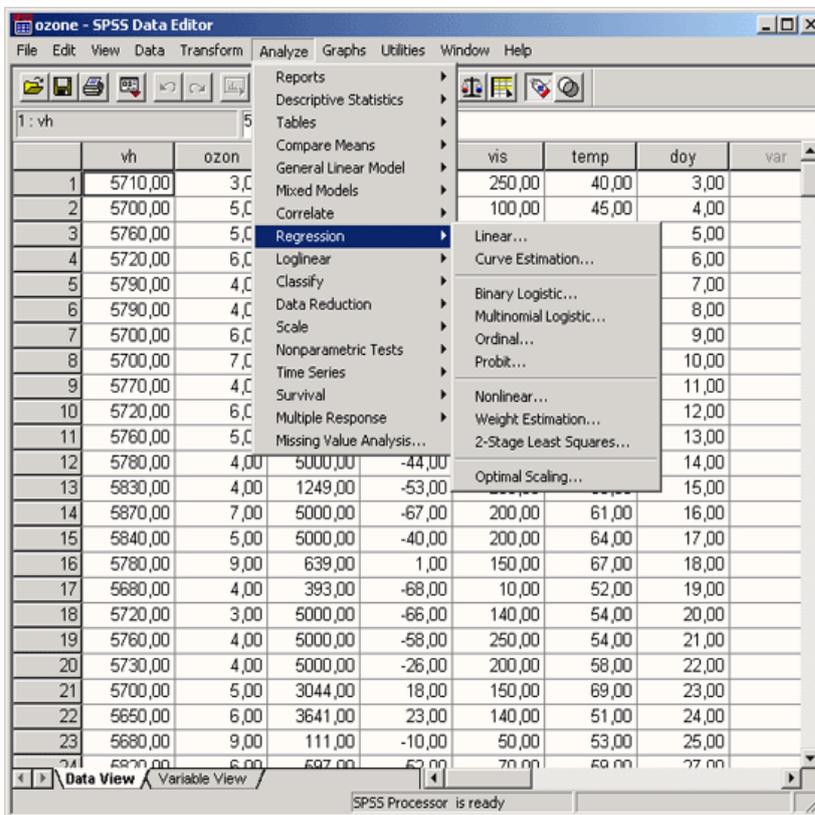
Полученные результаты содержат: корреляционный коэффициент Пирсона r , количество использованных пар значений переменных и вероятность ошибки p , соответствующей предположению о ненулевой корреляции. В приведенном примере присутствует сильная корреляция, поэтому все коэффициенты конечно же являются сверхзначимыми ($p < 0,001$).

При помощи щелчка на кнопке Options... (Опции) можно организовать расчёт среднего значения и стандартного отклонения для двух переменных. Дополнительно могут выводиться отклонения произведений моментов (значений числителя формулы для коэффициента корреляции) и элементы ковариационной матрицы (числитель, делённый на $n - 1$).

Регрессионный анализ служит для определения вида связи между двумя переменными и дает возможность для прогнозирования значения одной (зависимой) переменной, отталкиваясь от значения другой (независимой) переменной.

Чтобы вызвать инструменты для регрессионного анализа в SPSS, выберите в меню Analyze... (Анализ) Regression... (Регрессия).

Откроется соответствующее подменю.



Для проведения линейного регрессионного анализа зависимая переменная должна иметь интервальную (или порядковую) шкалу. В то же время, бинарная логистическая регрессия выявляет зависимость дихотомической переменной от некой другой переменной, относящейся к любой шкале. Если зависимая переменная является категориальной, но имеет более двух категорий, то здесь подходящим методом будет мультиномиальная логистическая регрессия

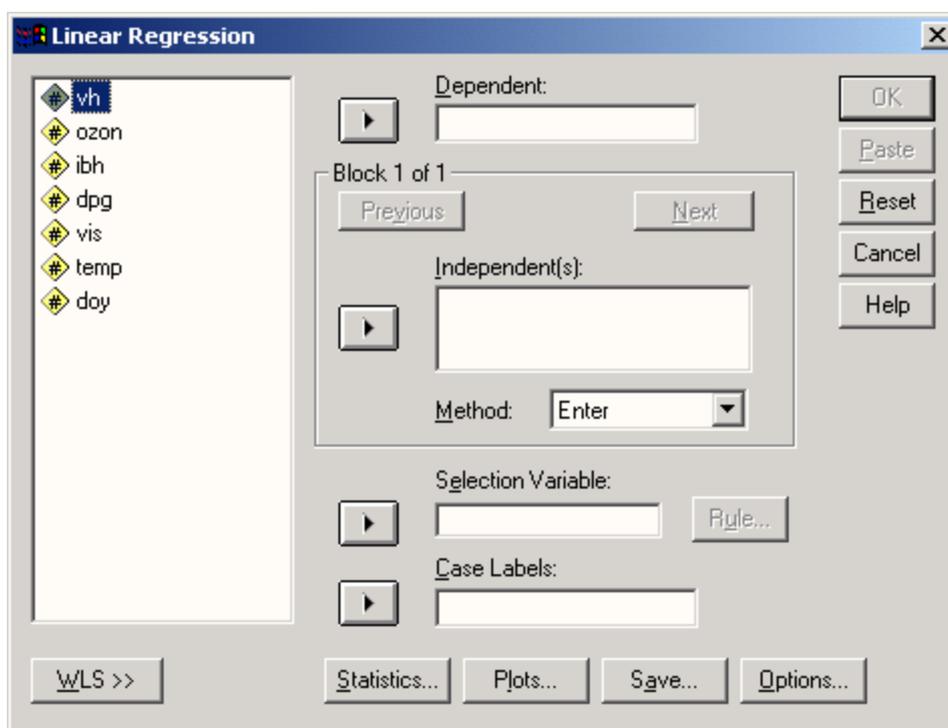
Откройте файл hyper.sav.

- Выберите в меню Analyze... (Анализ) Regression...(Регрессия) Linear... (Линейная) Появится диалоговое окно Linear Regression (Линейная регрессия).
- Перенесите переменную chol1 в поле для зависимых переменных и присвойте переменной chol0 статус независимой переменной.
- Ничего больше не меняя, начните расчёт нажатием ОК.

Вывод основных результатов выглядит следующим образом:

Model Summary (Сводная таблица по модели)

Model (Модель)	R	R Square (R-квадрат)	Adjusted R Square (Смещенный R-квадрат)	Std. Error of the Estimate (Стандартная ошибка оценки)
1	,861	,741	,740	25,26

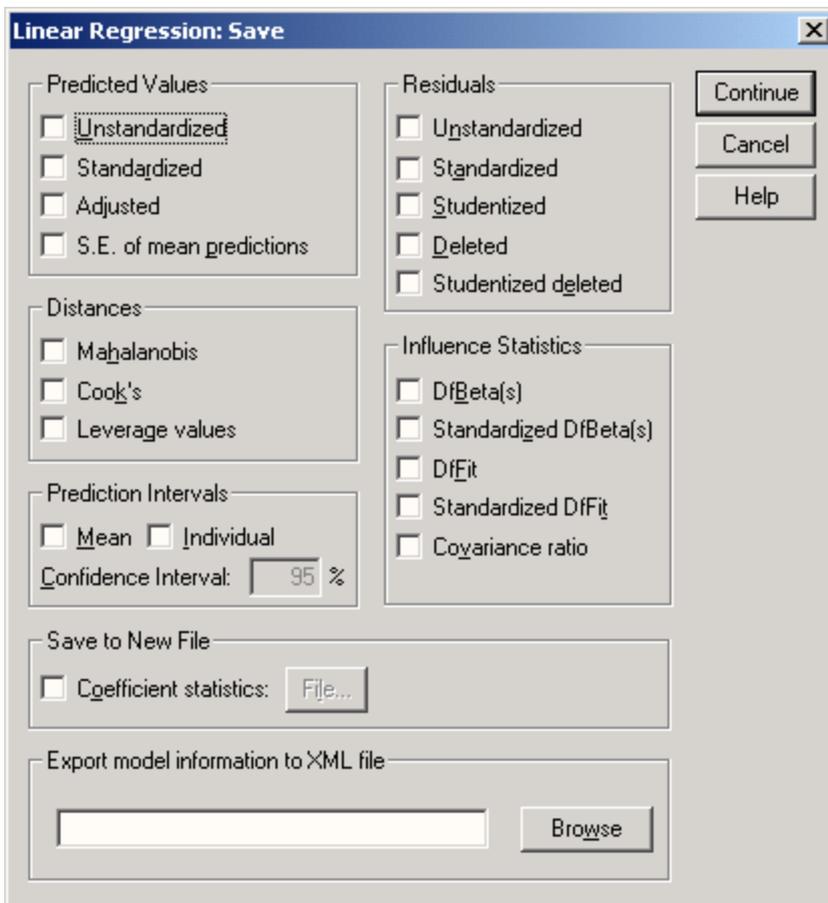


Многочисленные вспомогательные значения, рассчитываемые в ходе построения уравнения регрессии, можно сохранить как переменные и использовать в дальнейших расчётах.

Для этого в диалоговом окне Linear Regression (Линейная регрессия) щёлкните на кнопке Save (Сохранить).

Откроется диалоговое окно Linear Regression: Save (Линейная регрессия: Сохранение).

Интересными здесь представляются опции Standardized (Стандартизированные значения) и Unstandardized (Нестандартизированные значения), которые находятся под рубрикой Predicted values (Прогнозируемые величины). При выборе опции «Нестандартизированные значения» будут рассчитываться значения y , которые соответствуют уравнению регрессии. При выборе опции «Стандартизированные значения» прогнозируемая величина нормализуется. SPSS автоматически присваивает новое имя каждой новообразованной переменной, независимо от того, рассчитываете ли Вы прогнозируемые значения, расстояния, прогнозируемые интервалы, остатки или какие-либо другие важные статистические характеристики. Нестандартизированным значениям SPSS присваивает имена pre_1 (predicted value), pre_2 и т.д., а стандартизированным zpr_1.



Щёлкните в диалоговом окне Linear Regression: Save (Линейная регрессия: Сохранение) в поле Predicted values (Прогнозируемые значения) на опции Unstandardized (Нестандартизированные значения).

Подтвердите нажатием Continue (Далее) и в заключение ОК.

Вы увидите, что в редакторе данных была образована новая переменная под именем `pre_1` и добавлена в конец списка переменных в файле. Для объяснения значений, находящихся в переменной `pre_1`, возьмём случай 5. Для случая 5 переменная `pre_1` содержит нестандартизированное прогнозируемое значение 263,11289. Это прогнозируемое значение слегка отличается в сторону увеличения от реального показателя содержания холестерина, взятого через один месяц (`chol1`) и равного 260. Нестандартизированное прогнозируемое значение для переменной `chol1`, так же как и другие значения переменной `pre_1`, было вычислено исходя из соответствующего уравнения регрессии.

Если мы в уравнение регрессии

$$\text{chol1} = 0,863 \cdot \text{chol0} + 34,546$$

подставим исходное значение для `chol0` (265), то получим $\text{chol1} = 0,863 \cdot 265 + 34,546 = 263,241$

Небольшое отклонение от значения, хранящегося в переменной `pre_1`, объясняется тем, что SPSS использует в расчётах более точные значения, чем те, которые выводятся в окне просмотра результатов. На этом этапе мы ещё раз проиллюстрируем возможность использования регрессии в качестве прогноза.

Добавьте для этого в конец файла `hyper.sav`, ещё два случая, используя фиктивные значения для переменной `chol0`. Пусть, к примеру, это будут значения 282 и 314.

Мы исходим из того, что нам неизвестны значения показателя холестерина через месяц после начала лечения, и мы хотим спрогнозировать значение переменной chol1.

Оставьте предыдущие установки без изменений и проведите новый расчёт уравнения регрессии.

В конце списка переменных добавится переменная рге_2. Для нового добавленного случая (№175) для переменной chol1 будет предсказано значение 277,77567, а для случая №176 — значение 305,37620.

3.2. Построение регрессионной прямой

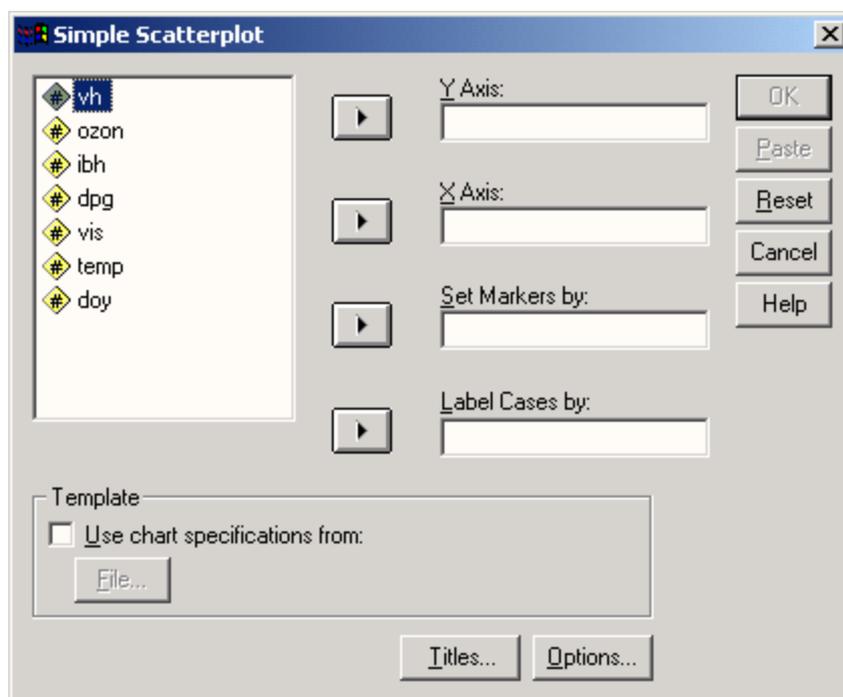
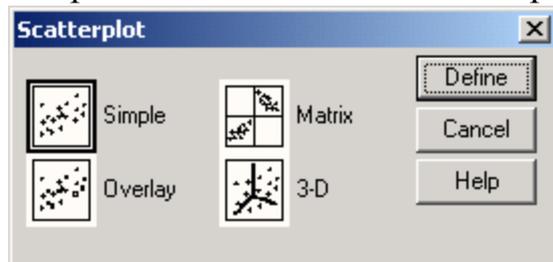
Чтобы на диаграмме рассеяния изобразить регрессионную прямую, поступите следующим образом:

Выберите в меню следующие опции: Graphs ... (Графики) Scatter plots... (Диаграммы рассеяния).

Откроется диалоговое окно Scatter plots... (Диаграммы рассеяния).

В диалоговом окне Scatter plots... (Диаграмма рассеяния) оставьте предварительную установку Simple (Простая) и щёлкните на кнопке Define (Определить).

Откроется диалоговое окно Simple Scatter plot (Простая диаграмма рассеяния).



Перенесите переменную chol1 в поле оси Y, а переменную chol0 в поле оси X. Подтвердите щелчком на ОК. В окне просмотра результатов появится диаграмма рассеяния. Щёлкните дважды на этом графике, чтобы перенести его в редактор диаграмм. Выберите в редакторе диаграмм меню Chart... (Диаграмма) Options...

(Опции). Откроется диалоговое окно Scatterplot Options (Опции для диаграммы рассеяния).

В рубрике Fit Line (Приближенная кривая) поставьте флажок напротив опции Total (Целиком для всего файла данных) и щёлкните на кнопке Fit Options (Опции для приближения). Откроется диалоговое окно Scatterplot Options: Fit Line (Опции для диаграммы рассеяния: приближенная кривая)

Подтвердите предварительную установку Linear Regression (Линейная регрессия) щелчком Continue (Далее) и затем на ОК.

Закройте редактор диаграмм и щёлкните один раз где-нибудь вне графика.

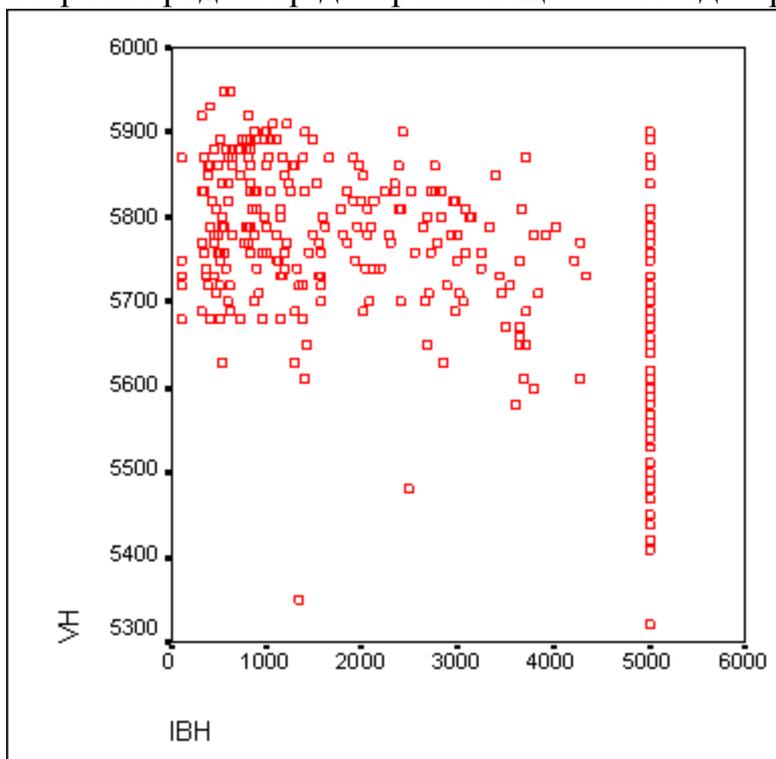
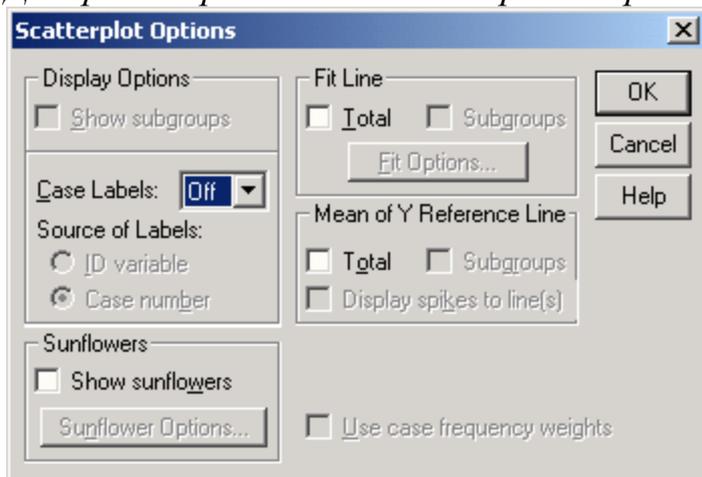
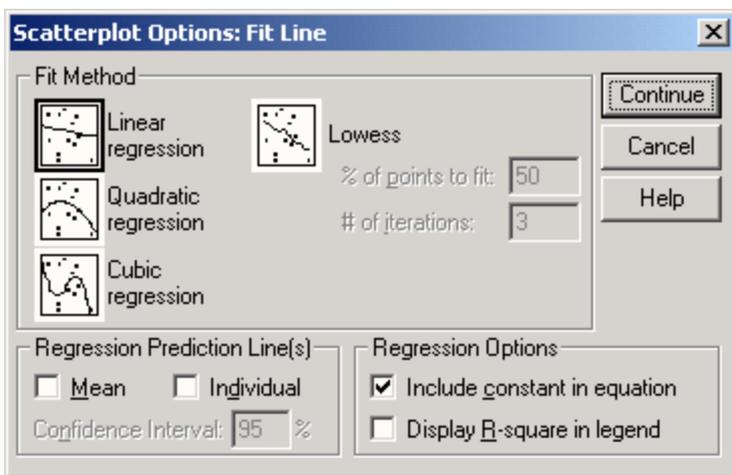


Диаграмма рассеяния в окне просмотра.

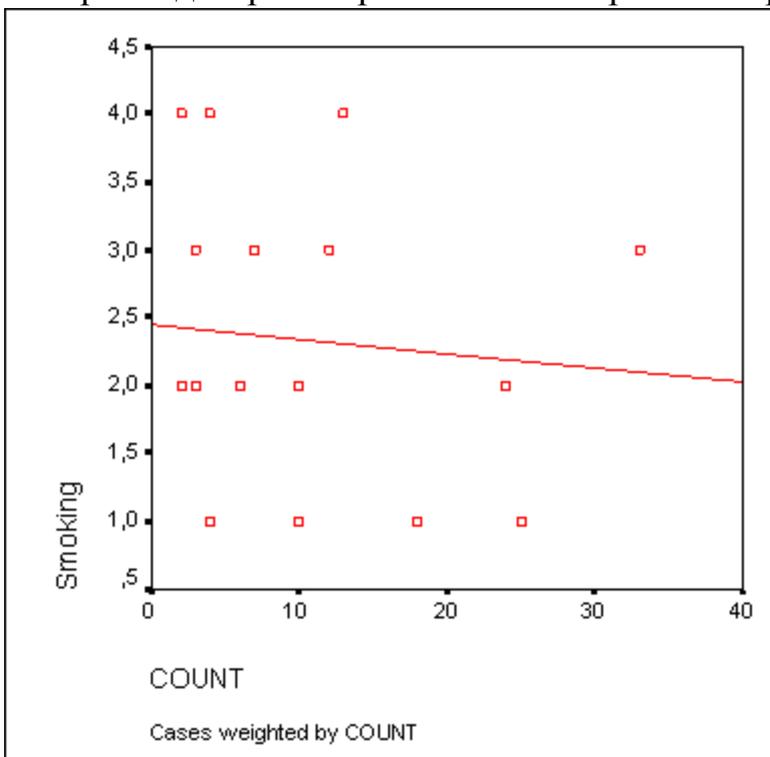


Диалоговое окно Scatterplot Options (Опции для диаграммы рассеяния).



Диалоговое окно *Scatterplot Options: Fit Line* (Опции для диаграммы рассеяния: приближенная кривая).

Теперь на диаграмме рассеяния отображается регрессионная прямая.



4. Список использованных источников

1. <http://ru.wikipedia.org/wiki/SPSS>
2. Тюрин Ю. Н. Анализ данных на компьютере [Текст] : учебное пособие для вузов / Ю. Н. Тюрин, А. А. Макаров .— 4-е изд., перераб .— Москва : ФОРУМ, 2008 .— 368 с : ил..

Розенцвайг А.К Исавнин А.Г

Статистика

Выполнение основных статистических расчетов с использованием пакета
SPSS

Учебно-методическое пособие

Подписано в печать 22.04.2019.

Формат 60x84/16. Печать ризографическая.

Бумага офсетная. Гарнитура «Times New Roman».

Усл.п.л. 1,5 Уч.-изд. л. 1,44

Тираж 100 экз. Заказ № 1246

Отпечатано в Издательско-полиграфическом центре
Набережночелнинского института
Казанского (Приволжского) федерального университета

423810, г. Набережные Челны, Новый город, пр.Мира, 68/19
тел./факс (8552) 39-65-99 e-mail: ic-nchi-kpfu@mail.ru