

## SPSS Statistics быстрый старт

Основным инструментом анализа и визуализации статистических данных для меня всегда был Excel. Я работаю с ним ежедневно. По нему написал больше всего заметок и прочитал наибольшее число книг. Пожалуй, лучшее сочетание статистики и Excel я нашел в книге [Левин. Статистика для менеджеров с использованием Microsoft Excel](#). Вторым инструментом, к которому я только прикоснулся, был R (см., например, [Алексей Шипунов. Наглядная статистика. Используем R!](#)). А недавно прочитал любопытную книгу Нил Дж. Салкинд. Статистика для тех, кто (думает, что) ненавидит статистику. В ней автор все примеры иллюстрирует в программе SPSS. Так что я решил попробовать и этот продукт.

На [сайте IBM](#) доступна пробная версия, которая будет работать на вашем ПК 14 дней. Регистрируетесь и скачиваете программу SPSS Statistics. При регистрации запомните пароль. Он вам пригодится для входа в программу. После запуска появляется приветственное окно:

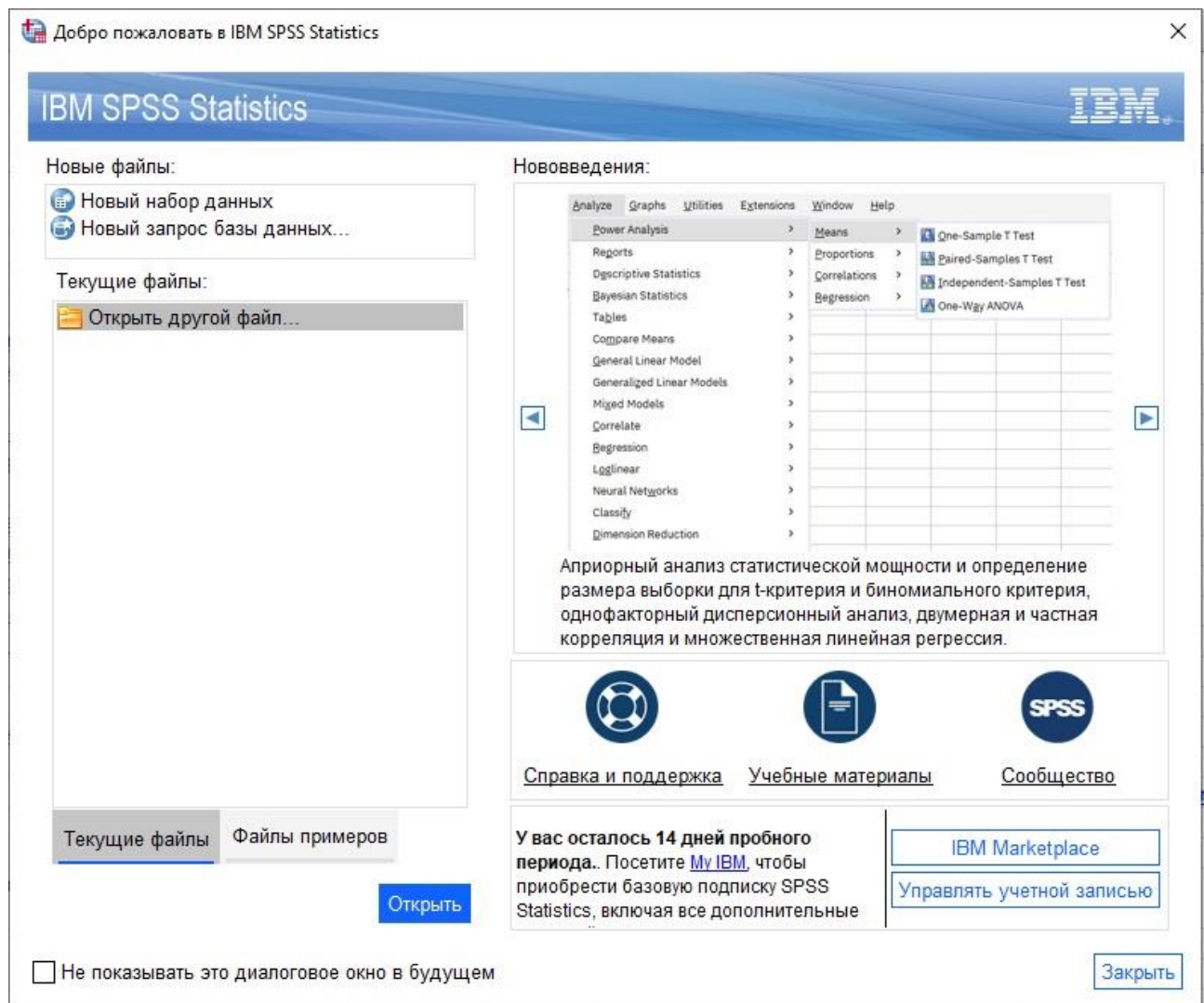


Рис. 1. Приветственное окно SPSS

Если вы не хотите видеть этот экран каждый раз при запуске SPSS, то в левом нижнем углу окна кликните *Не показывать это диалоговое окно в будущем*.

Кликните *Закреть* в правом нижнем углу экрана. Появится окно *Редактор данных*. По виду и функционалу *Редактор* похож на электронную таблицу, как лист Excel.

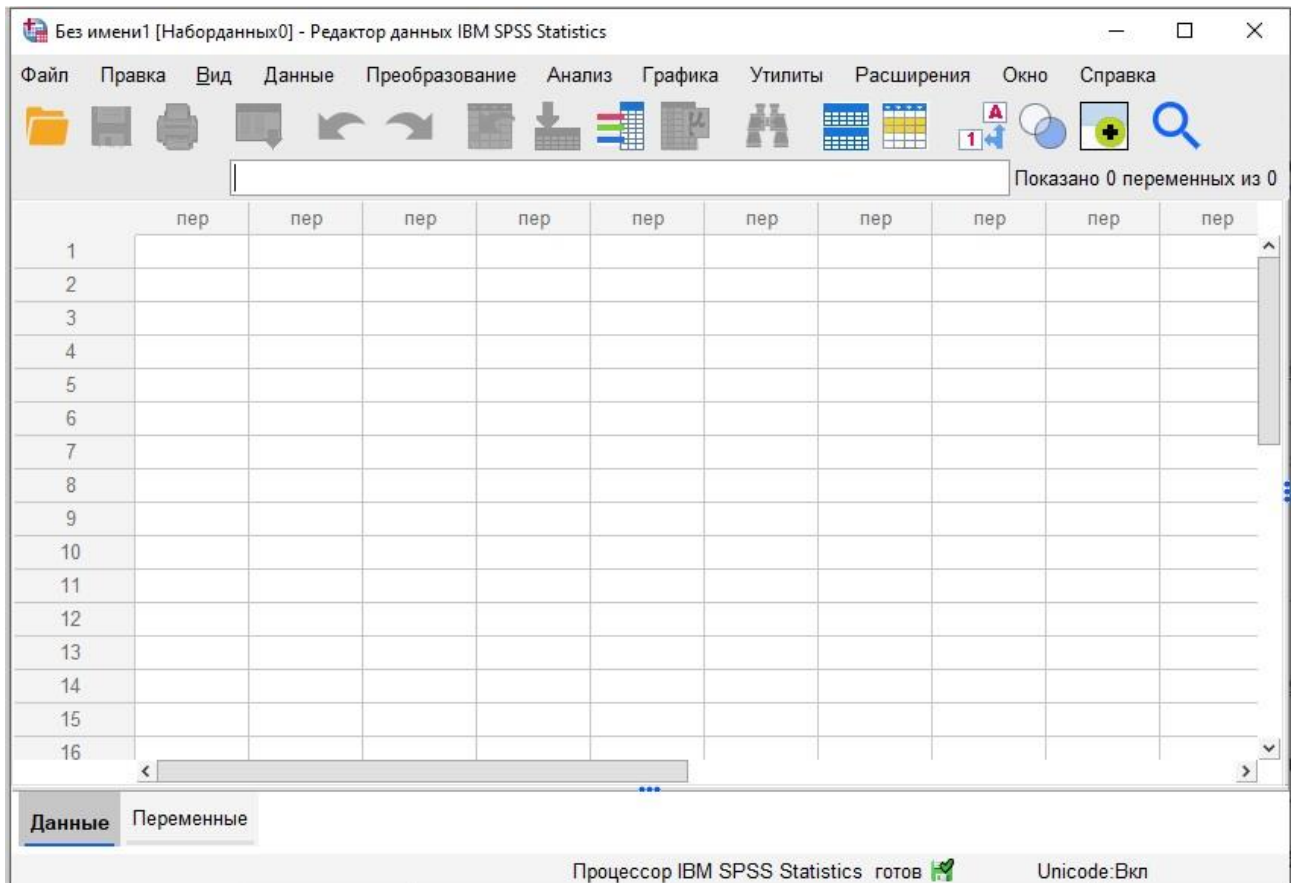


Рис. 2. Редактор данных

Хотя этого и не видно, когда SPSS открывается в первый раз, но есть еще одно открытое (хотя и неактивное) окно. Это *Окно вывода* (Viewer). Оно показывает создаваемые вами статистические результаты и графики. Набор данных создается при помощи *Редактора данных*, а после анализа или построения графиков вы изучаете результаты анализа в *Окне вывода*.

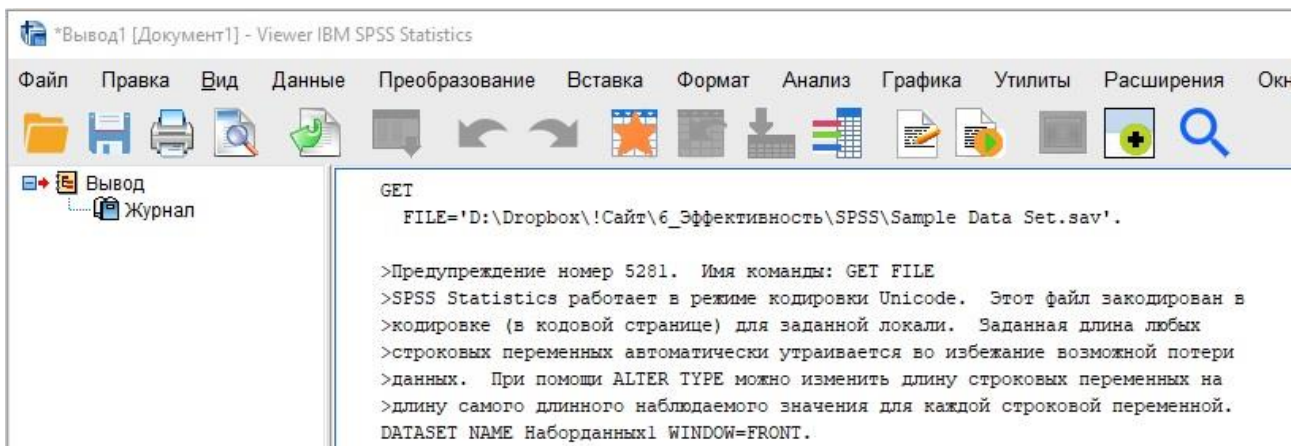


Рис. 3. Окно вывода

#### Панель инструментов и строка состояния

Если вы хотите узнать, что делает иконка на панели инструментов, просто наведите на нее указатель мыши. Некоторые кнопки на панели инструментов затенены. Это означает, что они не активны.

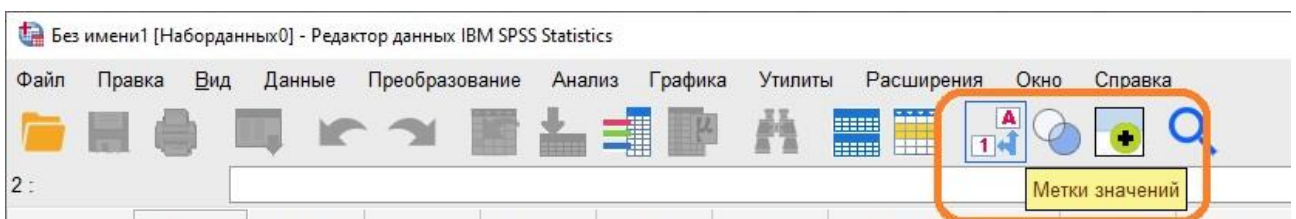


Рис. 4. Панель инструментов

В нижней части окна расположена *Строка состояния*. Она показывает, какие действия выполняет SPSS. Например, сообщение «Процессор IBM SPSS Statistics готов» говорит о том, что SPSS готова к вашим указаниям или вводу данных.

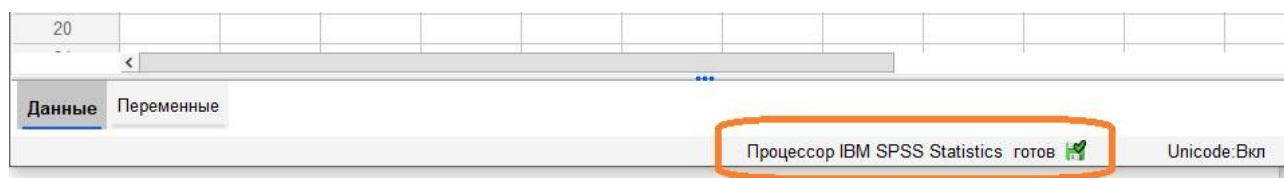


Рис. 5. Строка состояния

### Использование справки

Справка настолько подробна, что может указать вам путь, даже если вы новичок в работе с программой. Меню *Справка* содержит 10 разделов.

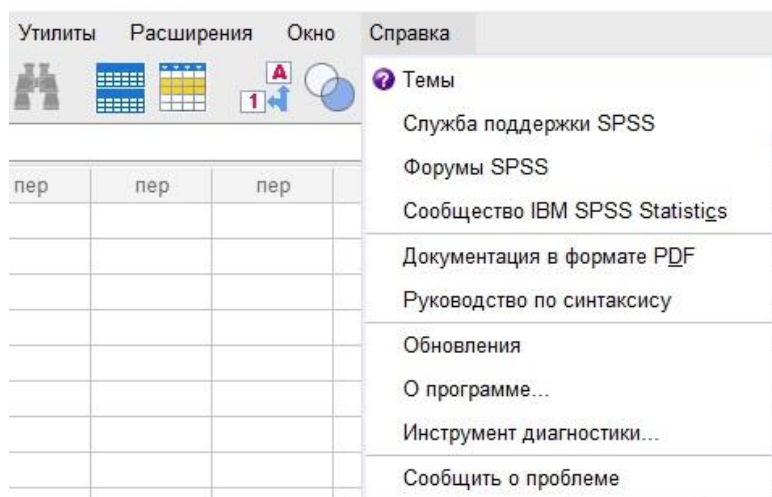


Рис. 6. Меню справки

Нажмите *Темы*, и перейдете в браузер на [страницу центра знаний IBM](#) на русском языке. Здесь представлена собственно справка, а также Учебное пособие, Разбор конкретных случаев, Инструктор по статистике, Разделы для подключаемых модулей Python и R.

### Открытие файла

Вы можете импортировать данные из Excel, или ввести значения в таблицу, после чего сохранить в новом файле SPSS, или открыть готовый файл. В этой заметке мы используем файл Sample Data Set.sav. Пройдите по меню *Файл* → *Открыть* → *Данные*. Выберите файл. Данные загрузятся в окно редактора:

Sample Data Set.sav [Наборданных1] - Редактор данных IBM SPSS Statistics

Файл Плавка Вид Данные Преобразование Анализ Графика Утилиты Расширения Окно Справка

Показано 5 переменных из 5

	ID	Gender	Treatment	Test1	Test2	пер	пер	пер	пер
1	1	1	1	98	32				
2	2	2	2	87	33				
3	3	2	1	89	54				
4	4	2	1	88	44				
5	5	1	2	76	64				
6	6	1	1	68	54				
7	7	2	1	78	44				
8	8	2	2	98	32				
9	9	2	2	93	64				
10	10	1	2	76	37				
11	11	2	1	75	43				
12	12	2	1	65	56				
13	13	1	1	76	78				
14	14	2	1	78	99				
15	15	2	1	89	87				
16	16	2	2	81	56				
17	17	1	1	78	78				
18	18	2	1	83	56				
19	19	1	1	88	67				
20	20	2	1	90	88				
21	21	1	1	93	81				
22	22	1	2	89	93				
23	23	2	2	86	87				
24	24	1	1	77	80				
25	25	1	1	89	99				
26									
27									

Данные Переменные

Процессор IBM SPSS Statistics готов Unicode:Вкл

Рис. 7. Данные загружены из файла в окно редактора

### Таблица и диаграмма

Допустим, мы хотим посчитать, сколько мужчин и женщин находится в нашей выборке, и вывести результат в виде столбчатой диаграммы. В окне *Редактора данных* пройдите по меню *Анализ* → *Описательные статистики* → *Частоты*. В открывшемся окне *Частоты*, выберите *Gender*, нажмите кнопку для переноса переменной в правое окно (или дважды кликните на *Gender*). Нажмите кнопку *Диаграммы*. Выберите *Столбчатые*. Нажмите *Продолжить*. Нажмите *Ok*.

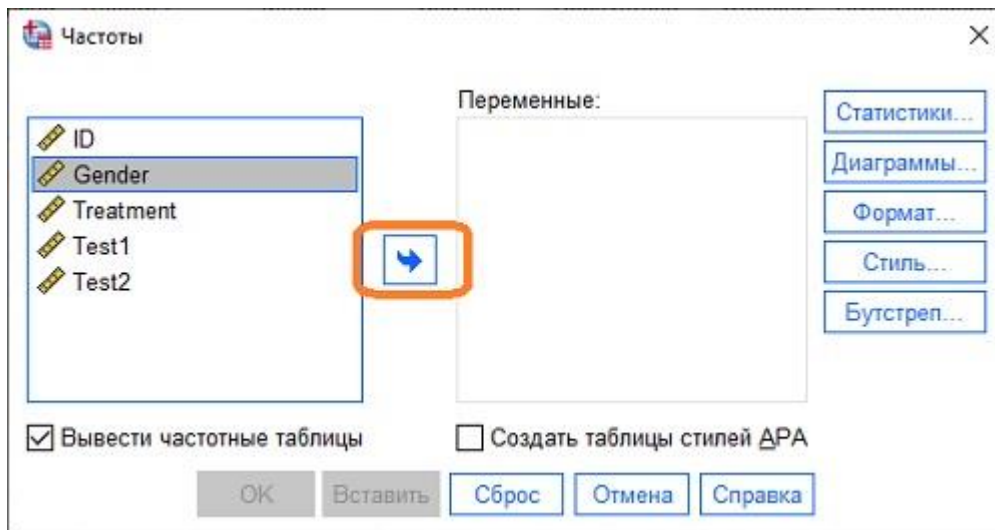


Рис. 8. Выбор переменной для анализа частоты

В окне вывода появится таблица и диаграмма:

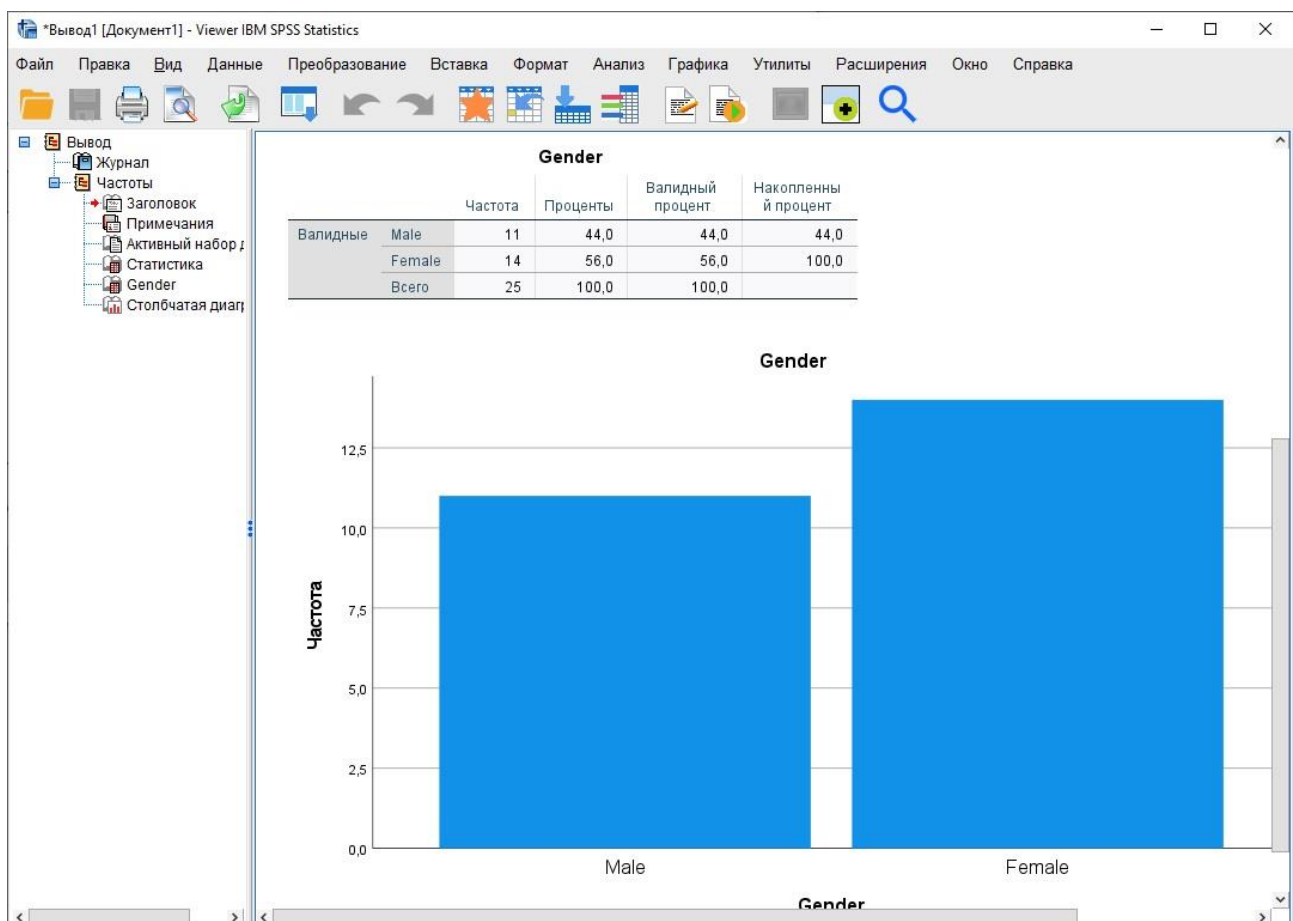


Рис. 9. Таблица и диаграмма в окне вывода

### Оценка t-критерия

Давайте проверим, отличаются ли средние значения результатов Test 1 у мужчин и женщин. Этот анализ основан на t-критерии для независимых выборок. В редакторе данных пройдите по меню *Анализ* → *Сравнение средних* → *T-критерий для независимых выборок*. В открывшемся окне переместите переменную *Test1* в область *Проверяемые параметры*, а переменную *Gender* – в область *Группировать по*:

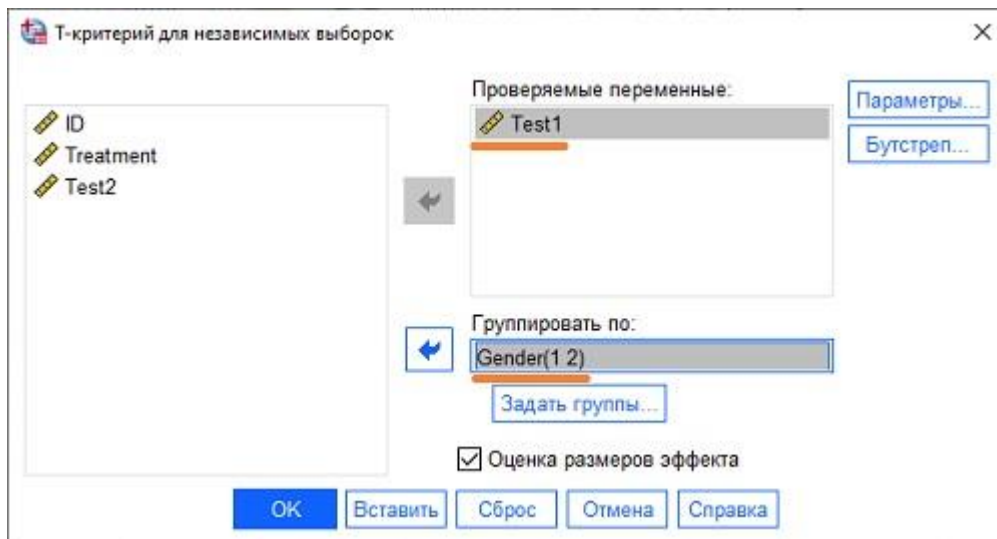


Рис. 10. Настройка расчета t-критерия для независимых выборок

Нажмите *Ok*. Программа сформирует таблицы проверки по t-критерию, и покажет их в окне вывода:

**Статистика группы**

Gender	N	Среднее	Стандартная отклонения	Стандартная средняя ошибка
Test1 Male	11	82,55	9,234	2,784
Test1 Female	14	84,29	8,389	2,242

**Критерий для независимых выборок**

		Критерий равенства дисперсий Ливина		t-критерий для равенства средних						
		F	знач.	T	ст.св.	знач. (двухсторонняя)	Средняя разность	Среднеквадратичная ошибка разности	95% доверительный интервал для разности	
									Нижняя	Верхняя
Test1	Предполагаются равные дисперсии	,714	,407	-,493	23	,627	-1,740	3,532	-9,047	5,566
	Не предполагаются равные дисперсии			-,487	20,532	,632	-1,740	3,575	-9,185	5,704

**Размеры эффектов независимых выборок**

		Стандартиза тор <sup>a</sup>	Точечная оценка		
			95% доверительный интервал		
			Нижняя	Верхняя	
Test1	d Козна	8,766	-,199	-,988	,595
	Коррекция Хеджеса	9,066	-,192	-,955	,576
	Дельта Гласса	8,389	-,207	-,997	,590

a. Этот знаменатель используется при оценке размеров эффектов.  
d Козна использует среднеквадратичное отклонение пула.  
Коррекция Хеджеса использует среднеквадратичное отклонение пула плюс фактор коррекции.  
Дельта Гласса использует среднеквадратичное отклонение выборки для контрольной группы.

Рис. 11. Результаты проверки t-критерия для независимых выборок

T-тест показал, что различие между мужчинами и женщинами при прохождении Test1 незначимо.

### Создание и редактирование файла данных

Давайте создадим набор данных, который только что загрузили из файла Sample Data Set.sav.

Сначала определим переменные, а затем введем данные. В окне *Редактора данных* пройдите по меню *Файл* → *Создать* → *Данные*. Откроется новое окно *Редактора данных*. Обратите внимание, что окно открылось на вкладке *Переменные* (SPSS подсказывает, что сначала надо заняться ими).

Если вы поместите курсор в первую ячейку в колонке *Имя*, введёте любое имя и нажмёте *Enter*, SPSS для всех характеристик переменной автоматически проставит значения по умолчанию (см. строку 1 на рис. ниже).

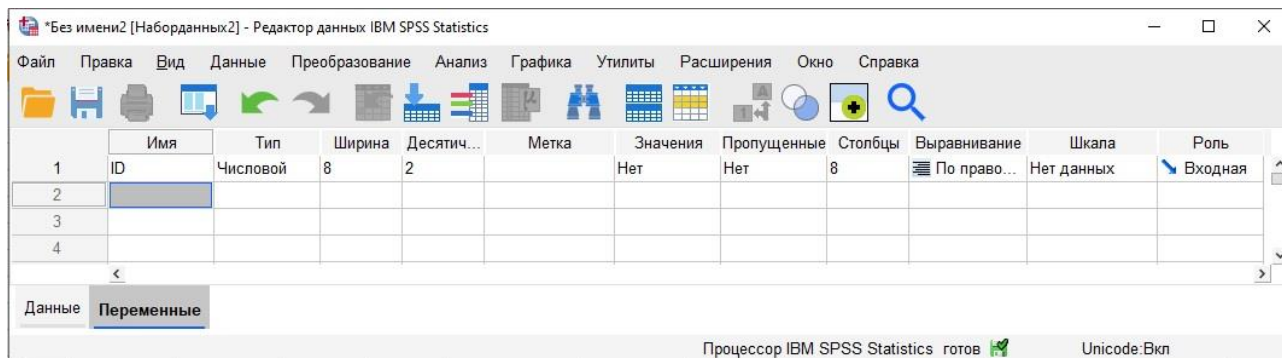


Рис. 12. Параметры по умолчанию

Подробнее о параметрах переменной:

- *Имя* переменной должно начинаться с буквы, иметь длину не более 64 символов, не содержать пробелы, подробнее см. [здесь](#);
- *Тип*:

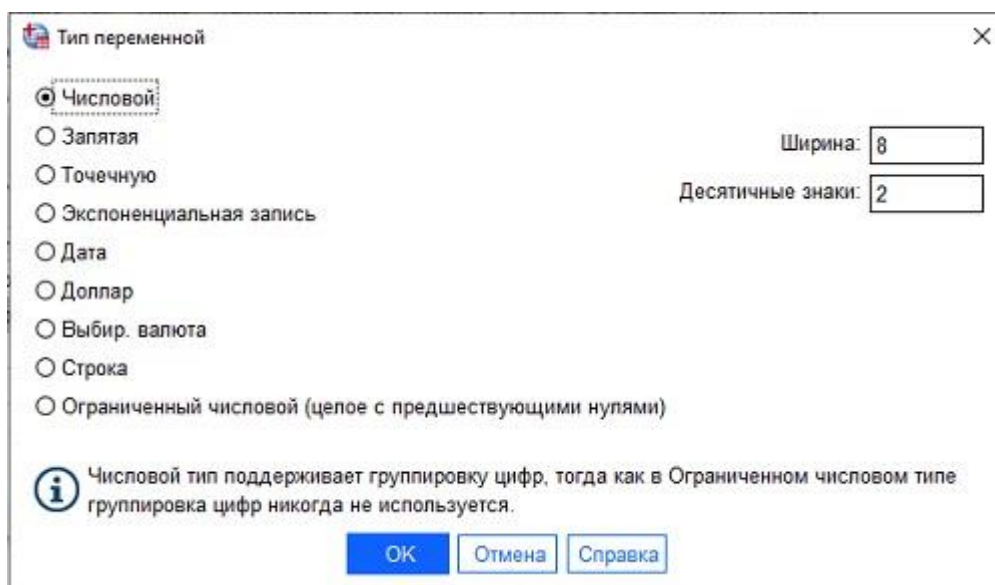


Рис. 13. Типы переменных

- *Ширина* задает количество символов в столбце, содержащем данную переменную;
- *Десятичные* определяет количество десятичных знаков;
- *Метка* задает метку переменной длиной до 256 символов;
- *Значения* устанавливает соответствие числовых значений и категорий; например, 1 для мужчин и 2 для женщин;

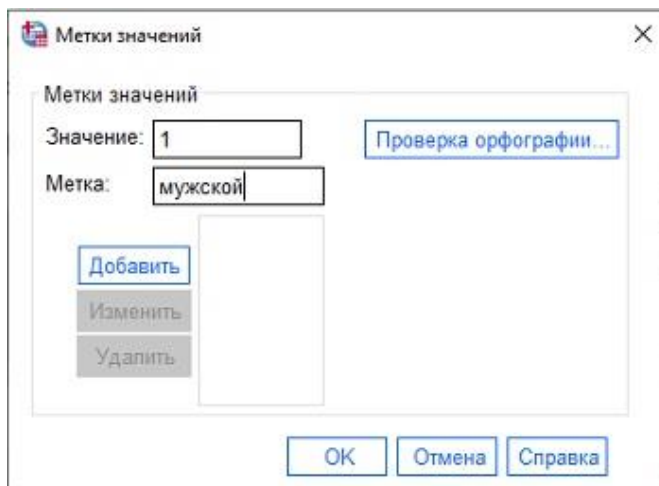


Рис. 14. Введение значений для категориальных переменных

Категорийные переменные в SPSS можно вводить в виде текстовых строк, например, Male и Female, а можно назначить им значения. Когда дело дойдет до анализа, окажется, что очень трудно работать с нечисловыми записями. Но при визуальном просмотре файла, наоборот, имена (метки) нагляднее. Окно *Метки значений* позволяет вводить в SPSS числа, а выводить на экран метки: и волки сыты, и овцы целы. Чтобы это работало, находясь в *Редакторе данных* на закладке *Данные*, перейдите в меню *Вид*, и поставьте галочку напротив *Метки значений*.

- *Пропущенные* – указывает, как обращаться с пропущенными значениями;

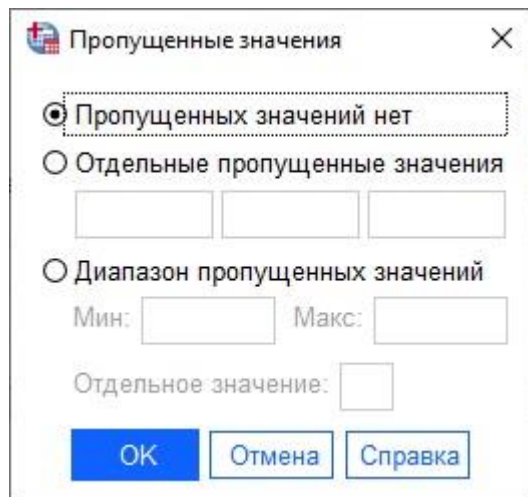


Рис. 15. Управление пропущенными значениями

- *Ширина столбца* определяет количество символов, выделенное для переменной в окне представления данных;
- *Выравнивание* – определяет, как будут выровнены данные в ячейке (влево, вправо, по центру);
- *Мера* – определяет шкалу измерения, которая лучше всего описывает переменную (номинальная, порядковая или интервальная);
- *Роль* – определяет роль, которую играет переменная в анализе (входная, целевая и т.д.).

Определите следующие переменные:

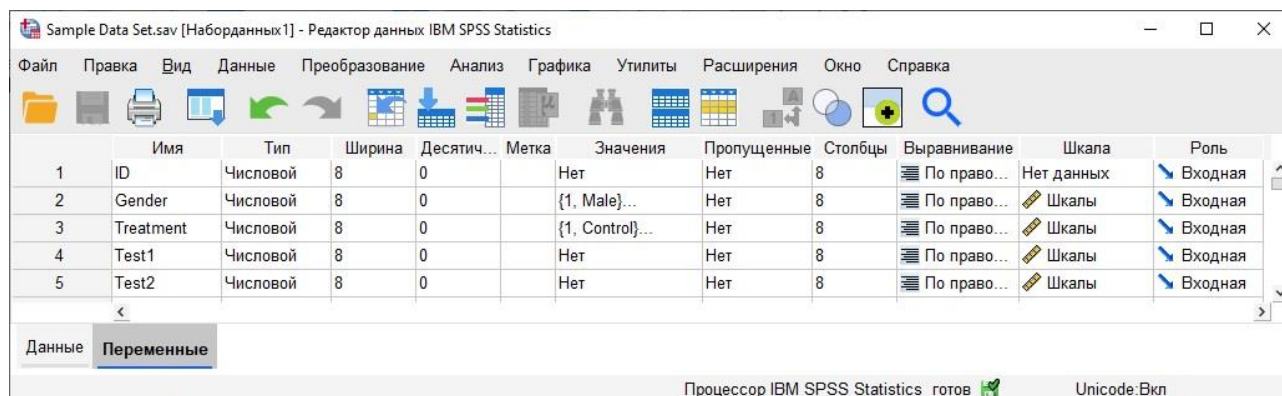


Рис. 16. Пять переменных в окне *Редактора данных* на закладке *Переменные*

Теперь вы можете переключиться на вкладку *Данные* и просто ввести все данные, которые представлены на рис. 7.

### Печать из SPSS

Чтобы распечатать весь файл данных или его часть:

- 1) убедитесь, что файл, который вы хотите напечатать, находится в активном окне;
- 2) кликните *Файл* → *Печать*;
- 3) откроется диалоговое окно печати (рис. 17);
- 4) выберите, что вы хотите распечатать: файл целиком или выделенный фрагмент (если предварительно фрагмент не был выбран, эта опция неактивна); нажмите *Ок*.



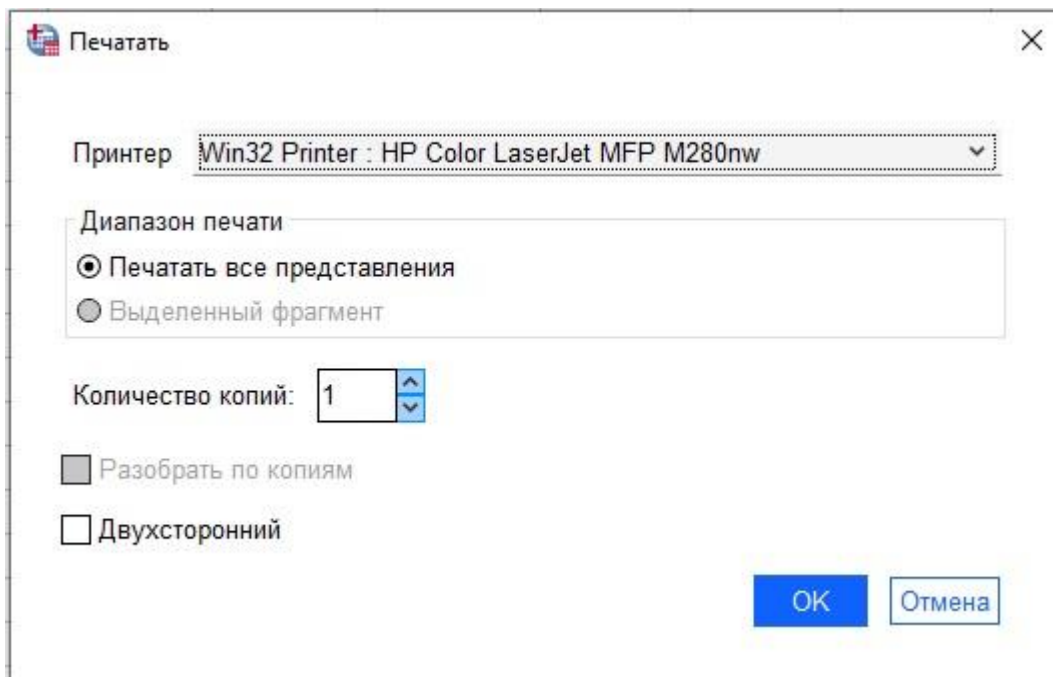


Рис. 17. Диалоговое окно печати

Вместо выбора принтера можно задать создание файл \*.pdf.

### Создание диаграммы в SPSS

Воспользуемся данными из файла Sample Data Set.sav. Откройте файл, на вкладке *Данные*, кликните меню *Графика*. Выберите одну из опций: *Мастер диаграмм*, *Панель выбора диаграмм*, *Устаревшие диалоговые окна*. Последняя опция позволяет выбрать один из стандартных типов диаграмм. Первые две опции проведут вас по пути создания диаграммы, наиболее подходящей к выбранным данным.

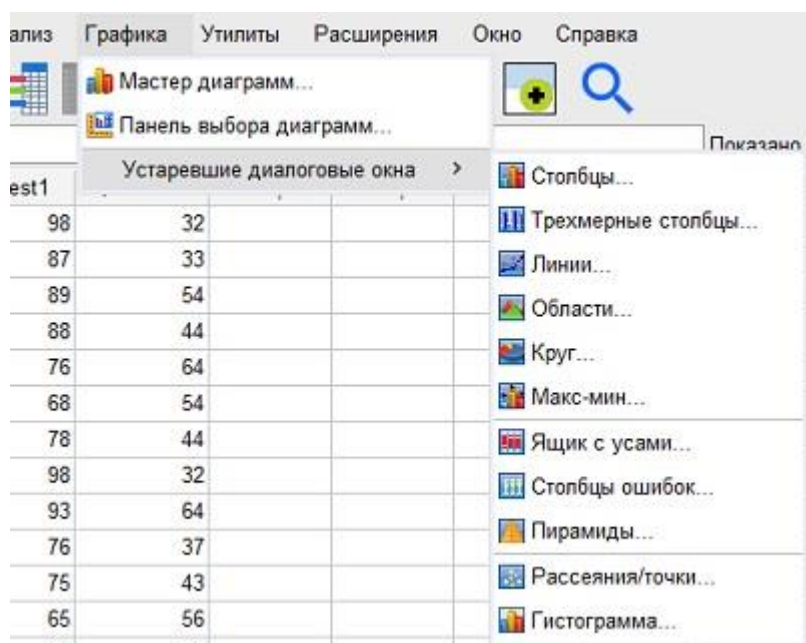


Рис. 18. Типы диаграмм

Выберите *Столбцы*, откроется диалоговое окно, предлагающее несколько вариантов оформления:

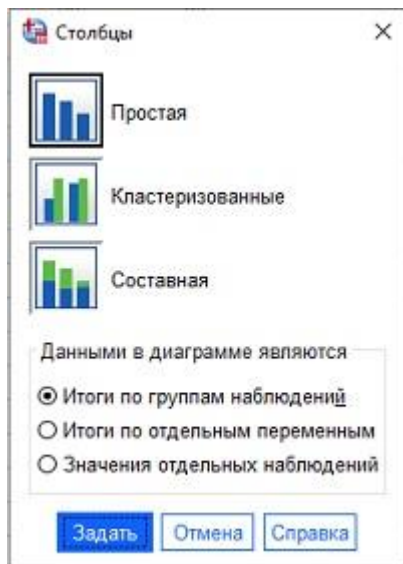


Рис. 19. Виды столбчатой диаграммы

Выберите *Простая* и *Итоги по группам наблюдений*. Нажмите *Задать*. Откроется окно *Простые столбцы*. Задайте, что будет анализировать диаграмма:

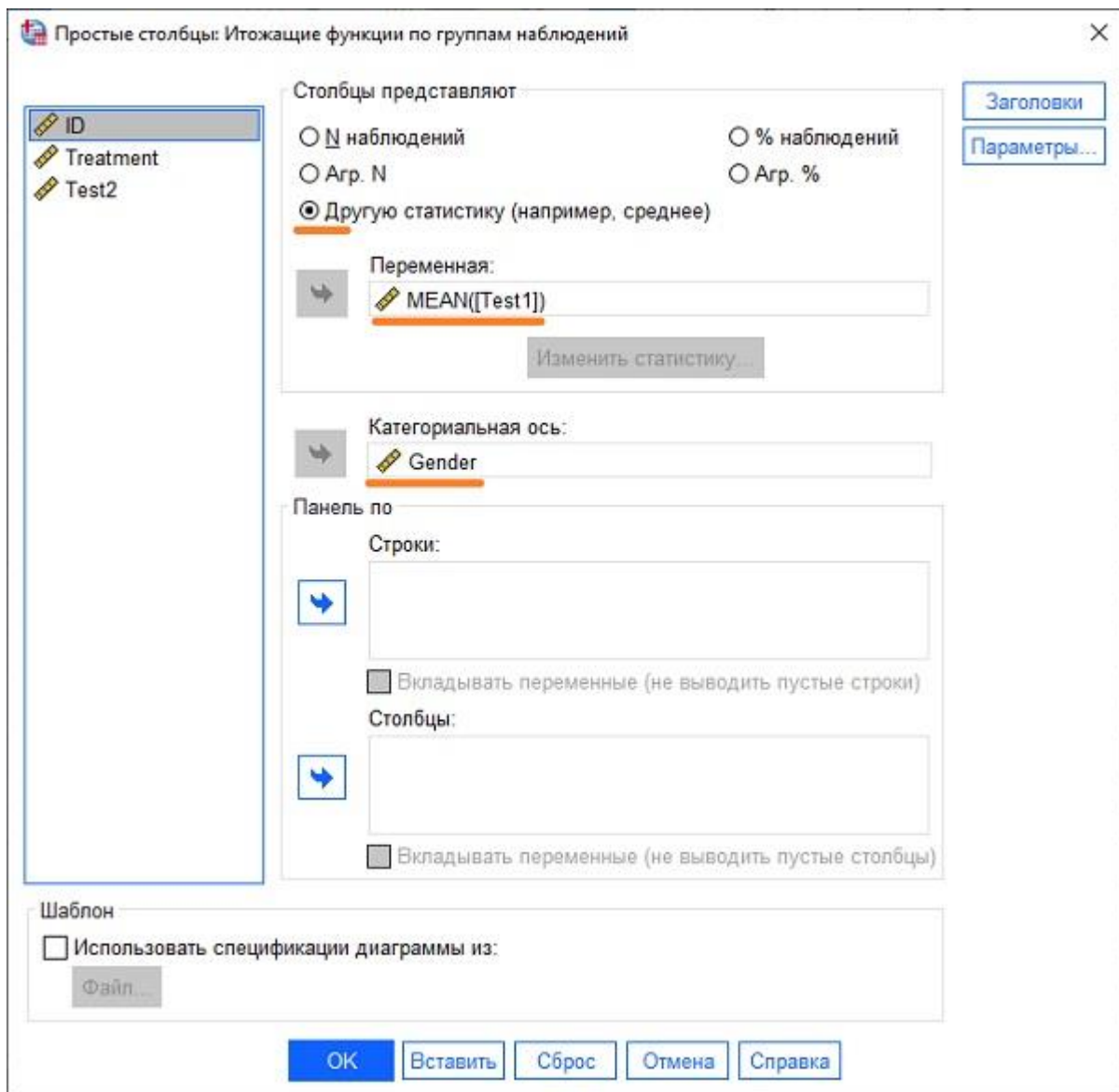


Рис. 20. Параметры аналитики диаграммы

Нажмите *Ok*. В окне вывода появится диаграмма: среднее значение Test1 отдельно по полу:

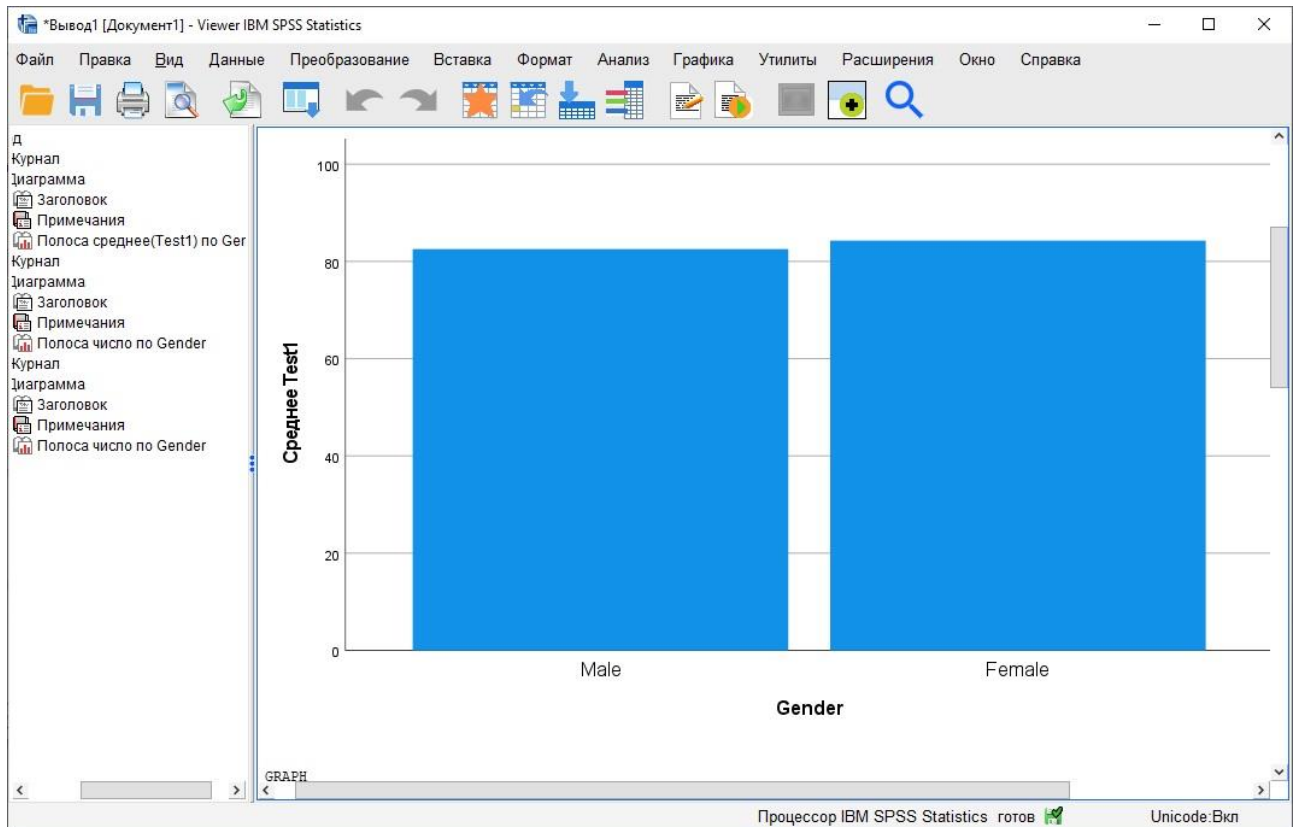


Рис. 21. Средние результаты Теста 1

#### *Сохранение диаграммы*

Диаграмма является частью окна вывода. В этом окне сохраняется любой выполняемый вами анализ. Диаграмма не является самостоятельной сущностью, и ее нельзя сохранить в качестве таковой. Для того чтобы сохранить диаграмму, вам нужно сохранить содержимое всего окна вывода. Для этого:

- 1) кликните *Файл* → *Сохранить*;
- 2) задайте имя для окна вывода и папку;
- 3) нажмите *Ok*; вывод сохранится в файле с расширением \*.sps.

#### *Редактирование диаграммы*

Для изменения диаграммы используйте *Редактор диаграмм*. Чтобы вызвать его дважды кликните на диаграмме в окне вывода.

Чтобы добавить заголовок кликните на соответствующей иконке на панели инструментов *Редактора диаграмм*:

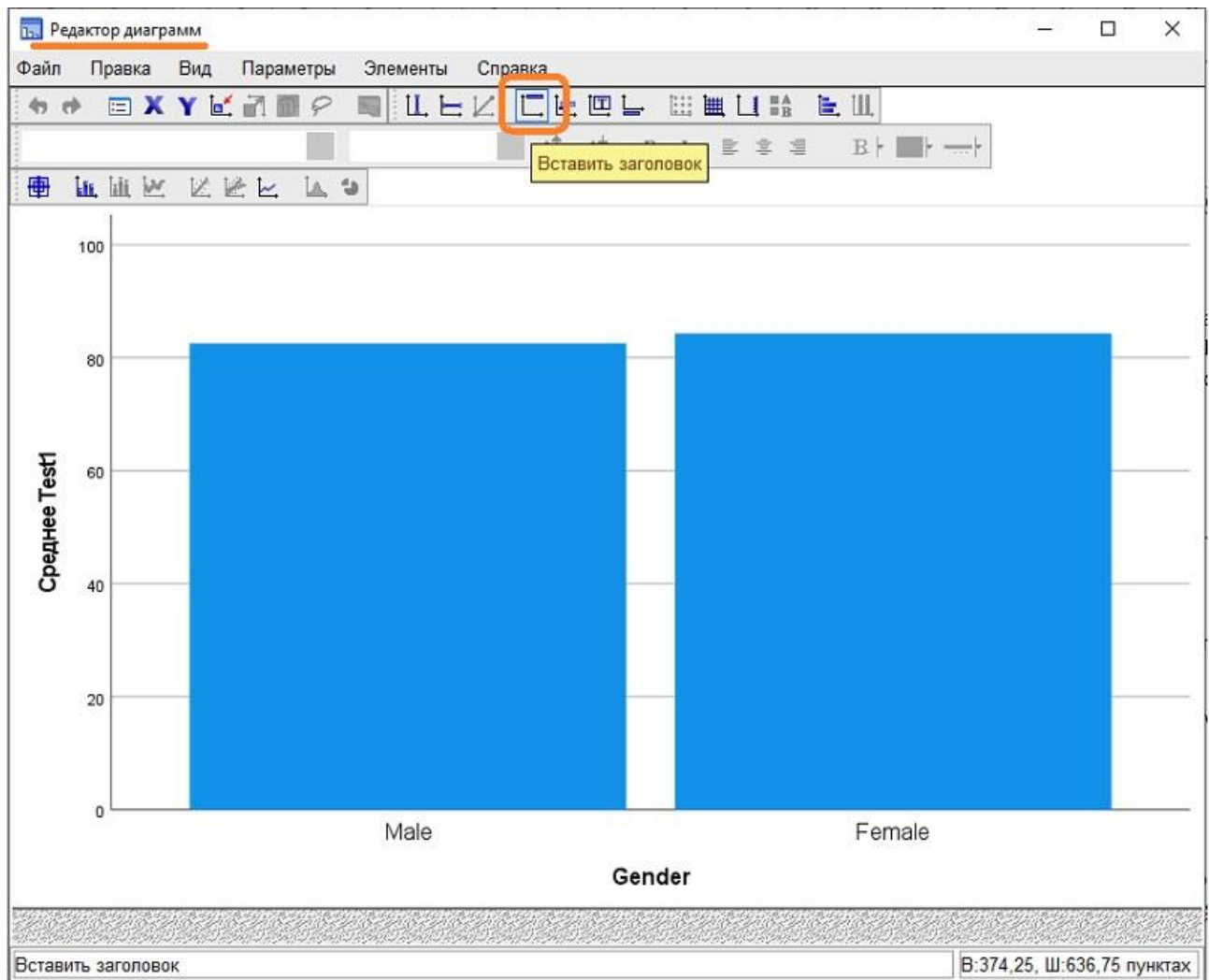


Рис. 22. Кнопка *Редактор диаграмм*

На диаграмме появится область для ввода заголовка и окно *Свойства*, где можно выбрать шрифты, границы и заливку. Для добавления подзаголовка (или даже нескольких) кликните на иконке *Вставить заголовок* повторно.

Для изменения любого элемента дважды щелкните на нем. Можно отдельно щелкнуть на названии оси, подписях и самой оси. В первых двух случаях можно будет отредактировать шрифты и стили оформления, а в последнем – масштаб.

Чтобы выйти из *Редактора диаграмм* просто кликните на крестике окна или пройдите *Файл* → *Закреть*.

#### *Описание данных*

**Частоты и сопряженные таблицы.** Частоты подсчитывают количество случаев возникновения определенного значения. Сопряженные таблицы позволяют подсчитать количество случаев возникновения определенного значения с разбивкой по одной или более категориям, например, по полу и возрасту. Для вычисления частот перейдите в окно *Редактора данных*, кликните *Анализ* → *Описательные статистики* → *Частоты*. Откроется диалоговое окно *Частоты* (см. выше рис. 8). Дважды щелкните мышью по переменным, для которых вы хотите посчитать частоты. В нашем случае это Test 1 и Test 2:

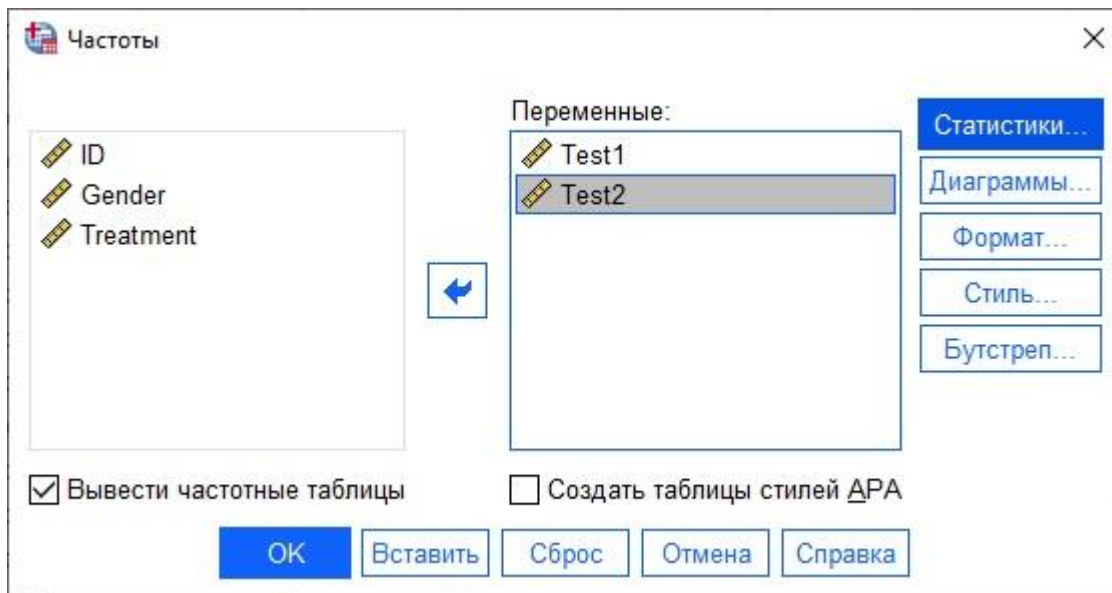


Рис. 23. Диалоговое окно *Частоты*

Щелкните по кнопке *Статистики*. Откроется диалоговое окно *Частоты: статистики*. В разделе *Разброс* отметьте *Стандартное отклонение*. В разделе *Положение центра распределения* отметьте *Среднее*:

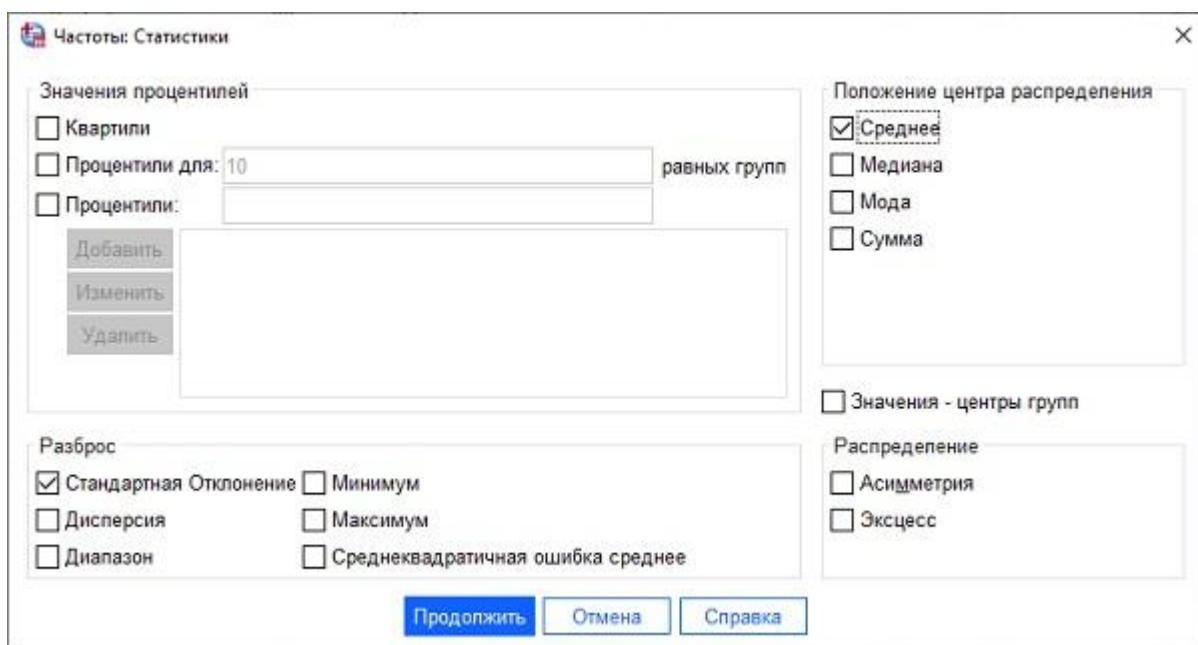


Рис. 24. Диалоговое окно *Частоты: Статистики*

Нажмите *Продолжить*, а затем *Ok*.

В окне вывода появится три таблицы: обобщенная, и подробная для каждой переменной – Test 1 и Test 2:

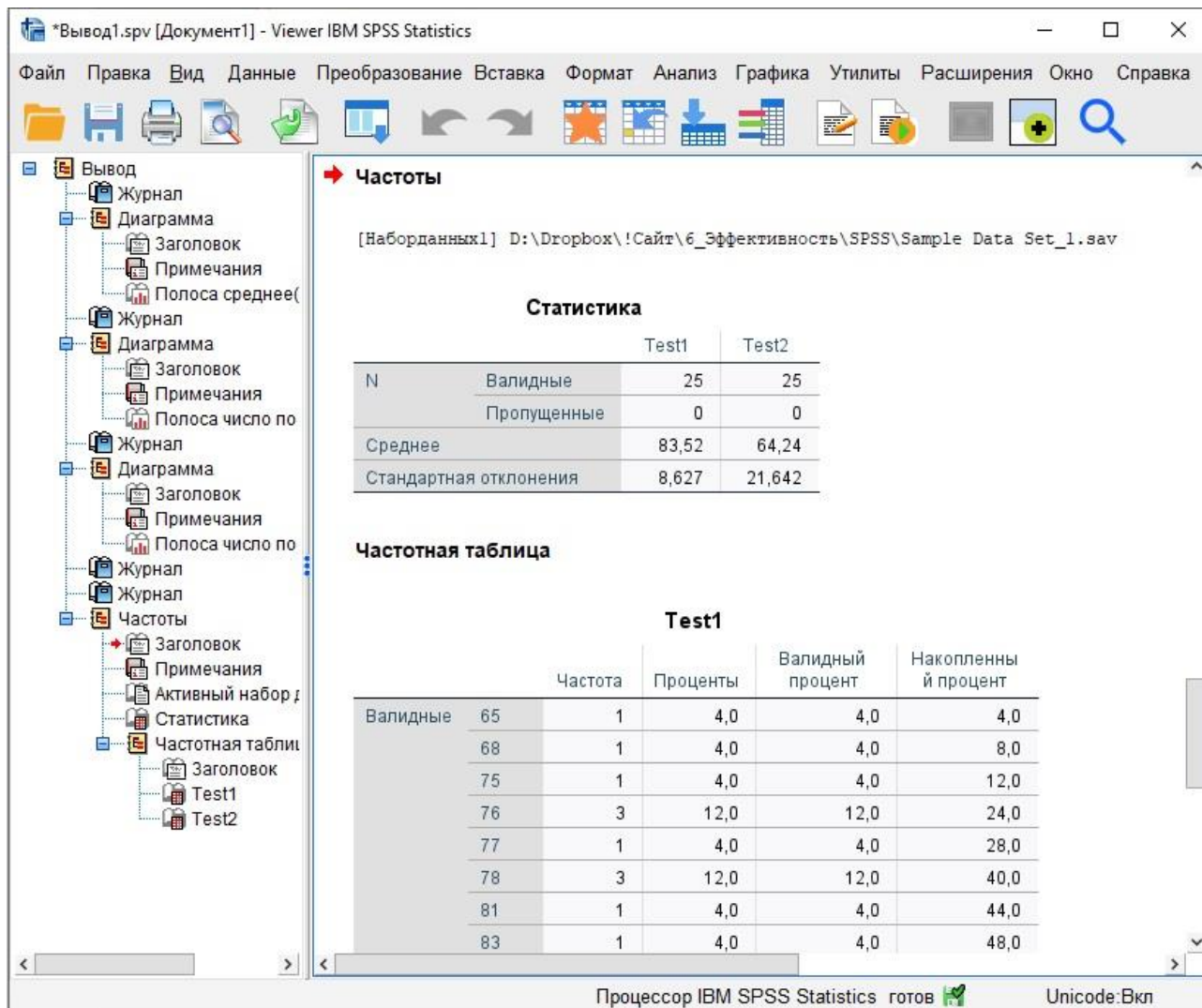


Рис. 25. Обобщенная статистика частот и фрагмент подробной таблицы частот Test 1

### Выход из SPSS

Кликните *Файл* → *Выход*. SPSS позаботится о том, чтобы сохранить все не сохраненные ранее или отредактированные окна, а затем закроется.

Только что вы кратко познакомились с SPSS. Однако эти навыки ничего не значат, если вы не понимаете смысла того, что делаете. Так что не восхищайтесь своими или чужими навыками пользования такими программами, как SPSS. Восхищайтесь, когда люди могут рассказать, что означает тот или иной вывод и какой ответ он дает на поставленный вопрос. И особенно восхищайтесь, если вы сами можете это сделать!