

Барышев В.В.  
Попучиев В.В.



**Статистическая оценка  
событий в клинических  
наблюдениях**

Обнинск – 2009

УДК 57.087.1

ББК 22.172

ISBN

Барышев В.В., Попучиев В.В.

Статистическая оценка событий клинических наблюдений.

Настоящее пособие позволяет максимально упростить применение статистического анализа при проведении медицинских исследований и оценке результатов клинических наблюдений с современных позиций доказательной медицины.

Пособие представлено в виде иллюстрированного обзора составления баз данных с последующей их обработкой и интерпретацией полученных результатов с помощью статистических пакетов Statistica и SPSS.

Книга предназначена для практических врачей, научных работников, организаторов здравоохранения, а также может быть полезна студентам медицинских и биологических вузов.

Рецензенты:

доктор медицинских наук, профессор МРНЦ РАМН, И.А. Гулидов

доктор медицинских наук, МРНЦ РАМН А.П.Бирюков

ISBN

© Барышев В.В., Попучиев В.В., 2009

© Типография

## Оглавление

Оглавление .....	3
Введение .....	4
Основы анализа времени до наступления события и статистические термины .....	4
Составление базы данных в рабочей книге приложения Excel Microsoft Office .....	6
Интерпретация полученных данных в пакете SPSS .....	14
Краткий терминологический словарь .....	34
Список литературы .....	37



**Современные ушные  
и носовые тампоны ПВА  
— эксклюзивные поставки  
из Германии**

**Подробная информация на стр. 38**

общество с ограниченной ответственностью

**СОЮЗ**

410002, г. Саратов, ул. Мичурина, 150/154  
тел.: (8452) 234-000  
www.pvatampons.ru

## Введение

Медицинская статистика — необходимый инструмент оценки клинических и экспериментальных исследований, для адекватного и информативного представления читателю полученных результатов.

Публикации, которые могут быть полезны для статистического анализа в биологии и медицине в настоящее время многочисленны и вполне доступны. Некоторые наиболее полезные, на наш взгляд, прилагаются списком в конце книги. Однако, многие из них требуют математического склада мышления или значительного времени для оценки клинических и экспериментальных данных. Одним из важных показателей эффективности работы лечебного учреждения, либо используемой методики является время до наступления события (рецидив опухоли, смерть, выздоровление и т.д.).

В настоящее время, читателю доступна подробная и хорошо иллюстрированная книга О.Ю. Ребровой «Статистический анализ медицинских данных. Применение пакета прикладных программ STATISTICA». Вместе с тем существует ряд мелких вопросов, ответы на которые необходимо постигать самостоятельно, методом проб и ошибок, что отнимает массу времени. Кроме того, нам не удалось сы найти подробного, иллюстрированного описания работы в статистическом пакете SPSS касательно оценки вероятности до наступления исхода.

Цель данного пособия — максимально упростить применение статистического анализа при проведении медицинских исследований и оценке результатов практической деятельности врачей и научных работников с современных позиций доказательной медицины.

Пособие представляет краткий иллюстрированный обзор составления баз данных с последующей их обработкой в лицензионных статистических пакетах Statistica и SPSS. Начальный раздел посвящен краткому обзору основ анализа времени до наступления исхода. В последующем рассматривается вопрос формирования данных для анализа информации, которую можно вносить в рабочий лист статистического пакета, либо приложение Excel Microsoft Office. В этом приложении достаточно просто редактировать, сортировать и экспортировать записи. Итогом будет универсальная таблица, которую можно открыть практически в любом статистическом пакете. Функция сортировки доступна и в статистических пакетах, однако для её понимания и осуществления, на наш взгляд, требуется гораздо больше времени.

## Основы анализа времени до наступления события и статистические термины.

Анализ времени до наступления события (исхода) — статистические процедуры для анализа времени между точкой включения участника в исследование и наступлением изучаемого исхода.

Особенность исследований времени до наступления исхода заключается в том, что сами исходы могут наблюдаться на протяжении очень длительного периода времени, либо не наступить совсем. Поэтому дождаться изучаемого исхода у всех больных достаточно трудно. В связи с этим исследуемые наблюдения подразделяются на три типа:

- Наблюдения с наступившим исходом — **нецензурируемые** (завершенные или полные), когда точно известны дата начала наблюдения и дата исхода.
- Наблюдения, где исход не наступил на момент окончания исследования (проведения анализа) — **цензурированные** (не завершенные или не полные).
- Наблюдения, в которых исход не зафиксирован по причине выбывания объекта наблюдения по причинам не связанным с наступлением исхода. Временной интервал при этом известен — так же, **цензурированные** (неполные) наблюдения.

Еще одна особенность исследований связанных с анализом времени до наступления исхода заключается в том, что объекты исследования должны включаться в определенный момент, который будет являться критерием включения. Например, время окончания лечения, день установления диагноза, момент констатации периода ремиссии заболевания или день хирургического вмешательства.

Не менее важно соблюдать требования к численности исследуемой группы для более равномерного выбывания в ходе наблюдения и отсутствие изменений в тактике ведения, либо диагностики в исследуемой группе. Основной смысл определения объема выборки — обеспечение необходимой статистической чувствительности исследования. Определить необходимый объем выборок можно с помощью математических формул либо специальных номограмм, которые представлены в учебниках по статистике [5].

Итогом работы в данном разделе статистики будут являться:

- Влияние одного или нескольких факторов на время до наступления события
- Вероятность наступления исхода в определенный промежуток времени
- Определение статистической значимости различий между сравниваемыми группами

Таким образом, с помощью анализа времени до наступления события можно определить следующие показатели:

**Безрецидивная выживаемость** — (disease free survival) рассчитывается от даты констатации полной ремиссии до даты рецидива или последней явки больного. Безрецидивная выживаемость характеризует только группу больных, достигших полной ремиссии, и показывает, какая часть из этих больных имеет возможность прожить указанный срок без признаков возврата заболевания.



**Выживаемость, зависящая от заболевания** – (disease specific survival) рассчитывается от даты начала лечения до даты смерти от данной болезни или до даты последней явки больного. Этот показатель характеризует всю группу больных, начавших лечение, и показывает, какая часть больных могла бы прожить указанный срок, если бы не было осложнений лечения и смертей от осложнений лечения в полной ремиссии заболевания.

**Общая выживаемость** – (overall survival) рассчитывается от даты начала лечения до смерти от любой причины или до даты последней явки больного. Общая выживаемость характеризует всю группу больных, начавших лечение и показывает фактическую выживаемость на момент наблюдения.

**Выживаемость, свободная от неудач лечения** – (freedom from treatment failure) рассчитывается от начала лечения до любой «неудачи» или до даты последней явки больного. Под «неудачей» лечения подразумевается прогрессирование, отсутствие полной ремиссии после окончания лечения, рецидив, осложнение лечения вызвавшие его прекращение, смерть от любой причины. Этот параметр характеризует всю группу больных, начавших лечение, и показывает, какая часть из всей группы больных имеет возможность прожить указанный срок без признаков заболевания.

**Бессобытийная выживаемость** – (event free survival) рассчитывается от даты начала лечения до любого «отрицательного» события или до даты последней явки больного. Под «отрицательным» событием понимается: прогрессирование, отсутствие ремиссии после окончания лечения, осложнения лечения, вызвавшие его прекращение, рецидив, смерть от любой причины.

**Локальный контроль** – (local control) рассчитывается от даты констатации полной или частичной ремиссии до даты рецидива, возобновления прогрессирования или последней явки больного. Локальный контроль характеризует только группу больных, достигших полной или частичной ремиссии.

### Составление базы данных в рабочей книге приложения Excel Microsoft Office.

Открываем Excel Microsoft Office. С помощью меню **Файл**, создаем новую рабочую книгу с необходимым количеством листов (Рисунок 1). **Первая строка** содержит названия столбцов базы данных. Обратите внимание, что заголовки пишутся латиницей: **«patient» «fr» «chemo»** и т.д. Это же условие необходимо соблюдать при заполнении записей базы данных (заполнении ячеек информацией так же на латинице, либо цифрами). Данные правила позволяют в дальнейшем свободно конвертировать таблицу как в пакет **Statistica**, так и в **SPSS**.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	patient	mts	fr	chemo	metod	first_data	relapse_data	final_observatons	OV_month	OV_year	Censor_obsh	recid_m	recid_ar	Censor_recidiv	
2	1	no_mts	sup_1	1_kurs	metod_1	08.03.97	10.22.97	10.21.98	101		0.0	live	0	0.0	recidiv
3	2	mts	sup_1_15	2_kurs	metod_2	22.11.97	29.11.97	11.12.97	1	0.1	live	0	0.0	recidiv	
4	3	mts	sup_1_1	2_kurs	metod_1	29.12.91	23.01.92	23.01.92	1	0.1	live	1	0.1	no_recidiv	
5	4	mts	convex	1_kurs	metod_2	22.11.05	28.12.05	28.12.05	1	0.1	live	1	0.1	no_recidiv	
6	5	mts	sup_1_15	2_kurs	metod_3	04.11.04	14.12.04	14.12.04	1	0.1	live	1	0.1	no_recidiv	
7	6	mts	sup_1_1	2_kurs	metod_1	05.05.04	18.06.04	18.06.04	1	0.1	live	1	0.1	no_recidiv	
8	7	no_mts	sup_1_1	1_kurs	metod_1	16.07.01	30.08.01	30.08.01	1	0.1	live	1	0.1	no_recidiv	
9	8	mts	convex	1_kurs	metod_2	09.12.91	01.02.92	05.02.92	2	0.2	death	2	0.1	recidiv	
10	9	mts	sup_1_1	1_kurs	metod_1	18.12.98	10.02.99	10.02.99	2	0.1	live	2	0.1	no_recidiv	
11	10	no_mts	sup_1_1	1_kurs	metod_1	01.03.88	24.04.88	24.04.88	2	0.1	live	2	0.1	no_recidiv	
12	11	no_mts	sup_1_1	1_kurs	metod_1	29.10.92	24.12.92	24.12.92	2	0.2	live	2	0.2	no_recidiv	
13	12	mts	sup_1_15	2_kurs	metod_3	06.10.05	02.12.05	02.12.05	2	0.2	live	2	0.2	no_recidiv	
14	13	no_mts	sup_1_15	1_kurs	metod_3	03.09.90	01.11.90	16.07.91	10	0.3	death	2	0.2	recidiv	
15	14	mts	sup_1_15	2_kurs	metod_3	11.01.99	12.03.99	12.03.99	2	0.2	live	2	0.2	no_recidiv	
16	15	mts	convex	1_kurs	metod_2	10.11.96	14.01.97	14.01.97	2	0.2	live	2	0.2	no_recidiv	
17	16	mts	sup_1_1	2_kurs	metod_1	18.11.97	04.02.98	07.08.98	9	0.7	death	3	0.2	recidiv	
18	17	mts	sup_1_15	2_kurs	metod_3	06.03.02	24.05.02	24.05.02	3	0.2	live	3	0.2	no_recidiv	
19	18	mts	sup_1_15	1_kurs	metod_2	05.12.05	27.02.06	27.02.06	3	0.2	live	3	0.2	no_recidiv	
20	19	mts	convex	2_kurs	metod_2	06.05.99	16.08.99	08.09.99	4	0.3	live	3	0.3	recidiv	
21	20	mts	convex	1_kurs	metod_2	14.10.94	25.01.95	25.01.95	3	0.3	live	3	0.3	no_recidiv	
22	21	mts	sup_1_1	1_kurs	metod_1	05.02.90	21.05.90	27.07.90	6	0.5	death	4	0.3	recidiv	
23	22	mts	convex	1_kurs	metod_2	03.11.97	14.12.99	03.07.06	104	8.7	live	25	2.1	recidiv	
24	23	mts	sup_1_1	1_kurs	metod_1	28.02.96	23.06.96	31.12.96	10	0.9	death	4	0.3	recidiv	
25	24	mts	convex	2_kurs	metod_2	05.12.02	29.03.03	29.03.03	4	0.3	live	4	0.3	no_recidiv	
26	25	mts	sup_1_1	1_kurs	metod_1	15.06.89	19.10.89	19.10.89	4	0.3	live	4	0.3	no_recidiv	
27	26	mts	convex	1_kurs	metod_2	14.07.05	30.11.05	30.11.05	5	0.4	live	5	0.4	no_recidiv	
28	27	mts	sup_1_1	1_kurs	metod_1	27.03.89	21.08.89	21.08.89	5	0.4	live	5	0.4	recidiv	
29	28	mts	convex	1_kurs	metod_2	22.02.89	18.07.89	22.11.89	11	0.5	death	5	0.4	recidiv	

Рисунок 1. Заполнение данными рабочей таблицы Excel

Как было сказано выше, для работы в статистических модулях с целью определения вероятности наступления исхода обязательны следующие данные: начало наблюдения (включение в исследование), наступления события и показатель цензурирования. В нашем случае это **«first\_data»**, **«relapse\_data»** (дата наступления события – рецидива), **«final\_observatons»** (дата последнего наблюдения) и показатели цензурирования (для наступления рецидива **«Censor\_recidiv»** и показатель цензурирования для общей выживаемости **«Censor\_obsh»**). Показатель цензурирования указывает, произошло событие или нет. Например, в нашей базе данных (таблице) для общей выживаемости нецензурированное наблюдение будет отмечено в столбце **«final\_observatons»** как **«death»**, а для цензурированного **«live»** (Рисунок 1).

Результаты диагностики (в нашем случае столбец **«mts»** наличие или отсутствие метастазов), варианты стандартных методик лечения (столбец **«metod»** содержит, виды операций, эффективность которых нужно сравнить) и другие сведения о проведенном лечении (столбцы **«fr»** и **«chemo»** — вид фракционирования лучевой терапии и число курсов химиотерапии) помещаются в отдельные столбцы.

Следует сказать несколько слов о стандартизации при вводе данных. Так, если мы изучаем время до наступления рецидива опухоли (ис-

ход) в зависимости от наличия или отсутствия метастазов, то в соответствующем столбце «mts» для каждого наблюдения допустимы записи, либо «no\_mts», либо «mts». Это позволит статистическому пакету при подсчете правильно группировать больных по данному критерию. Если же хоть в одной из записей мы укажем пациенту без метастазов «нет\_mts», то сравниваться будут три группы: «no\_mts», «mts» и «нет\_mts», так как статистический пакет не «понимает», что «no\_mts» и «нет\_mts» в нашем случае одно и то же.

**Строки со второй** и далее содержат записи базы данных.

Например: пациент под №1 (выделен) не имел метастазов «no\_mts», методика лучевой терапии фракционирование «sup1\_1», получил 1 курс химиотерапии во время лучевого лечения «1\_kurs», хирургическое лечение проведено по методике «метод\_1», дата начала наблюдения 08.08.1989, дата рецидива 19.02.1992, дата последнего наблюдения 19.01.1998. Далее, для этого пациента, в столбце «OV\_month» в ячейки записи базы данных вводим формулу =ДНЕЙ360(F2;H2)/30. Эта формула отражает период наблюдения в месяцах (результат виден в ячейке) исходя из начальной (столбец F2) и конечной (столбец H2) даты наблюдения. Для второго пациента формула будет выглядеть так: =ДНЕЙ360(F3;H3)/30, для третьего =ДНЕЙ360(F4;H4)/30 и т.д. В столбце «OV\_year» содержится формула =2/12, которая позволяет получить срок наблюдения в годах путем деления срока наблюдения в месяцах на 12. Так же можно проследить сроки наблюдения в днях. В этом случае формула будет выглядеть так: =ДНЕЙ360(F2;H2). В столбце «Censor obsh» отмечаем цензурированные наблюдения (что произошло с пациентом: жив – «live» или умер – «death» или выпал из наблюдения «live») для общей выживаемости. Для оценки безрецидивной выживаемости потребуется дата полной ремиссии и дата развития рецидива. Если рецидив так и не развился, то в графе даты рецидива ставят дату последнего наблюдения и отмечают наблюдение как цензурированное или неполное (в нашем случае «no recidiv») Рисунок 1.

Как мы отметили выше, Таблица Excel хороша тем, что в ней достаточно просто редактировать и сортировать данные с помощью функции «Фильтр». Сортировка нужна для выбора требуемых наблюдений из общего массива базы данных. Подробнее о сортировке данных можно прочитать в соответствующих руководствах и справке по Excel Microsoft Office.

Данные введены и отсортированы. Теперь необходимо отключить фильтр в таблице ждем на панели инструментов «Фильтр». При этом исчезают соответствующие флажки в первой строке базы данных и на панели инструментов (Рисунок 2).

Сохраняем полученный результат в файл **stat.xls** в любом удобном месте на компьютере.

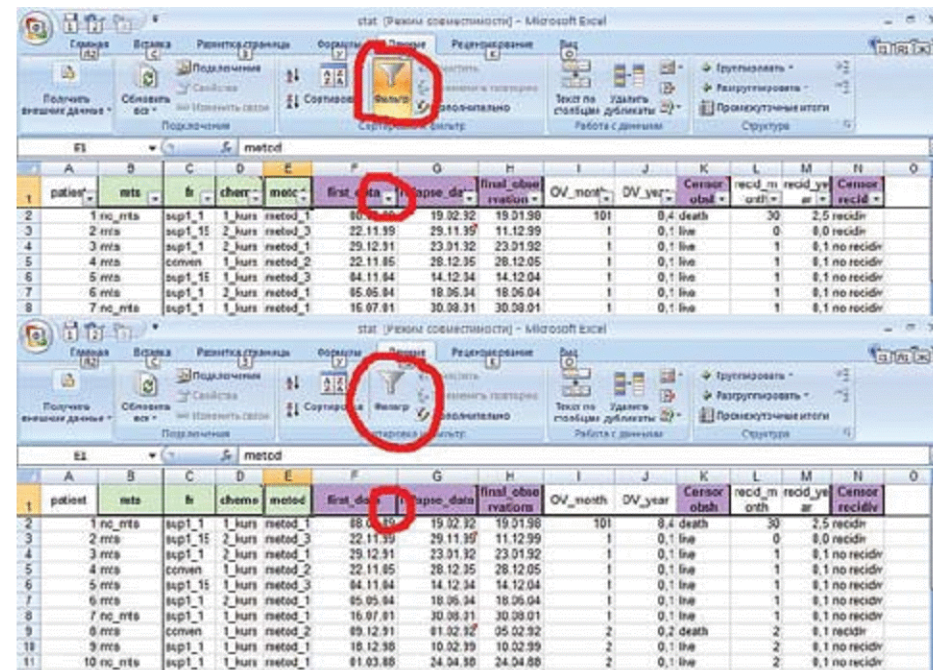


Рисунок 2. Сортировка данных с помощью. Функции «Фильтр» в приложении Excel.

## Анализ времени до наступления события в SPSS.

Открываем лицензионную программу SPSS. В меню **File** выбираем **Open Data** или выбираем команду открыть на панели инструментов (Рисунок 3). В окне «тип файлов» выбираем «\*.xls». Затем, выбираем нужный файл (в нашем случае **stat.xls**) рабочей книги Excel. Нажимаем «Открыть». При этом файл нашей рабочей книги (stat.xls) должен быть сохранен и закрыт. В противном случае программа SPSS выдаст сообщение об ошибке.

Если перечисленные условия выполнены, открывается диалоговое окно «Opening Excel Data Source» Рисунок 4. В окне **Worksheet** этого диалога выберете лист рабочей книги **Excel**, который необходимо импортировать. В нашем случае лист был всего один (Рисунок 1), поэтому в раскрывающемся меню отображается всего один лист рабочей книги. Данные в скобках означают строку, с которой начинаются записи базы **A1**, и число строк базы **N104**. В окне Range можно указать, какие строки рабочей книги Excel необходимо импортировать. Например, если мы напишем **A5:N20**, то программа импортирует 20 записей базы данных, начиная с 5 строки. Если это окно оставить пустым будут импортированы все записи листа рабочей книги. Нажимаем «OK». И затем сохраняем рабочую таблицу SPSS.



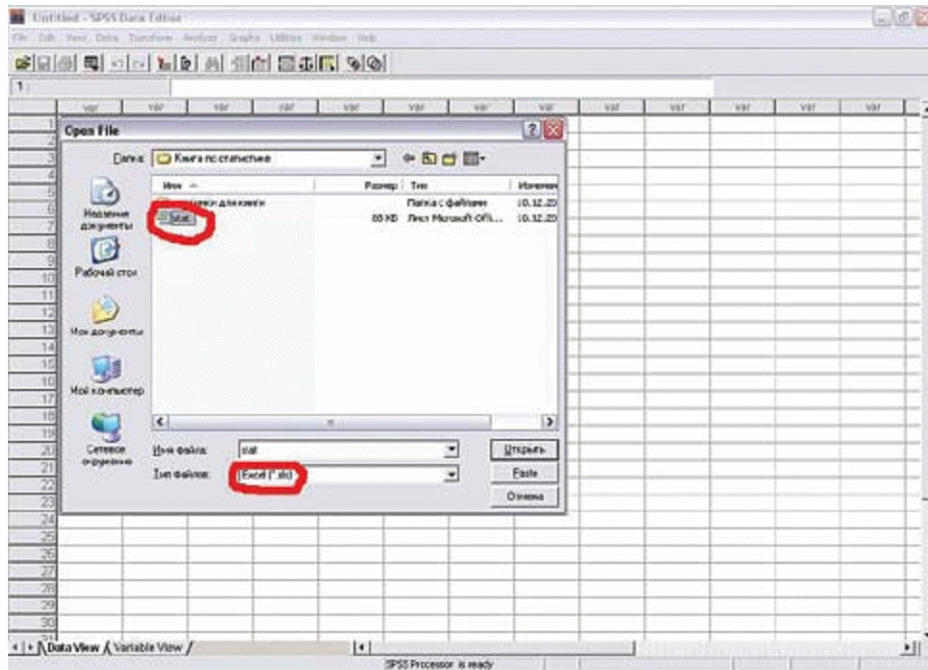


Рисунок 3 Импорт данных в статистический пакет SPSS

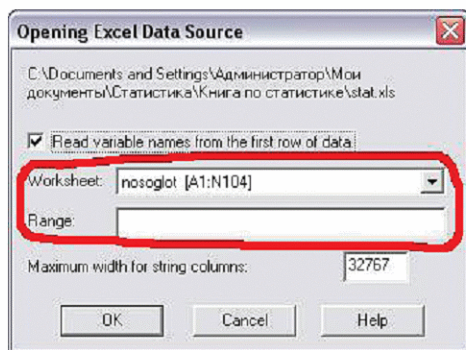


Рисунок 4. Выбор импортируемого листа из рабочей книги Excel.

Рисунок 5 демонстрирует, что структура таблицы SPSS идентична нашей таблице в Excel. При этом первая запись листа рабочей книги Excel импортируется в виде названия столбцов таблицы SPSS. Приступаем к обработке данных.

В строке меню выбираем пункт «**Analyze**» (Анализ), далее, в раскрывающемся меню выбираем «**Survival**» (Выживаемость) и следующим этапом методику анализа. В нашем примере это будет методика «**Kaplan-Meier**». Работа с помощью других методик раздела «**Survival**» аналогична.

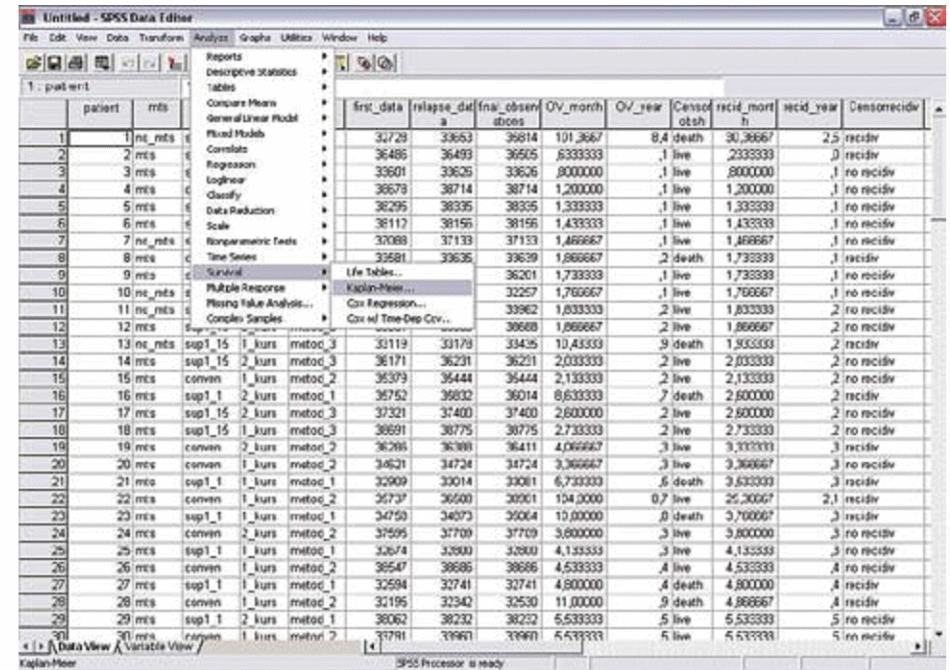


Рисунок 5. Выбор метода анализа в окне программы SPSS.

После выбора методики анализа открывается диалоговое окно методики «**Kaplan-Meier**» (Рисунок 6). В большом окне слева мы видим названия столбцов нашей базы данных. Для определения выживаемости отмечаем столбец «**OV\_year**» (время наблюдения в годах. Если необходимо время наблюдения в месяцах выбираем столбец «**OV\_month**»), выделяем курсором, нажимаем стрелку слева от окна Time и данные становятся доступны для дальнейшей работы. Об этом свидетельствуют заполненные поля в правой части Рисунок 6.

Аналогично указывается путь к данным для других диалоговых окон. В окне «**Status**» указываем столбец с данными цензурирования для общей выживаемости (столбец Censor obs). Чуть ниже располагается кнопка «**Define Event**» (Определить наступление события). При её нажатии открывается диалоговое окно (Рисунок 7), где в поле «**Single value**» необходимо указать индикатор для нецензурированных (полных или завершенных) наблюдений. В нашем случае «**death**», смерть пациента.

Нажимаем «**Continue**» и возвращаемся к оставшимся полям диалогового окна «**Kaplan-Meier**» (Рисунок 6).

В поле необходимо определить столбец, где содержатся метки исследуемых групп. Например, мы собираемся анализировать выживаемость пациентов разделенных на группы в зависимости от методики хирургического вмешательства. Эти данные содержатся в столбце



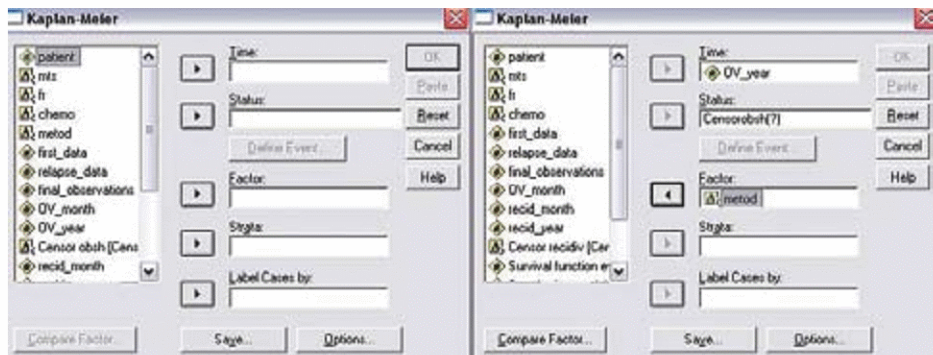


Рисунок 5. Выбор метода анализа в окне программы SPSS.

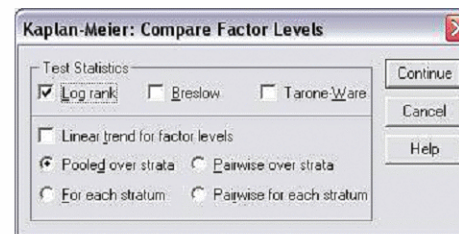


Рисунок 9. Выбор метода сравнения групп.

кой «**Compare Factor**» (Сравнить факторы) после чего отмечаем нужный нам метод сравнения (Рисунок 9) (в нашем случае логоранговый критерий) и его опции. Продолжаем работу кнопкой «**Continue**». Снова открывается диалоговое окно «**Kaplan-Meier**» (Рисунок 6).

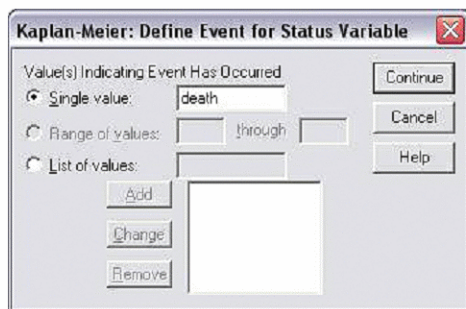


Рисунок 7. Окно определение индикатора наступления события.

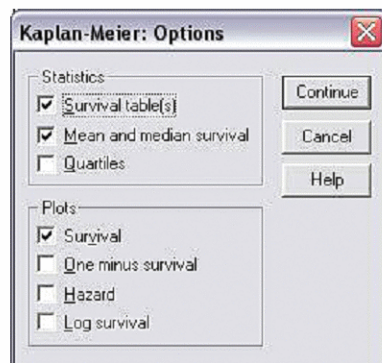


Рисунок 8. Окно выбора представления итоговых данных.

«**metod**». Если необходимо оценить выживаемость в группах в зависимости от режима фракционирования или методики химиотерапии, то мы выбираем столбцы «**fr**» или «**chemo**» соответственно. Практически все готово, остается лишь указать какие таблицы и графики следует создать в качестве итога анализа данных. Для этого нажимаем кнопку «**Options**» внизу справа диалогового окна «**Kaplan-Meier**» (Рисунок 6). В открывшемся окне (Рисунок 8) отмечаем флажки напротив «**Survival table**» (Таблица жизни), «**Mean and median survival**» (Средняя продолжительность и медиана жизни) и, чуть ниже, просим построить график (Plots) для наглядного представления выживаемости Survival. Здесь же можно определить и квантили (25% и 75%).

Когда необходимые таблицы и графики выбраны, нажимаем «**Continue**» и вновь возвращаемся в диалоговое окно «**Kaplan-Meier**» (Рисунок 6). При сравнении двух кривых выживаемости требуется определить достоверность их различий. Для этого можно воспользоваться кнопкой

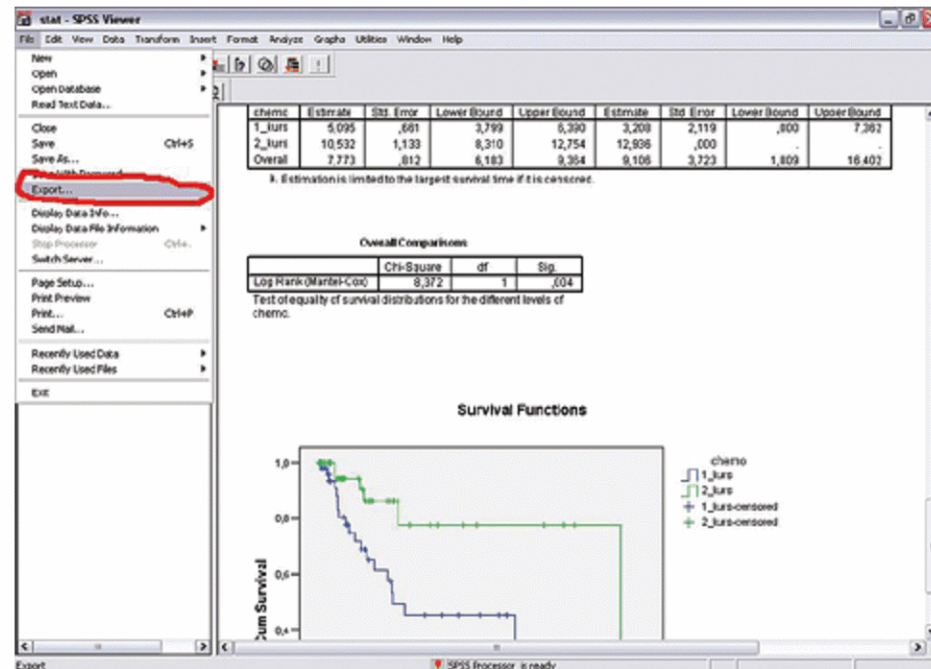


Рисунок 10. Окно вывода конечных результатов анализа.

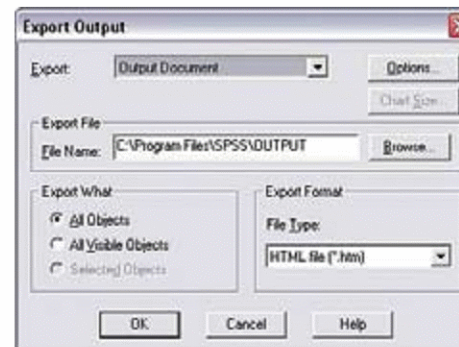


Рисунок 11. Окно выбора формата итогового документа.

Нажимаем ОК и переходим к интерпретации полученных данных. Следует заметить, что в открывшемся окне результатов (Рисунок 10), в меню File можно выбрать функцию экспорта данных.

При выборе функции экспорта можно определить варианты экспортируемого документа (Рисунок 11), путь, выбрать отдельные или все объекты конечного результата и, наконец, формат экспорта (доступны

форматы html, doc, ppt, txt, xls). В опциях этого окна (правая верхняя кнопка) можно определить параметры экспортируемых изображений.

### Интерпретация полученных данных в пакете SPSS.

Полученные результаты начинаются с таблицы исходных данных (Рисунок 12).

Из этой таблицы следует, что анализ проводился по группам в зави-

chemo	Total N	N of Events	Censored	
			N	Percent
1_kurs	56	20	36	64,3%
2_kurs	47	6	41	87,2%
Overall	103	26	77	74,8%

Рисунок 12. Краткая характеристика анализируемых групп.

симости от методики химиотерапии (столбец «chemo» исходной таблицы). Таких групп было две: «1\_kurs» и «2\_kurs» 56 и 47 человек соответственно. Общая численность изучаемых пациентов 103. Далее из таблицы видно, что общее число нецензурированных наблюдений в группе «1\_kurs» было 20, а цензурированных 36, что составило 64,3%. Аналогично интерпретируются данные для группы «2\_kurs» и общие (Overall) для обеих групп.

Следующая таблица («Survival Table», таблица жизни Рисунок 13) представляет результаты анализа по изучаемым группам. В данной таблице нас интересуют столбцы «Time» (Время наблюдения) и «Cumulative Proportion Surviving at the Time» (Кумулятивная выживаемость в определенный промежуток времени). Данную таблицу мы намеренно сократили для удобства восприятия. Зеленым и синим маркерами, отмечены данные для сравниваемых групп «1\_kurs» и «2\_kurs» соответственно.

Из представленных данных следует, что в первой группе (массив данных обведен зеленым маркером) пациентов «1\_kurs» пятилетняя выживаемость была равна 0,453 или 45,3%. Это следует из того, что на протяжении промежутка от 3,728 лет до 5,311 лет наблюдения выживаемость не менялась. Стандартная ошибка была равна 0,093. Следовательно, во второй изучаемой группе «2\_kurs» (массив данных обведен синим маркером) пятилетняя выживаемость составила 0,776 что соответствует 77,6%. Стандартная ошибка 0,101.

Отметим, что при выводе данных, полученных по методике Kaplan-Meier, в виде таблицы, отсутствует деление наблюдения на строгие временные промежутки, в отличие от методики «Life Tables» (Таблицы дожития).

Перейдем к рассмотрению следующей таблицы «Means and Medlans for Survival Time» (Средняя продолжительность и медиана времен жизни) (Рисунок 14). Здесь представлены данные для каждой группы в отдельности и обобщенные показатели (Overall) средней продолжительности жизни и медианы жизни.

chemo	Time	Status	Cumulative Proportion Surviving at the Time		N of Cumulative Events	N of Remaining Cases	
			Estimate	Std. Error			
1_kurs	1	0,000	live		0	56	
	2	,111	live		0	54	
	3	,122	live		0	53	
	38	2,103	death	,653	,080	13	18
	39	2,189	live			13	17
	40	2,428	death	,614	,084	14	16
	41	3,000	death	,576	,087	15	15
	42	3,117	live			15	14
	43	3,172	death	,535	,090	16	13
	44	3,208	death	,494	,092	17	12
45	3,728	death	,453	,093	17	11	
46	4,583	live			18	10	
47	5,311	live			18	9	
48	6,014	live			18	8	
2_kurs	1	0,000	live		0	46	
	2	,067	live			0	45
	3	,119	live			0	44
	30	2,239	live			4	17
	31	2,322	live			4	16
	32	2,367	live			4	15
	33	2,992	live			4	14
	34	3,094	live			4	13
	35	3,225	live			4	12
	36	3,256	live			4	11
	37	3,411	live			4	10
	38	3,453	death	,776	,101	5	9
	39	3,936	live			5	8
40	4,808	live			5	7	
41	5,133	live			5	6	

Рисунок 13. Таблицы жизни, составленные в программе SPSS.

chemo	Mean <sup>a</sup>				Median			
	Estimate	Std. Error	95% Confidence Interval		Estimate	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound			Lower Bound	Upper Bound
1_kurs	5,095	,681	3,799	6,390	3,208	,2119	,000	7,362
2_kurs	10,532	1,133	8,310	12,754	12,936	,000		
Overall	7,773	,812	6,183	9,364	9,106	,3723	1,809	16,402

a. Estimation is limited to the largest survival time if it is censored.

Рисунок 14. Данные медианы и средней продолжительности жизни.



Чаще всего в медицинских исследованиях оперируют понятием «**Медиана жизни**». Медиана продолжительности жизни время, в течение которого наступает исход (смерть, либо рецидив) в 50% наблюдений (соответственно, 50% живут дольше этого времени).

Представленные данные иллюстрируют, что в группе «1\_kurs» медиана жизни составила 3,208 лет (так как изначально для анализа мы брали столбец отражающий срок наблюдения в годах) при этом стандартная ошибка составила 2,119. Во второй изучаемой группе «2\_kurs» эти показатели составили 12,936 и 0,0 соответственно. Медиана выживаемости для общего числа пациентов (без деления на группы) составила 9,106. Стандартная ошибка 3,723.

Данные результата сравнения исследуемых групп представлены в таблице «**Overall Comparisons**» (Рисунок 15).

Overall Comparisons			
	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	8,372	1	,004

Test of equality of survival distributions for the different level chemo.

Рисунок 15. Результаты сравнения Log-Rank Test.

Из данной таблицы следует, сравнение групп производилось с помощью Log-Rank теста<sup>1</sup>. При этом значение хи-квадрат было равным 8,372. Значение  $p=0,004$ , что свидетельствует о достоверных различиях между исследуемыми группами.

И, наконец, заключительная часть представления полученных данных кривые выживаемости (Рисунок 16).

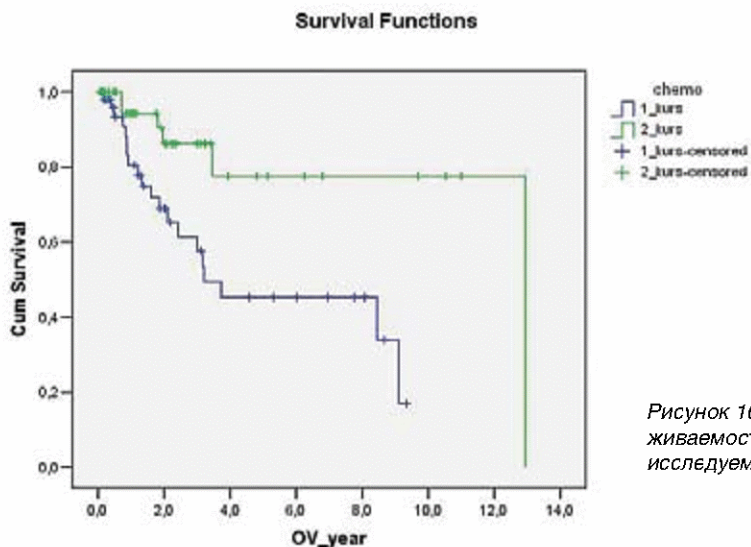


Рисунок 16. Кривые выживаемости для каждой исследуемой группы.

Подобные графики наглядно представляют показатели выживаемости в исследуемых группах: мы видим, что выживаемость больных в группе «2\_kurs» была выше, чем в группе «1\_kurs». Подобные графики легко экспортировать в другие приложения, например для демонстрации результатов в презентации Microsoft Office Power Point. При желании можно легко отредактировать график для большего удобства чтения (Рисунок 17).

Общая выживаемость больных в зависимости от вида химиотерапии

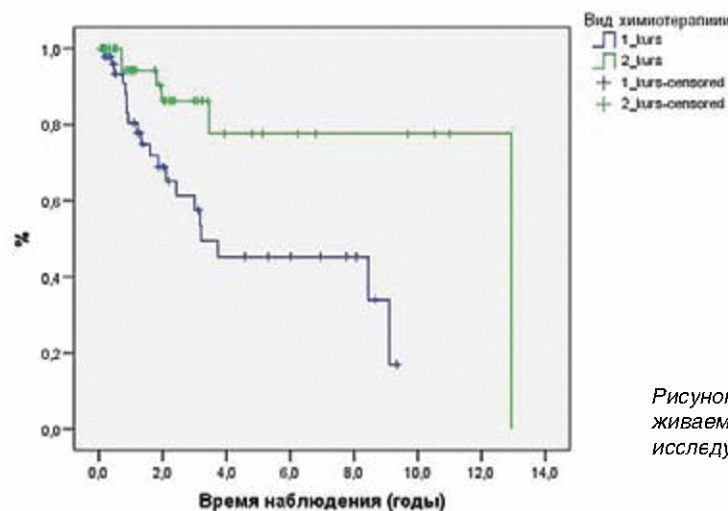
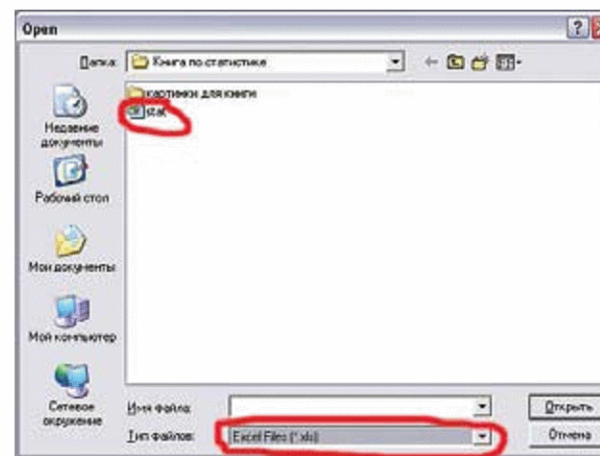


Рисунок 17. Кривые выживаемости для каждой исследуемой группы.

### Анализ времени до наступления события в Statistica.

Открываем лицензионный статистический пакет Statistica. В меню «File» выбираем команду «Open» (Рисунок 18).



Внизу открывшегося окна указываем тип открываемого файла (\*xls), затем из списка файлов выбираем знакомый нам файл stat.xls. После этого в следующем диалоговом окне (Рисунок 19) нам предлагается импортировать либо все листы

Рисунок 18. Окно выбора файла данных.



книги Excel (**Import all sheets a Workbook**), либо выбрать нужные. Флажком внизу окна выбор можно установить по умолчанию.

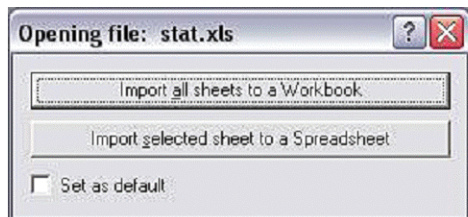
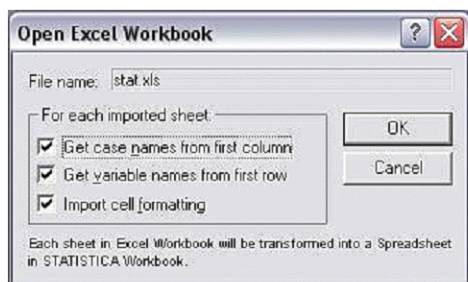


Рисунок 19. Окно выбора конвертируемых рабочих листов Excel.

Наша база данных состоит из одного листа, поэтому подходят оба варианта выбираем **Import all sheets a Workbook**.

В следующем диалоговом окне (Рисунок 20) есть возможность импортировать названия записей базы данных (например, порядковый номер пациентов) из первого столбца (**Get case names from first column**) и названия столбцов переменных из первой строки (**Get variable names from first row**) (Рисунок 21).



Кроме того можно импортировать формат ячеек Excel (**Import cell formatting**). Нажимаем ОК и переходим к импортированной базе данных статистики. Для начала сохраним полученную таблицу.

Рисунок 20. Окно дополнительных параметров конвертации.

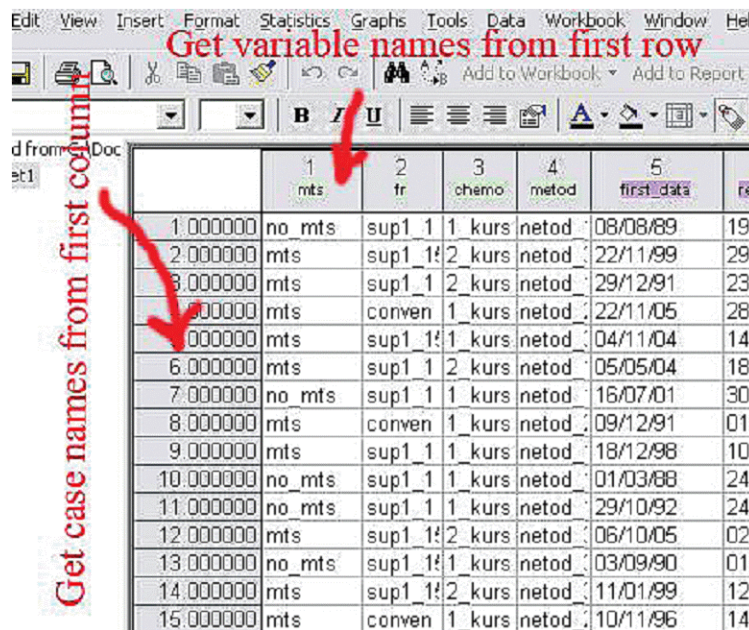


Рисунок 21. Вид импортированных данных в окне программы Statistica.

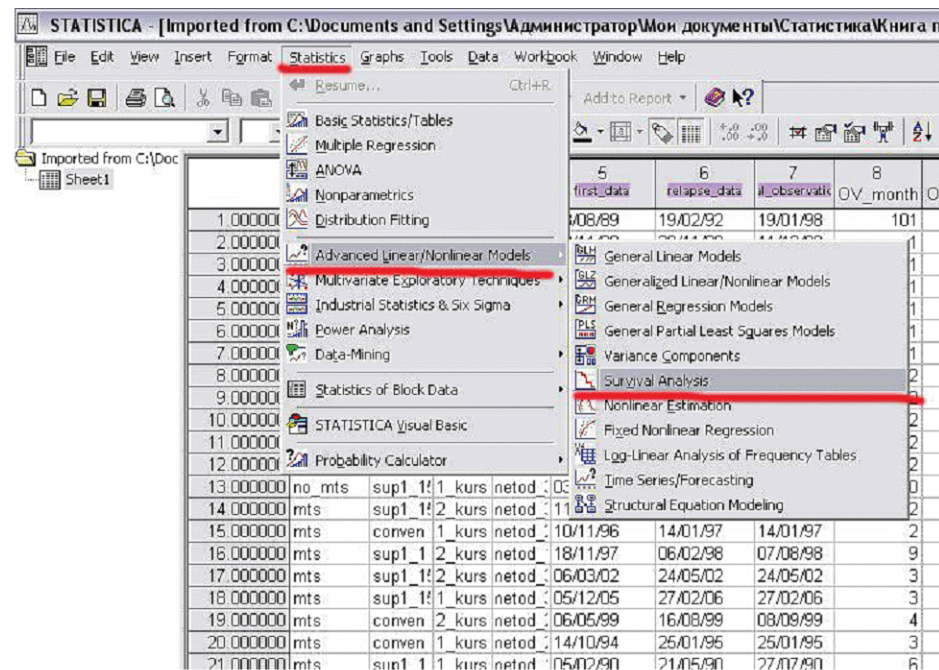


Рисунок 22. Выбор метода статистического анализа данных.

Теперь переходим к обработке данных таблицы (Рисунок 22). В меню **«Statistics»** (Статистики) выбираем **«Advanced Linear/Nonlinear Models»** (Первичные линейные и нелинейные модели), и далее **«Survival Analysis»** (Анализ выживания). В открывшемся окне **«Select Spreadsheet»** (Рисунок 23) выбираем необходимую рабочую таблицу.

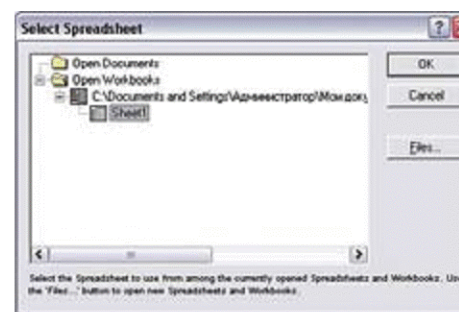


Рисунок 23. Выбор таблицы содержащей данные.

В нашем случае мы импортировали рабочую книгу Excel с одним листом, следовательно, рабочая книга Statistica будет содержать тоже один лист **Sheet1**. Нажимаем **OK**.

В следующем диалоговом окне (Рисунок 24) необходимо указать методику анализа. Начнем с **«Comparing multiple samples»** (Сравнение нескольких случаев). Данная методика позволяет сравнивать одновременно до 10 исследуемых групп. В этом же окне желательно определить опции со-