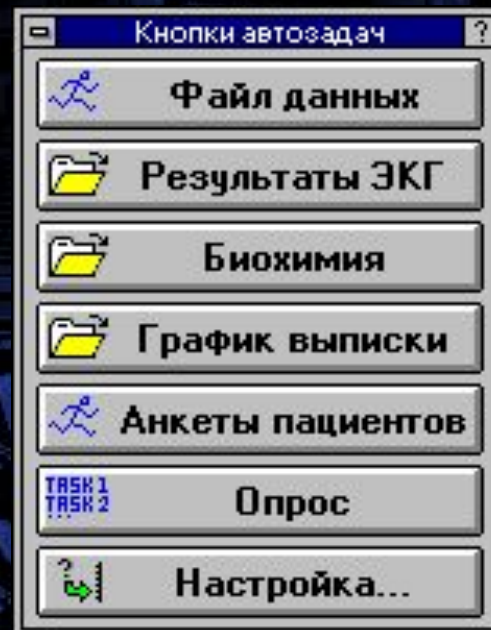
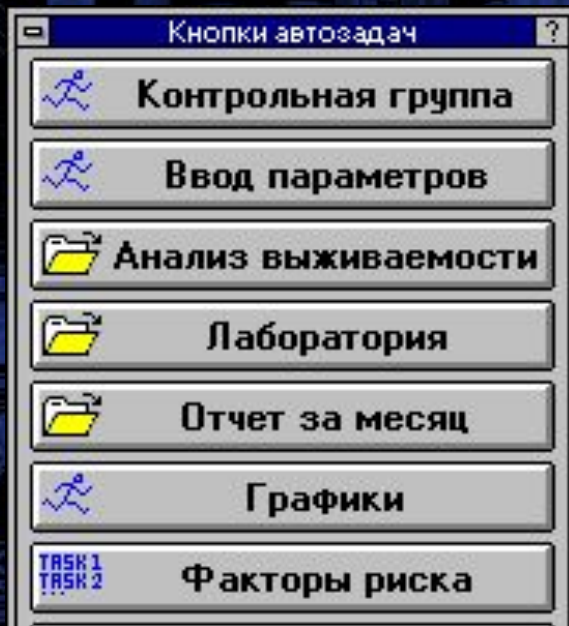


Примеры применения пакета *STATISTICA 5.5*

для статистического
анализа медицинской
информации

StatSoft Russia

Полный набор статистических методов, удобные средства визуализации и гибкий, полностью настраиваемый интерфейс позволяют эффективно использовать систему *STATISTICA* для анализа медицинской информации.



Основные этапы анализа данных

Подготовка данных: заполнение таблиц, импорт, проверка и сортировка.



Разведочный анализ: вычисление основных статистик и построение графиков.



Понижение размерности: факторный и корреляционный анализ.



Анализ зависимостей. Построение линейных и нелинейных моделей.

Примеры

- Статистический анализ результатов обследования полости рта и опроса людей пожилого и старческого возраста.
- Исследование статистических характеристик прибора для транскутанного измерения билирубина крови.
- Предсказание срока выздоровления по длительности пребывания в больнице.
- Анализ выживаемости пациентов после операции на сердце в различных госпиталях.

Пример 1. Изучение состояния полости рта пожилых людей

Данные: STOMAT.STA 282п * 512н

МЕТК ЗНАЧЕ	15 ОБР	16 УСЛ_ТР	17 ПРОФ_ВР	18 ВР_ПР	60 УДАЛЕН	61 КПУ
8	Начальное	Плохие	Есть	Некурящие	22	26
9	Начальное	Плохие	Нет	Курильщики	13	13
10	Высшее	Плохие	Есть	Курильщики	10	13
11	Начальное	Плохие	Есть	Курильщики	0	4
12	Среднее	Хорошие	Нет	Курильщики	4	5
13	Начальное	Плохие	Есть	Курильщики	0	2
14	Начальное	Плохие	Есть	Курильщики	26	26
15	Начальное	Хорошие	Нет	Некурящие	21	22
16	Среднее	Плохие	Нет	Некурящие	26	26
17	Среднее	Плохие	Есть	Курильщики	14	14
18	Начальное	Хорошие	Нет	Некурящие	19	23

В файле данных собраны результаты опроса и объективного обследования

Используемые методы анализа

t-критерий или дисперсионный анализ, если распределение исследуемой переменной близко к нормальному

Внутригрупповые описательные статистики и корреляции - Результаты

ЗАВИСИМЫЕ: 2 перемен.: И_ф_В ПАРОДОНТ

ГРУППИР.: 1-ВОЗР_ГР (3): 60-69 70-79 >=1
2-ПОЛ (2): Мужчина Женщины

Итоговая таблица средних

Подобранные двуходовые таблицы

Отображать длинные имена переменных

Отображать длинные метки значений

Маргинальные средние

Дисперсионный анализ

Апостериорные сравнения средних

Левена Брауна-Форсайта (ОД)

Внутригрупповые корреляции Опции

Переупорядочить факторы в таблице

Дисперсии

Медиана и квантили

Категоризованные диаграммы размаха

Категоризованные гистограммы

Категоризованные норм. вероятн. графики

Графики взаимодействия

Категоризованные диаграммы рассеяния

Графики средних и станд. отклонений

Дисперсионный анализ (stomat.sta)

Отмечены эффекты, значимые на уров. $p < .05000$

Перемен.	Сум. кв.эфф.	Сл. св.эфф.	Ср. кв.эфф.	F	p
И_ф_В	13.42426	5	2.684852	4.519954	.000503
ПАРОДОНТ	11.27024	5	2.254048	3.117164	.008905

График гистограммы с нормальным распределением (Женщины)

Используемые методы анализа

Непараметрические критерии, в том числе ранговый дисперсионный анализ

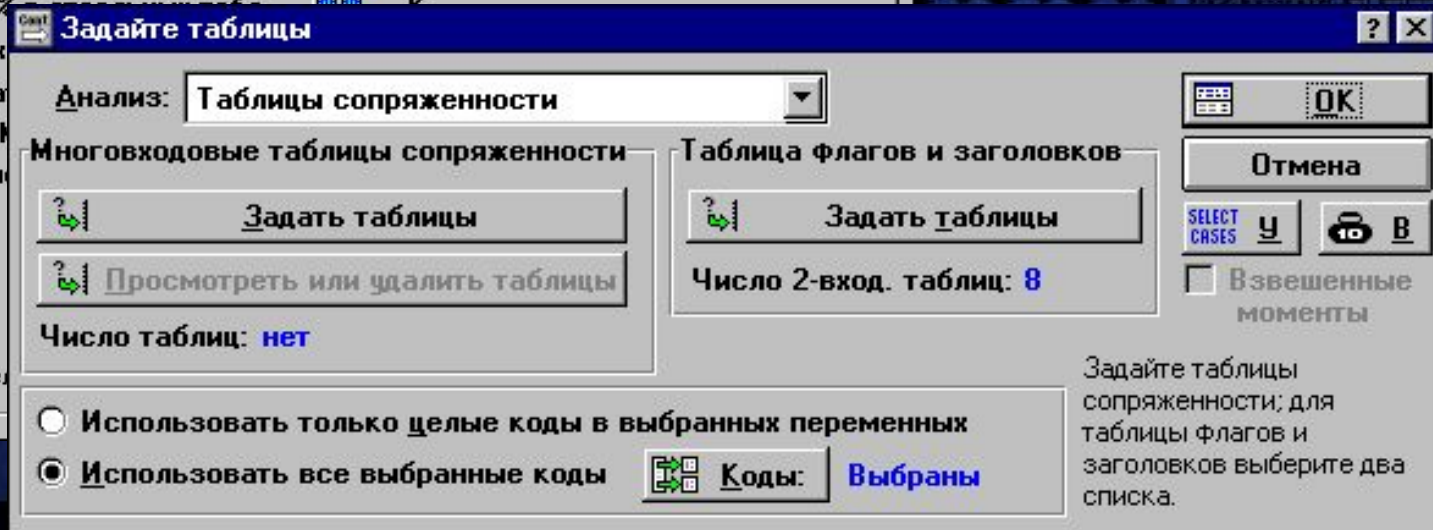
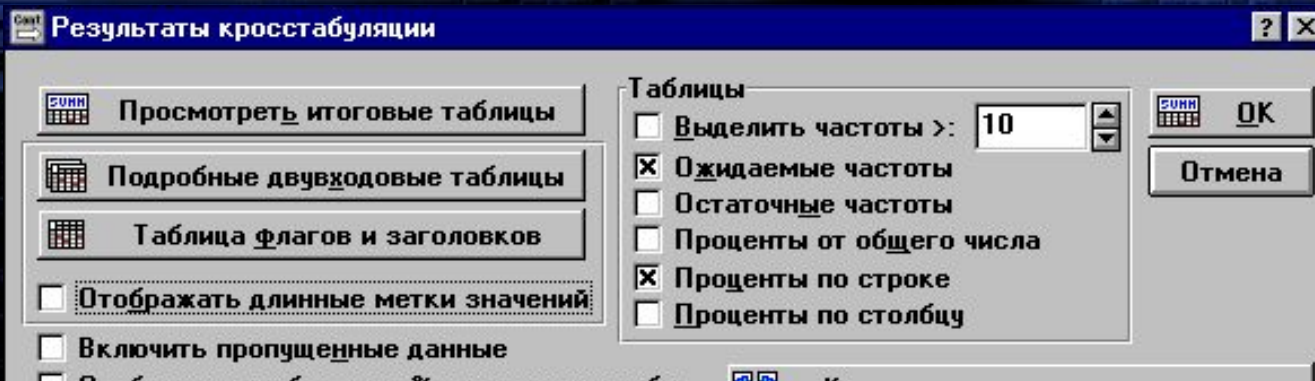
The image shows a screenshot of the SPSS software interface, specifically the "Непараметрические статистики" (Nonparametric Statistics) dialog box. The main dialog box has two radio buttons: "Непараметрические статистики" (selected) and "Подгонка распределения" (Distribution fitting). Below this, a list of nonparametric statistics is shown, with "Критерий знаков" (Sign Test) selected. To the right of the main dialog are buttons for "OK" and "Отмена" (Cancel). Below the main dialog, three sub-dialog boxes are open:

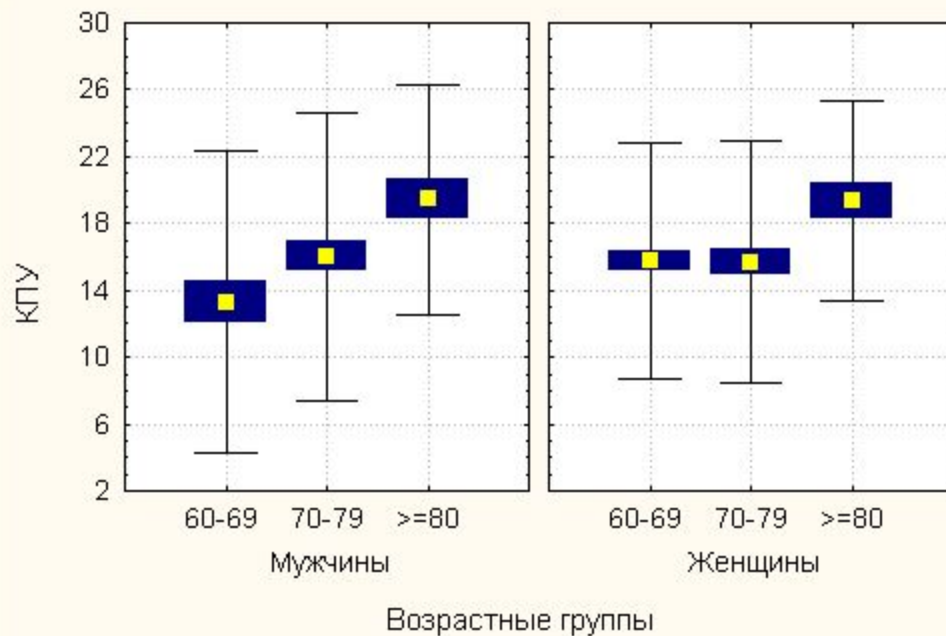
- Критерий знаков (Sign Test):** Shows "Переменные" (Variables) with two lists: "Список переменных 1: ПРИЧ_1-ПРИЧ_4" and "Список переменных 2: ПРИЧ_5-ПРИЧ_8". It also has a "Диаграмма размаха" (Boxplot) option and buttons for "OK", "Отмена", "SELECT CASES", and "CASES".
- ANOVA Краскела-Уоллиса и медианный тест (Kruskal-Wallis and Median Test):** Shows "СОЦ" (SOC) as the variable and has "OK" and "Отмена" buttons.
- Критерий знаков (Sign Test):** (Another instance) with "Переменные" and "Диаграмма размаха" options, and buttons for "OK", "Отмена", "SELECT CASES", and "CASES".

The background shows a blurred view of the SPSS main window and other data-related windows.

Используемые методы анализа

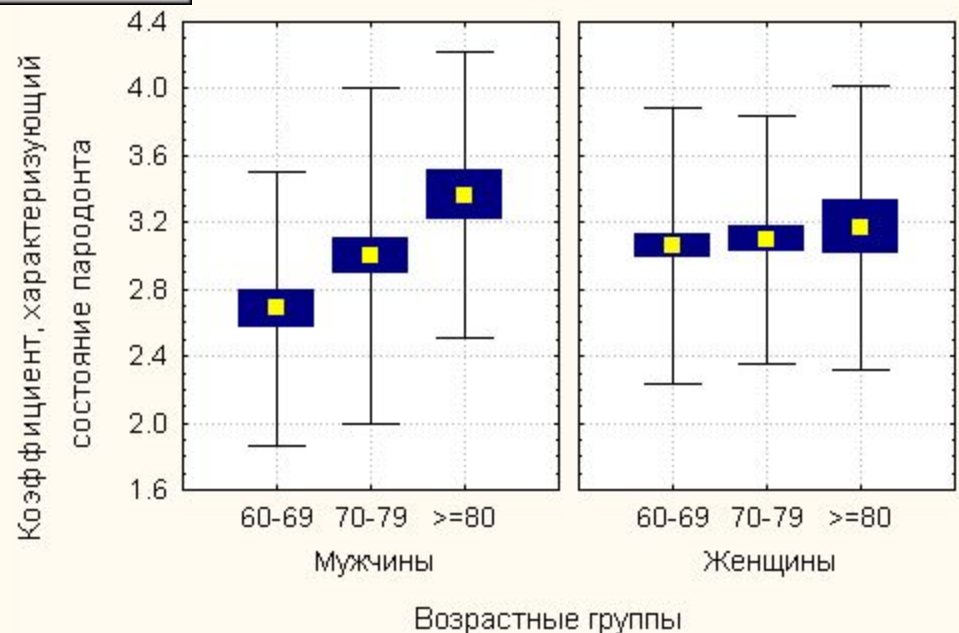
Анализ таблиц сопряженности для категориальных переменных





Характер изменения средних значений индекса КПУ различается в группах мужчин и женщин

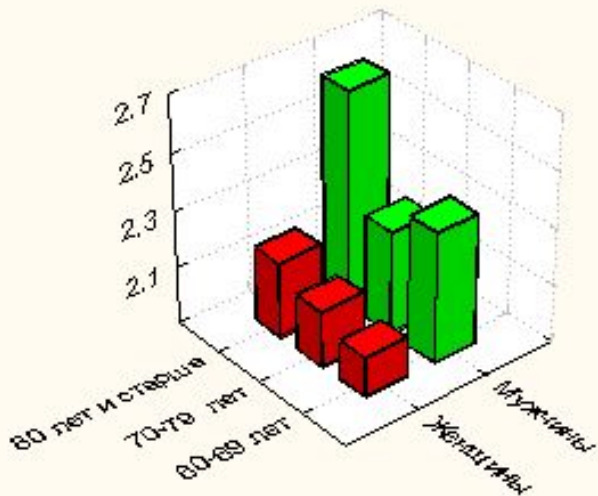
Состояние пародонта у мужчин заметно ухудшается с возрастом, а у женщин практически не изменяется



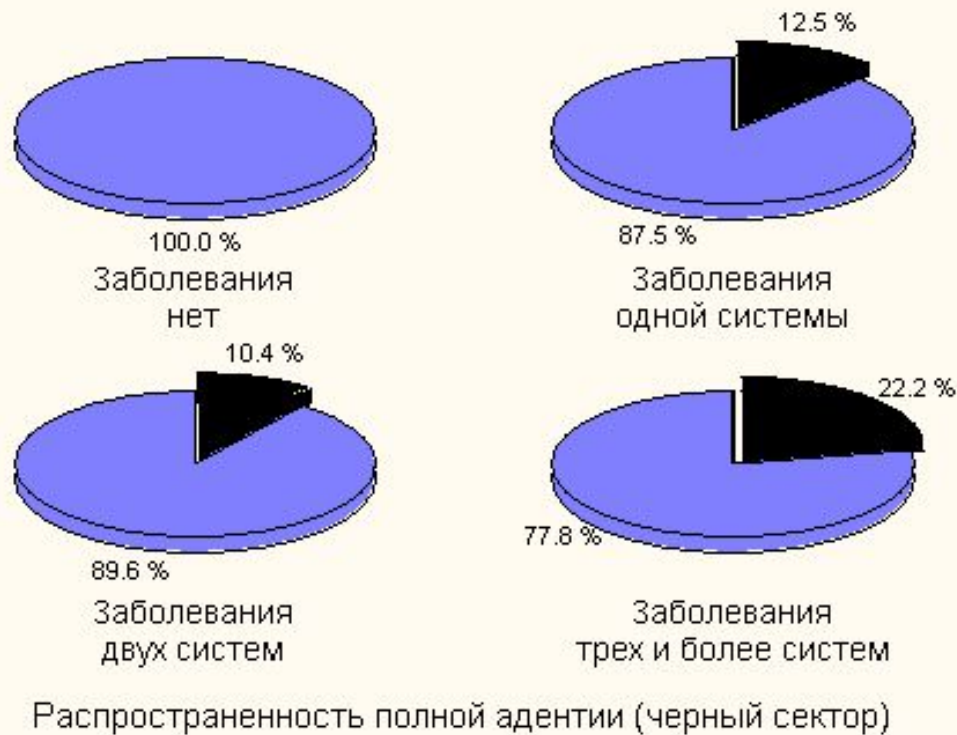
Наблюдается высокая распространенность патологии корня даже среди тех курильщиков, кто выполняет рекомендации врачей.



Гигиенический индекс Федорова-Володкиной



Средние значения гигиенического индекса Федорова-Володкиной заметно больше у мужчин, особенно в самой старшей возрастной группе.



**Статистический анализ таблиц частот
 позволяет установить взаимосвязь
 между числом соматических
 заболеваний и распространенностью
 полной вторичной адентии.**

Пример 2. Исследование прибора для измерения билирубина крови

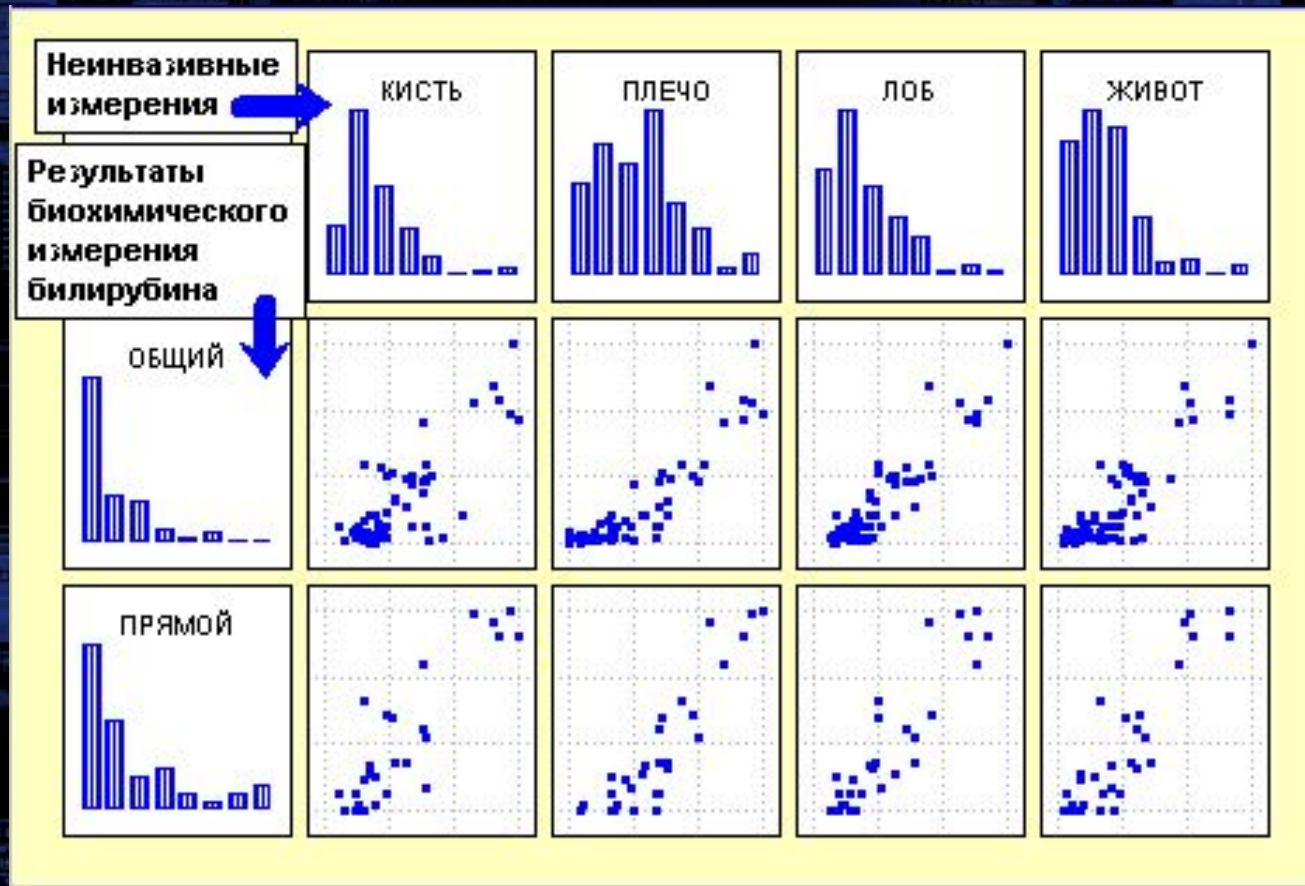
Данные: BIL_DATE.STA 28п * 201н

Текст: Неинвазивные измерения билирубина в сравнении с биохимическим анализом крови

	1 МЕСТО	2 ГРУППА	3 ВОЗРАСТ	4 БИОХИМИЯ	9 ПЛЕЧО	10 ЛОБ
92	Склиф	Взрослые	48.0	6.0	14.4	17.2
93	Склиф	Взрослые	24.0	8.0	17.8	18.2
94	Склиф	Взрослые	50.0	10.0	19.8	22.8
95	Склиф	Взрослые	32.0	12.0	20.8	22.8
96	Детская	Дети	14.0	12.4	34.0	18.0
97	Детская	Дети	12.0	13.5		
98	Детская	Дети	10.0	13.6	27.0	22.0
99	Детская	Дети	7.0	13.9		
100	Склиф	Взрослые	57.0	14.0	22.6	27.8
101	Склиф	Взрослые	51.0	14.0	22.0	25.6
102	Детская	Дети	2.5	15.0	18.0	16.0
103	Детская	Дети	8.0	15.0	40.0	24.0
104	Склиф	Взрослые	21.0	15.0	19.4	23.6
105	Детская	Дети	7.0	15.5	28.0	19.0
106	Склиф	Взрослые	29.0	17.0	21.4	23.6
107	Склиф	Взрослые	34.0	17.0	26.8	26.4
108	Склиф	Взрослые	27.0	17.2	17.4	27.4

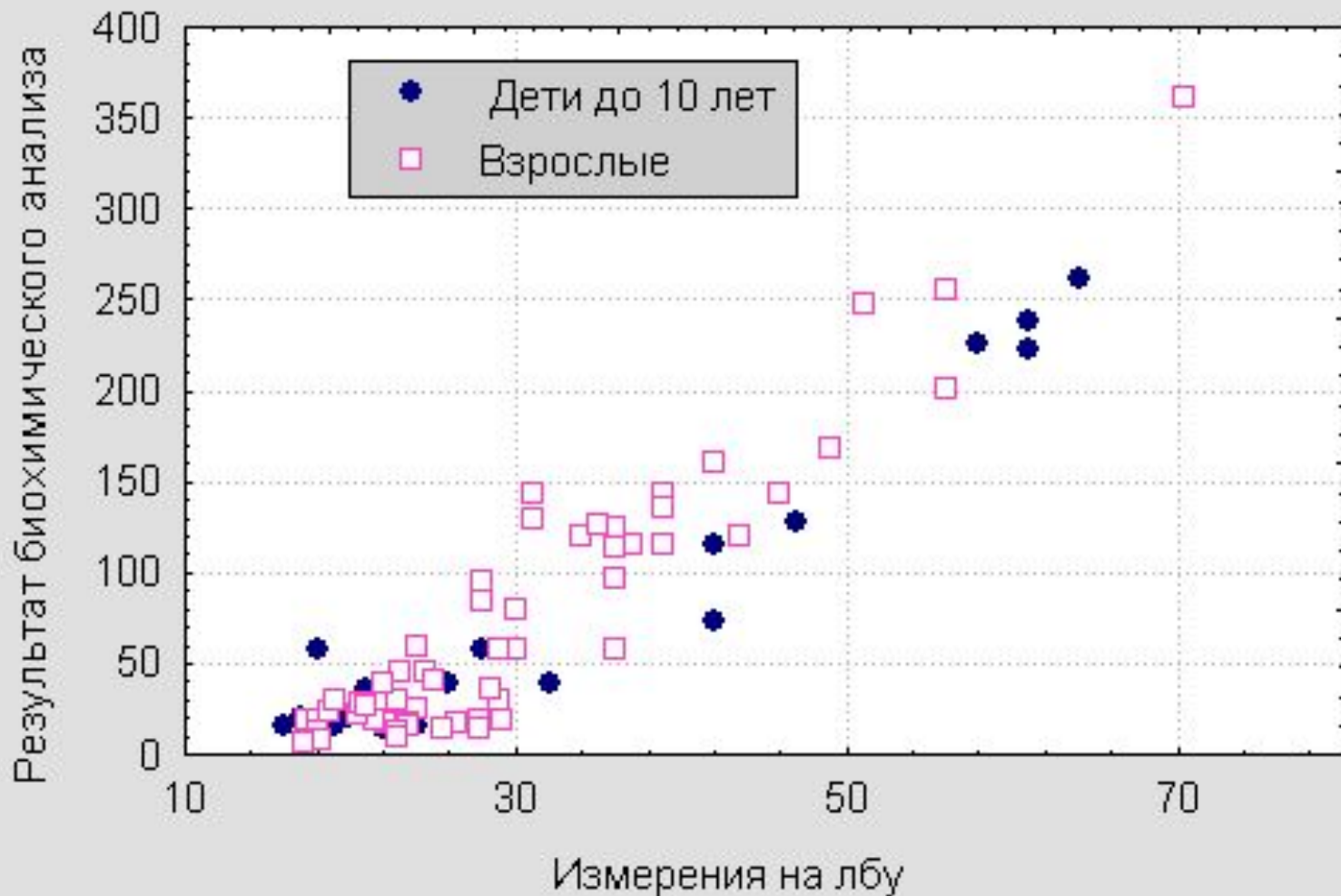
Проводились измерения транскутанного билирубина и биохимический анализ крови у пациентов разных возрастных групп.

Визуализация корреляционных связей



Измерения
транскутанного
билирубина в
разных точках
тела человека
коррелируют с
данными
биохимического
анализа.

Исследование зависимостей для разных возрастных групп



Построение регрессионной модели

С помощью модулей «Множественная регрессия» и «Нелинейное оценивание» построены регрессионные модели, которые позволяют предсказывать результат биохимического анализа по данным транскутанных измерений билирубина в нескольких точках тела.

Определение модели

Переменные

Независимые: 7 9-11 16-17
Зависимые: 4-5

Процедура: Пошаговая с исключением

Свободный член: Добавить в модель

Толерантность: .00010 (Введите 0.0, чтобы задать минимум=1.e-25)

Гребневая регрессия; лямбда: .100

Пошаговая множественная регрессия:
F-включить: 11.00 F-исключить: 10.00
Число шагов: 10

Отображение результатов: Только итоги

Пакетная обработка/печать
 Печатать результаты анализа остатков

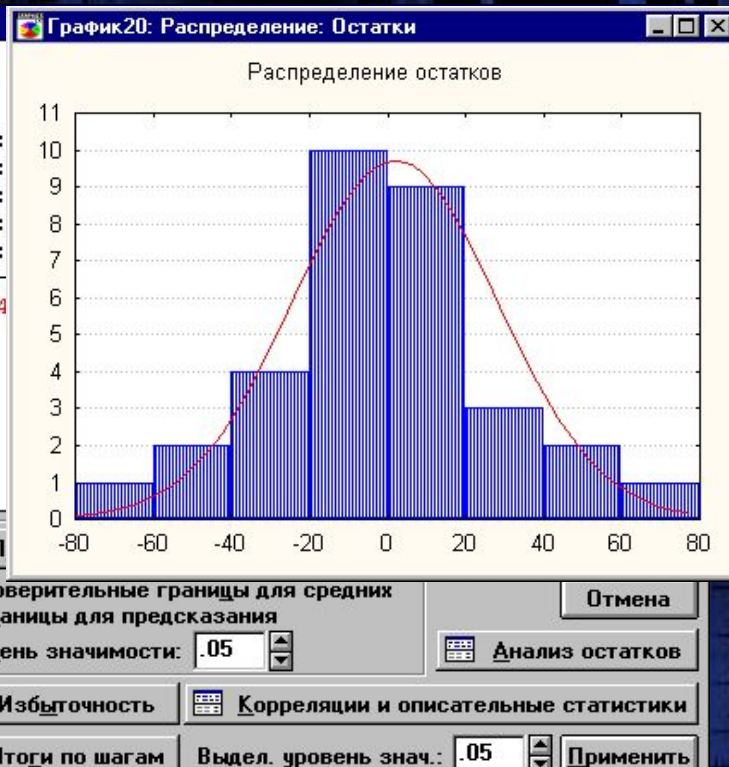
Результаты множественной регрессии

Результаты множ. регрессии (Mar)

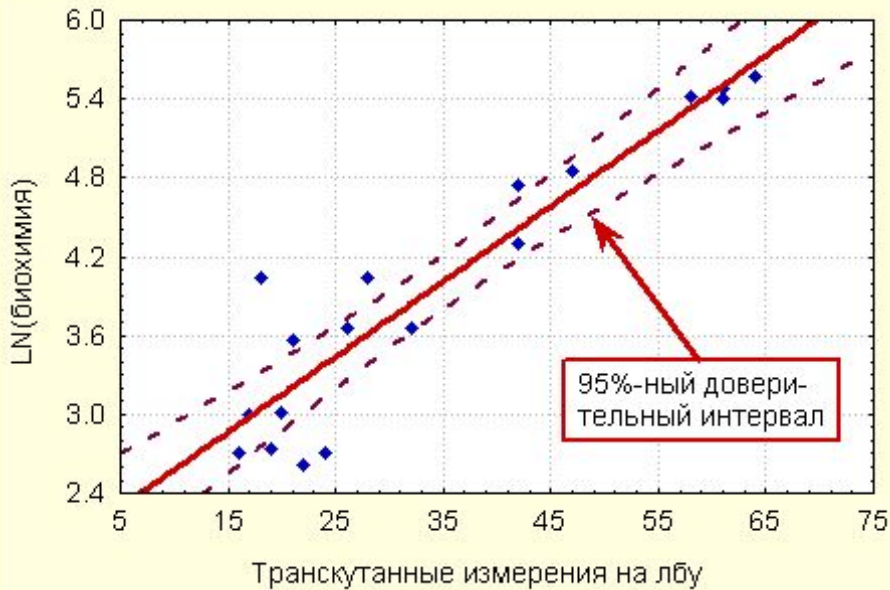
Зав.перем.: **БИОХИМИЯ** Множеств. R:
R2:
Число набл.: 31 скоррект.R2:
Стандартная ошибка оценки:
Своб.член: -107.5301501 Ст.ошибка:

ПЛЕЧО бета=.517 ЛАДОНЬ бета=.4

(выделены значимые бета)

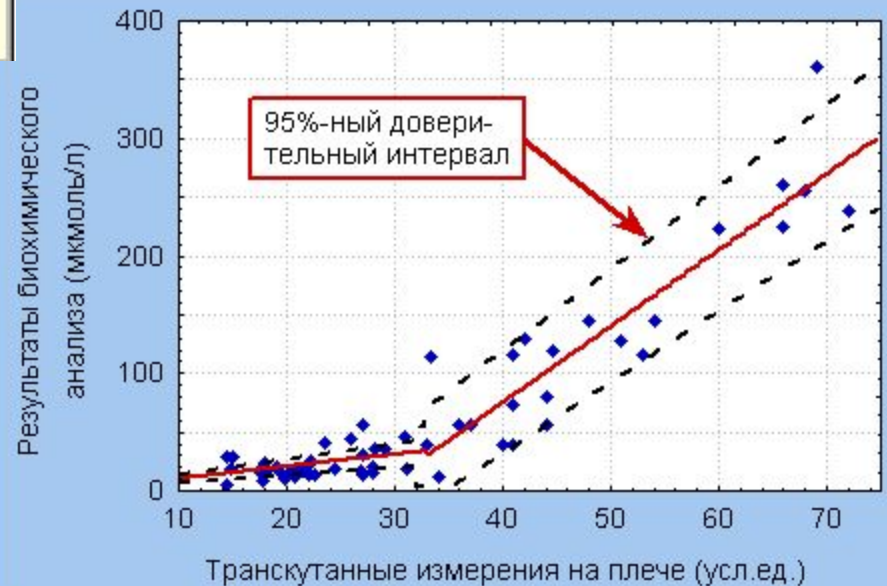


Результаты



Экспоненциальная модель для вычисления общего билирубина у детей до 10 лет включительно.

Кусочно-линейная модель для вычисления общего билирубина у пациентов старше 10 лет.



Пример 3. Предсказание выздоровления пациентов

Предположим, нужно изучить связь между возможностью выздоровления тяжело больного после продолжительной болезни и сроком его пребывания в больнице.

Данные: PATIENTS.STA...

ЧИСЛОВЫЕ ЗНАЧЕНИЯ	1	2
	ДНИ	ПРОГНОЗ
Пациент А	2	54
Пациент В	5	50
Пациент С	7	45
Пациент D	10	37
Пациент E	14	35
Пациент F	19	25
Пациент G	26	20
Пациент H	31	16
Пациент I	34	18
Пациент J	38	13
Пациент K	45	8
Пациент L	52	11
Пациент M	53	8
Пациент N	60	4

Нелинейное оценивание

USER Регрессия, определяемая пользователем

- Логит регрессия
- Пробит регрессия
- Регрессия экспоненциального роста
- Кусочно-линейная регрессия

OK

Отмена

Данные

SELECT CASES Y B

Этапы построения модели

- Выберите метод оценивания.
- Задайте оцениваемую функцию.
- Задайте функцию потерь.
- Запустите алгоритм оценивания.

Оценивание модели

Модель : $v2=b0*\exp(b1*v1)$
Число оцениваемых параметров модели: 2
Функц.потерь: $(OBS-PRED)**2$
Зависимая переменн.: PROGMAS

Независимые переменные: DATA
ИД удаляются построчно
Число допустимых набл.: 15

Метод оценивания: **Квази-ньютоновский**

Асимптотические стандартные ошибки
 Эта для аппроксимации конечными разностями, 1.E-8

Максимальное число итераций: 50
Критерий сходимости: .0001

Начальные значения: **Опред.пользов.**
Начальный размер шага: **.50 для всех параметр.**

Оцениваемая функция и функция потерь

Оцениваемая функция:
 $v2=b0*\exp(b1*v1)$

Функция потерь:
 $L = (OBS-PRED)**2$

ОК
Отмена
Открыть
Сохранить как
Переменные

Оцениваемая функция: 'Завис.перемен.' = выражение; напр.: $v2=constant+param*v3$
Функция потерь: L = выражение; напр.: $L=(obs-pred)**2$; Допустимые операторы:
+ - * ** / < > >= <= <> = () ; Ссылайтесь на переменные по их имени или номеру в файле данных; напр.: $v3=b1*v4$ или $COST=b1*SIZE$
Все нераспознанные имена рассматриваются в качестве параметров; например:
 $v3=const+param*v4$ Используйте стандартную математическую запись; например:
 $v3=b1*v1/3e+2$; Константы: $Pi=3.14...$; $Euler=2.71...$; напр.: $v3=b*Euler*v3$; Функции:
abs arcsin cos exp log log2 log10 sign sin sinh sqrt tan; Логические операции:
истина=1, ложь=0; напр.: $v2=b1*v3*(v1<0)+b2*v3*(v1>=0)$; В тексте функции потерь:
PRED = предсказанное значение, OBS = наблюдаемое значение; По умолчанию,
в функция потерь равна сумме квадратов остатков, т.е.: $L=(OBS-PRED)**2$
Пример 1: $Energy=b0*Mass*b1**2+b2$ $L=v5*(OBS-PRED)**2$
Пример 2: $v4=\exp(a+b1*v4)/(1+\exp(a+b1*v4))$ $L=Weight*abs(OBS-PRED)$

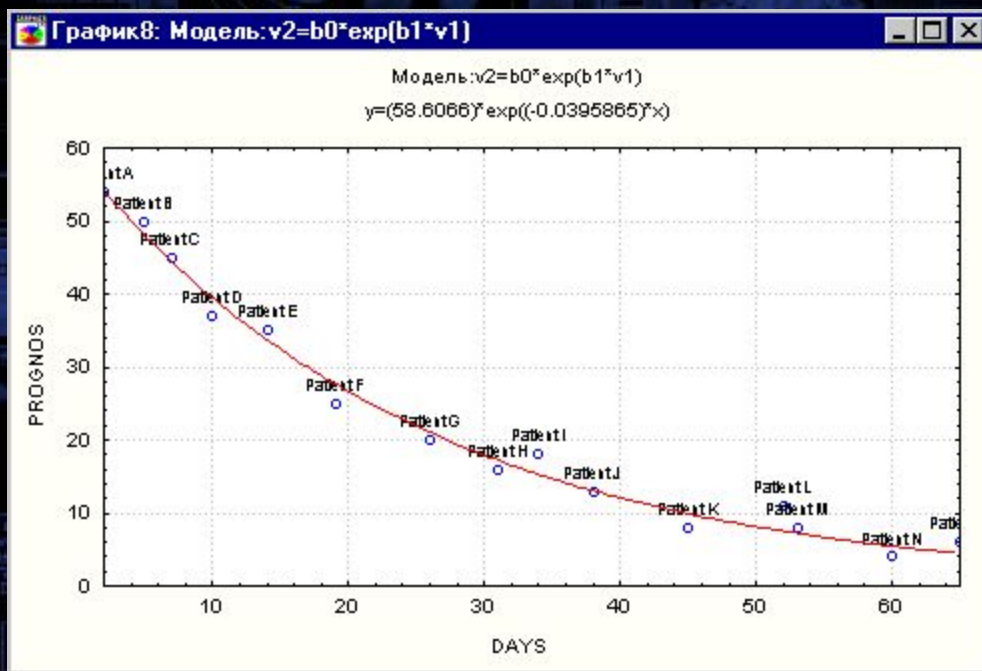
Анализ результатов

- Если процесс сошелся, нажмите ОК.
- Просмотрите оценки параметров.
- Постройте графики.

Оценивание параметров			
Итерация	Потери	Параметры	
* 21	49.4618	58.6340	-.03963
* 22	49.4618	58.6340	-.03963
* 23	49.4593	58.6070	-.03959
* 24	49.4593	58.6066	-.03959
* 25	49.4593	58.6066	-.03959
* 26	49.4593	58.6066	-.03959
* 27	49.4593	58.6066	-.03959

Процесс оценивания параметров сошелся

Отмена



Модель: $y_2 = b_0 \cdot \exp(b_1 \cdot v_1)$ (patients.sta)	
НЕЛИН. ОЦЕНИВ.	Зав.перем: PROGNOS Пот.:(OBS-PRED) ² Ок.потери: 49.45929871 R=.99371 Объясне
N=15	B0
Оценка	58.60662
	B1
	-.039587

Модель позволяет спрогнозировать срок госпитализации!

Пример 4. Анализ выживаемости пациентов после операции

Данные: HEART.STA 11п * 65н

Heart transplant data from Crowley and Hu, stratified

	1	2	3	4	5	6	7	8	9	10	11
ТЕК	MONTH_1	DAY	YEAR_1	MONTH_2	DAY	YEAR_2	CENSORED	AGE	ANTI	MISMA	HOSPITAL
3НВ											
1	JANUARY	6	68	JANUARY	21	68	CENSORED	54	0	1.11	HILLVIEW
2	MAY	2	68	MAY	5	68	CENSORED	40	0	1.66	HILLVIEW
3	AUGUST	31	68	MAY	17	70	COMPLETE	51	0	1.32	HILLVIEW
4	AUGUST	22	68	OCTOBER	7	68	COMPLETE	42	0	.61	ST_AND
5	SEPTEMBER	9	68	JANUARY	14	69	CENSORED	48	0	.36	ST_AND
6	OCTOBER	5	68	DECEMBER	8	68	COMPLETE	54	0	1.89	ST_AND
7	OCTOBER	26	68	JULY	7	72	COMPLETE	54	0	.87	BINER
8	NOVEMBER	22	68	AUGUST	29	69	COMPLETE	49	0	1.12	BINER
9	NOVEMBER	20	68	DECEMBER	13	68	CENSORED	56	0	2.05	HILLVIEW
10	FEBRUARY	15	69	FEBRUARY	25	69	COMPLETE	55	1	2.76	HILLVIEW
11	FEBRUARY	8	69	NOVEMBER	29	71	COMPLETE	43	0	1.13	BINER
12	MARCH	29	69	MAY	7	69	COMPLETE	42	0	1.38	HILLVIEW
13	APRIL	13	69	APRIL	13	71	COMPLETE	58	0	.96	ST_AND
14	JULY	16	69	NOVEMBER	29	69	COMPLETE	52	1	1.62	ST_AND



Данные о выживаемости пациентов в разных госпиталях после операций по пересадке сердца взяты из работы Кроули.



Этапы построения модели


Результаты анализа Каплана-Мейера


Перемен.: число дней, вычисленное по данным
Индикатор цензурирования: CENSORED


Общее число допустимых наблюдений : 65
нецензур. : 29 (44.62%) цензур. : 36 (55.38%)

 **Анализ выживаемости** 

 **Процентили функции выживаемости** 

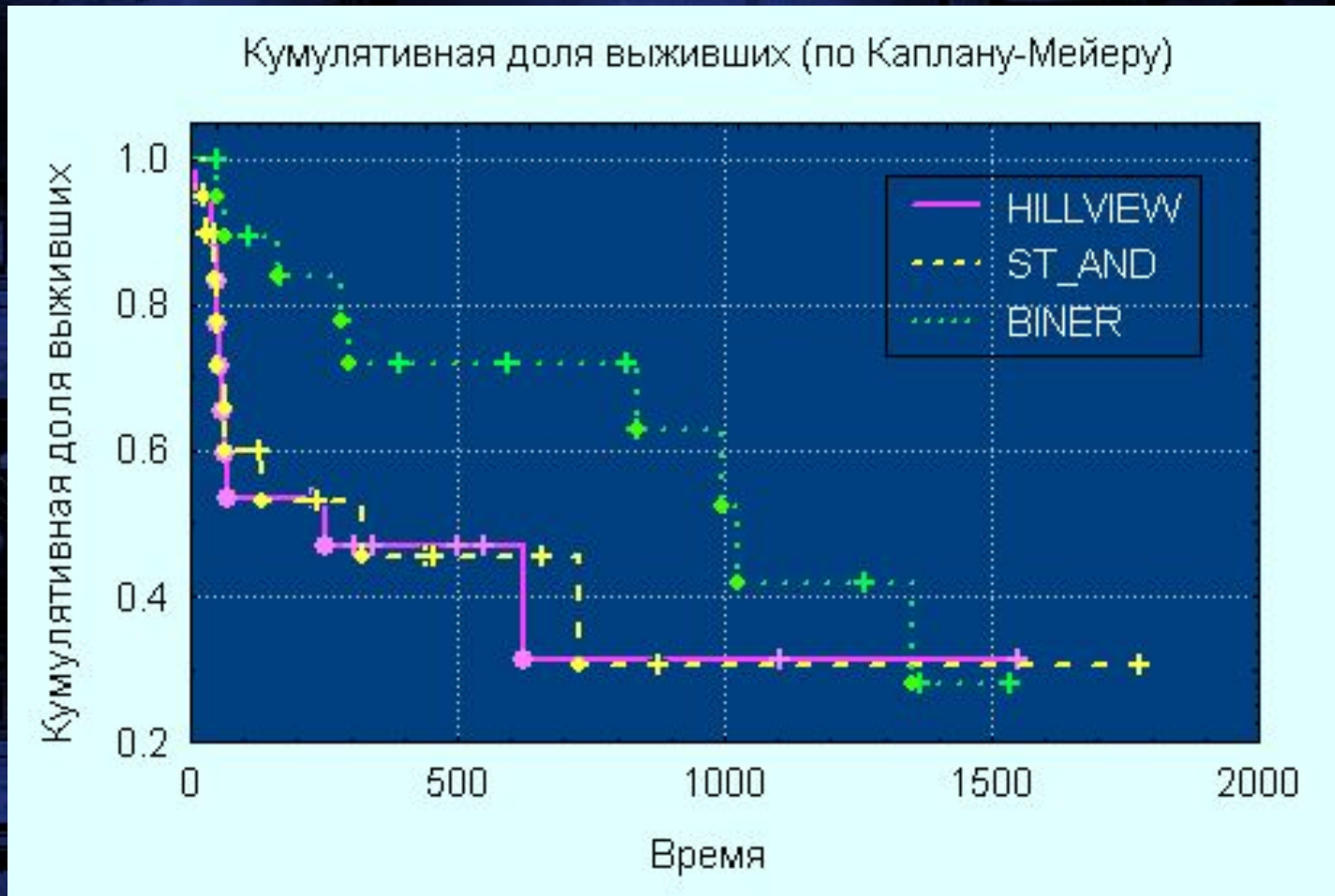
 **График времен жизни и кумулятивной доли выживших**

 **График логарифмов времен жизни и кумул. доли выживших**

 **График лог. времен жизни и лог. кумул. доли выживших**

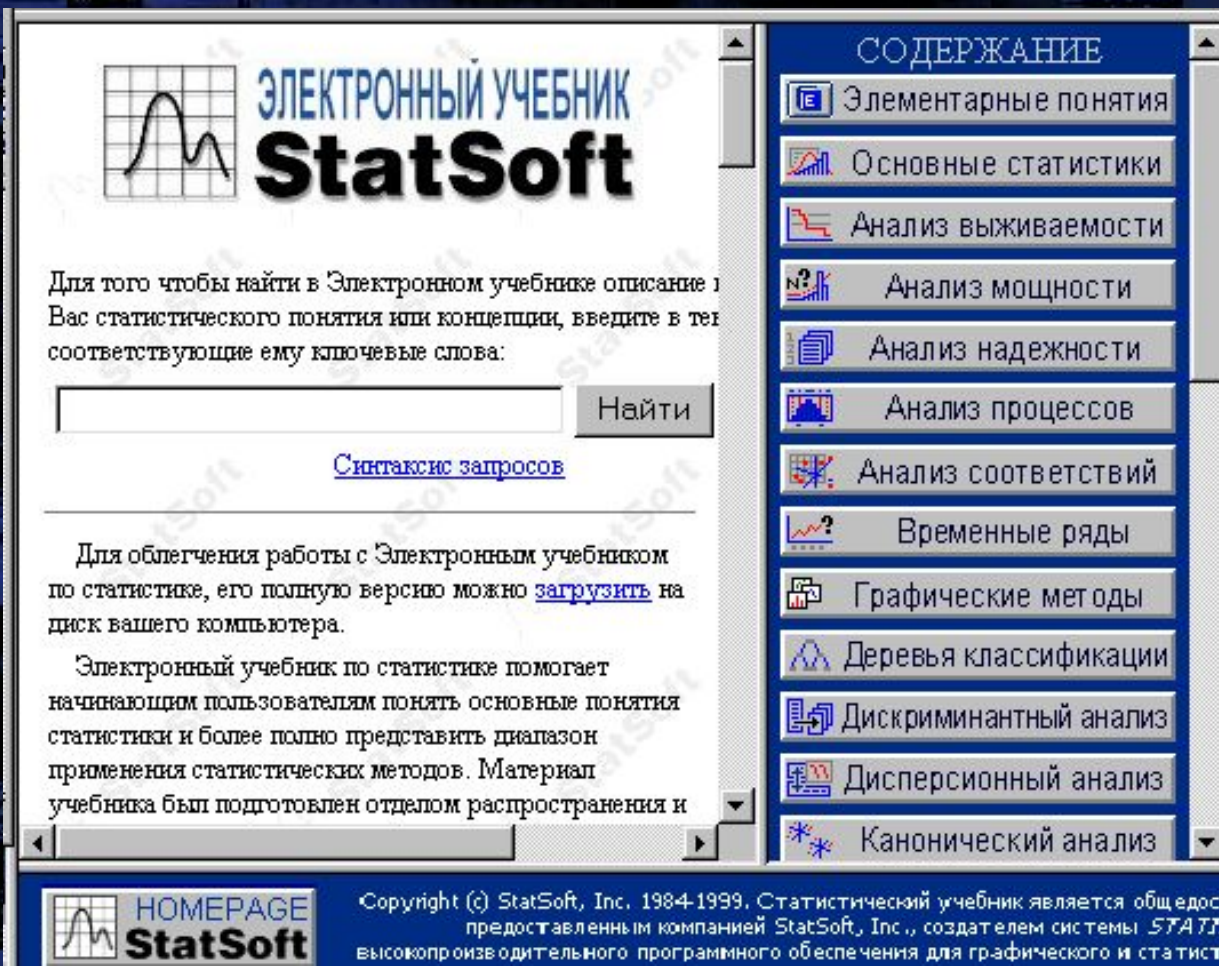
Выбор способа представления результатов

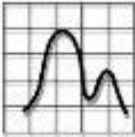
Результаты



Функция выживаемости отдельно для разных госпиталей.

Бесплатный электронный учебник по статистике компании StatSoft доступен в интерактивном режиме и в виде архива на сайте www.statsoft.ru



 **ЭЛЕКТРОННЫЙ УЧЕБНИК**
StatSoft

Для того чтобы найти в Электронном учебнике описание и Вас статистического понятия или концепции, введите в тек соответствующие ему ключевые слова:


[Синтаксис запросов](#)

Для облегчения работы с Электронным учебником по статистике, его полную версию можно [загрузить](#) на диск вашего компьютера.

Электронный учебник по статистике помогает начинающим пользователям понять основные понятия статистики и более полно представить диапазон применения статистических методов. Материал учебника был подготовлен отделом распространения и

СОДЕРЖАНИЕ

- Элементарные понятия
- Основные статистики
- Анализ выживаемости
- Анализ мощности
- Анализ надежности
- Анализ процессов
- Анализ соответствий
- Временные ряды
- Графические методы
- Деревья классификации
- Дискриминантный анализ
- Дисперсионный анализ
- Канонический анализ

 **HOME PAGE**
StatSoft

Copyright (c) StatSoft, Inc. 1984-1999. Статистический учебник является общедоступным, предоставленным компанией StatSoft, Inc., создателем системы STATISTICAL SOFTWARE FOR WINDOWS, высокопроизводительного программного обеспечения для графического и статистического анализа данных.

Учебник содержит разделы по методам статистического анализа данных и предназначен в первую очередь для тех, кто не является специалистом по математической статистике.