МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Тульский государственный университет»

Институт прикладной математики и компьютерных наук
Кафедра «Прикладная математика и информатика»

Утверждено на заседании кафедры
«Прикладная математика и информатика»
24 января 2022 г., протокол № 5

Заведующий кафедрой

_____М.В. Грязев

МЕТОДИЧЕСКИЕ УКАЗАНИЯ
по выполнению лабораторных работ
по дисциплине (модулю)
«Компьютерные и информационные технологии
в науке и производстве»

основной профессиональной образовательной программы
высшего образования – программы магистратуры

по направлению подготовки
01.04.02 Прикладная математика и информатика

с направленностью (профилем)
Искусственный интеллект в кибербезопасности

Форма обучения: очная

Идентификационный номер образовательной программы: 010402-02-22

Тула 2022 год

Разработчик методических указаний

Родионова Г.А., доцент каф. ПМиИ, к.т.н., доцент

(ФИО, должность, ученая степень, ученое звание)

(подпись)

Лабораторная работа №1
по дисциплине Компьютерные и информационные технологии

Пакет STATISTICA. Обработка данных. Вероятностные характеристики

Цель и задачи работы: Изучение способов задания исходных данных в статистическом пакете. Определение вероятностных характеристик. Подбор закона распределения.

1 Задание

- Ввести исходные данные по своему варианту.
- Найти вероятностные характеристики вариационного ряда.
- Построить частотные диаграммы.
- Подобрать закон распределения вариационного ряда (с использованием удаления подозрительных наблюдений).

2 Выполнение

1) Введение исходных данных.

Электронная таблица в пакете STATISTICA состоит из строк и столбцов. Столбцы таблицы называются *Variables* (переменные), а строки — *Cases* (случаи, наблюдения).

✓ Создадим новый файл с данными при помощи меню *File* (файл). Выберем команду *New* из меню *File*. В появившемся диалоговом окне выделим вкладку *Spreadsheet* (таблица) и укажем *Namber of variables* (число переменных) – 1, и *Namber of cases* (число случаев) – 72. Нажмем ОК.

✓ Для сохранения файла выберем команду *Save* (сохранить) из меню *File*.

✓ Для ввода данных в таблицу установим указатель мыши на ячейку, в которую необходимо ввести данные. Щелкнем мышью, для того чтобы сделать эту ячейку активной, командой *Copy* (копировать) скопируем столбцы с их содержимым из файла задания (вариант 4). Командой *Paste* (вставить) вставим столбцы с данными в таблицу.

2) Нахождение вероятностных характеристик вариационного ряда.

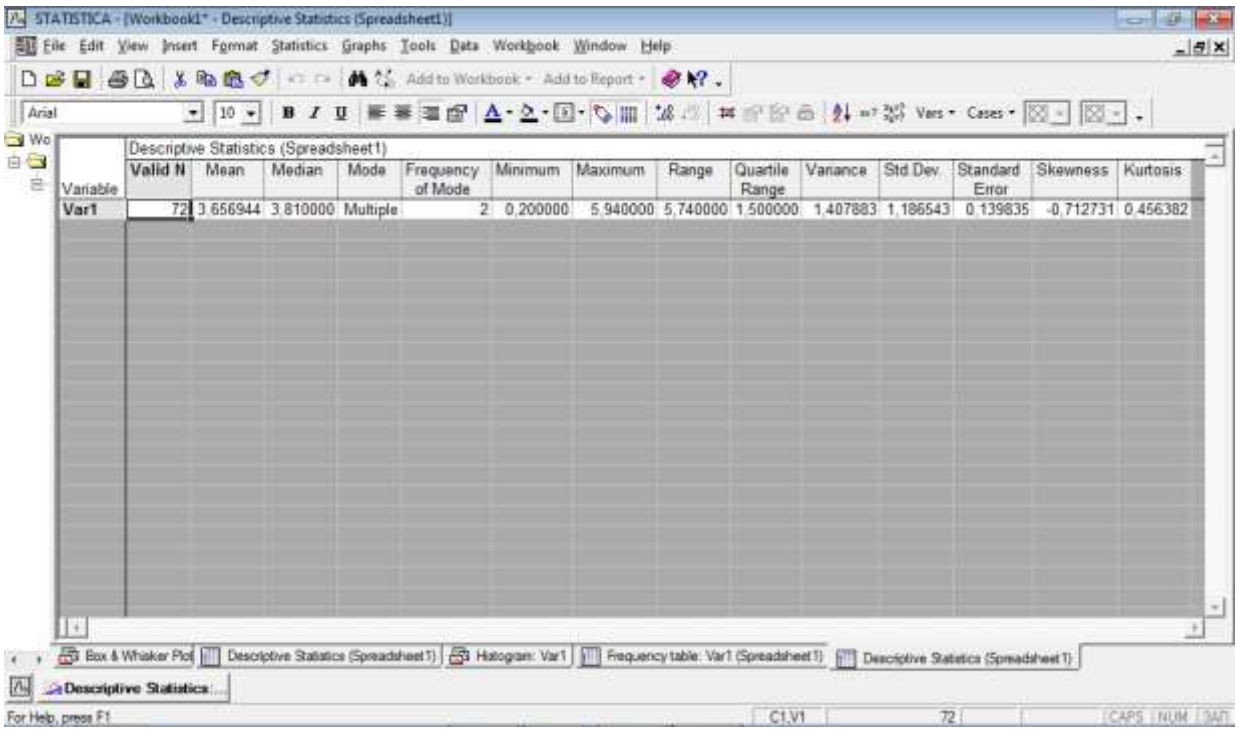
Для нахождения основных вероятностных характеристик вариационного ряда используем модуль *Descriptive statistics* (описательные статистики).

✓ Для запуска программы в верхнем меню *Statistics* выберем команду *Basic Statistic Tables* (основные статистики/таблицы).

✓ В появившемся меню выберем команду *Descriptive statistics* (описательные статистики).

✓ Для выбора переменной нажмем кнопку *Variables* и в открывшемся окне щелкнем на имени переменной *Var1* .

✓ Для просмотра результатов нажмем кнопку *Summary: Descriptive statistics*. Откроется таблица с основными статистиками.



Variable	Valid N	Mean	Median	Mode	Frequency of Mode	Minimum	Maximum	Range	Quartile Range	Variance	Std Dev.	Standard Error	Skewness	Kurtosis
Var1	72	3,656944	3,810000	Multiple	2	0,200000	5,940000	5,740000	1,500000	1,407883	1,186543	0,139835	-0,712731	0,456382

Mean (среднее арифметическое) — показывает центральное положение (центр) переменной и рассматривается совместно с доверительным интервалом. Доверительный интервал представляет интервал значений вокруг оценки, где с данным уровнем доверия находится «истинное» (неизвестное) среднее генеральной совокупности. Вычисление доверительных интервалов основывается на предположении нормальности наблюдаемых величин.

Альтернативной оценкой среднего являются *median* (медиана) и *mode* (мода). Медиана — это квантиль, соответствующая вероятности 0,5, т.е. это значение, которое разбивает выборку на две равные части по количеству элементов.

Мода — это значение переменной, соответствующее наибольшей частоте появления переменной в выборке.

Std.dev. (стандартное отклонение) — это корень квадратный из суммы квадратов отклонений значений переменной от среднего значения, деленное на n - 1.

Std.eir.of mean (стандартная ошибка среднего) — это стандартное отклонение, деленное на корень квадратный из объема выборки.

Varience (коэффициент вариации) — это отношение стандартного отклонения к среднему.

Minimum (минимум) или *Maximum* (максимум) — это соответственно минимальное или максимальное значение выборки.

Range (размах) — это разность между максимальным и минимальным значениями выборки.

Quartiles range (квартальный размах) равен разности значений верхней и нижней квартилей, т.е. это интервал, содержащий медиану, в который попадает 50% выборки.

Skewness (асимметрия) — это мера симметричности распределения. Если распределение симметрично, то асимметрия равна нулю, если асимметрия существенно отличается от 0, то распределение несимметрично. Нормальное и равномерное распределения абсолютно симметричны.

Kurtosis (эксцесс) — мера остроты пика распределения. Если распределение нормальное, то эксцесс равен 0. Если эксцесс положителен, то пик заострен, если отрицателен, то пик закруглен.

3) Построение частотных диаграмм.

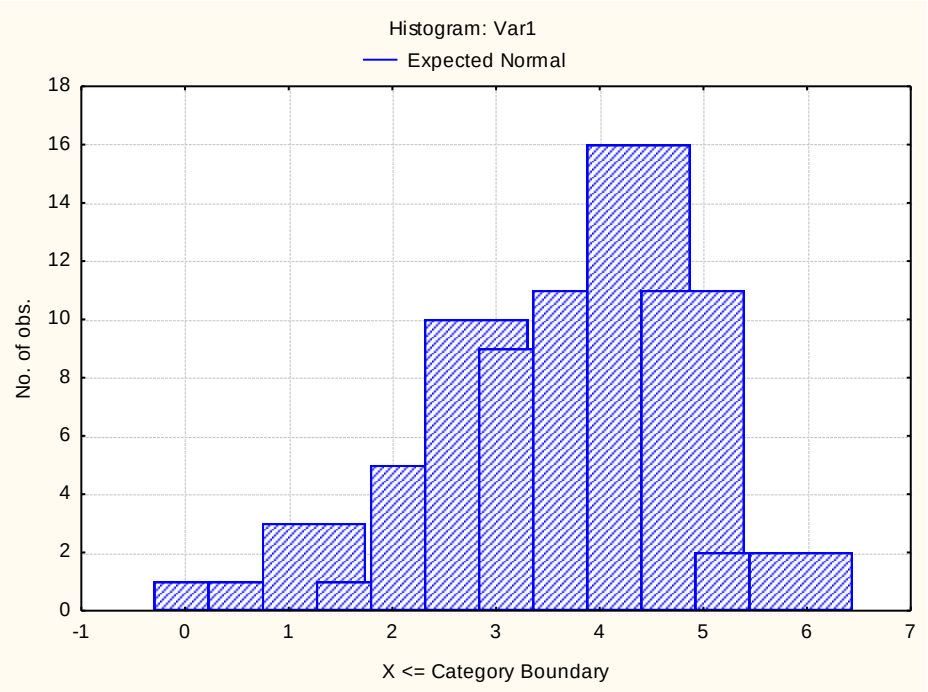
Таблицы частот, или одновходовые таблицы, представляют собой простейший метод анализа категориальных (номинальных) переменных.

✓ Чтобы открыть диалоговое окно *Frequency tables* (таблицы частот), из стартовой панели *Basic Statistics/Tables* выберем команду *Frequency tables*. С помощью кнопки *Variables* на вкладке *Quick* открывается диалоговое окно выбора одного списка переменных для анализа.

Summary: Frequency tables. Открываются итоговые таблицы частот для выбранных переменных.

	Frequency table: Var1 (Spreadsheet1)			
From To	**Count**	Cumulative Count	Percent	Cumulative Percent
-,060909<x<=,4609091	1	1	1,38889	1,3889
,4609091<x<=,9827273	1	2	1,38889	2,7778
,9827273<x<=1,504545	3	5	4,16667	6,9444
1,504545<x<=2,026364	1	6	1,38889	8,3333
2,026364<x<=2,548182	5	11	6,94444	15,2778
2,548182<x<=3,070000	10	21	13,88889	29,1667
3,070000<x<=3,591818	9	30	12,50000	41,6667
3,591818<x<=4,113636	11	41	15,27778	56,9444
4,113636<x<=4,635455	16	57	22,22222	79,1667
4,635455<x<=5,157273	11	68	15,27778	94,4444
5,157273<x<=5,679091	2	70	2,77778	97,2222
5,679091<x<=6,200909	2	72	2,77778	100,0000
Missing	0	72	0,00000	100,0000

Histograms. Программа строит последовательность гистограмм для выбранной переменной.



4) Подбор закона распределения вариационного ряда (с использованием удаления подозрительных наблюдений)

При обработке экспериментальных данных иногда возникает необходимость аппроксимировать эмпирическое распределение тем или иным известным законом распределения. Для этой цели в STATISTICA предназначен модуль *Distribution Fitting* (подгонка распределения).

✓ Чтобы запустить модуль *Distribution Fitting*, необходимо в главном меню *Statistics* выбрать одноименную команду. В открывшемся окне *Distribution Fitting* указажем природу случайной величины, т.е. *Continuous Distribution* (непрерывная), а также предполагаемый закон распределения, которому случайная величина подчиняется. Выберем, например, Normal и нажмем ОК.

✓ В открывшемся окне укажем переменную *Var1*. На вкладке *Quick* есть две кнопки: *Summary: Observed and expected distributions* (результат: наблюдаемые и ожидаемые распределения) и *Plot of Observed and expected distributions* (график наблюдаемых и ожидаемых распределений).

✓ Если нажать на кнопку *Summary: Observed and expected distributions*, то программа отобразит численные характеристики в виде таблицы. Каждая строка этой таблицы характеризует интервал, в который попадают значения исследуемой переменной.

Variable: Var1, Distribution: Normal (Spreadsheet1)
Chi-Square = 5,68727, df = 5 (adjusted) , p = 0,33785

Upper Boundary	Observed Frequency	Cumulative Observed	Percent Observed	Cumul. % Observed	Expected Frequency	Cumulative Expected	Percent Expected	Cumul. % Expected	Observed-Expected
<= 0,00000	0	0	0,00000	0,0000	0,07402	0,07402	0,10280	0,1028	-0,07402
0,50000	1	1	1,38889	1,3889	0,20677	0,28079	0,28718	0,3900	0,79323
1,00000	2	3	2,77778	4,1667	0,62428	0,90507	0,86706	1,2570	1,37572
1,50000	2	5	2,77778	6,9444	1,58212	2,48719	2,19739	3,4544	0,41788
2,00000	1	6	1,38889	8,3333	3,36568	5,85287	4,67456	8,1290	-2,36568
2,50000	4	10	5,55556	13,8889	6,01031	11,86319	8,34766	16,4766	-2,01031
3,00000	10	20	13,88889	27,7778	9,00996	20,87315	12,51384	28,9905	0,99004
3,50000	10	30	13,88889	41,6667	11,33859	32,21174	15,74804	44,7385	-1,33859
4,00000	9	39	12,50000	54,1667	11,97869	44,19043	16,63708	61,3756	-2,97869
4,50000	17	56	23,61111	77,7778	10,62370	54,81413	14,75514	76,1307	6,37630
5,00000	8	64	11,11111	88,8889	7,90962	62,72375	10,98558	87,1163	0,09038
5,50000	6	70	8,33333	97,2222	4,94359	67,66734	6,86610	93,9824	1,05641
6,00000	2	72	2,77778	100,0000	2,59375	70,26109	3,60243	97,5848	-0,59375
6,50000	0	72	0,00000	100,0000	1,14235	71,40343	1,58659	99,1714	-1,14235
< Infinity	0	72	0,00000	100,0000	0,59657	72,00000	0,82857	100,0000	-0,59657

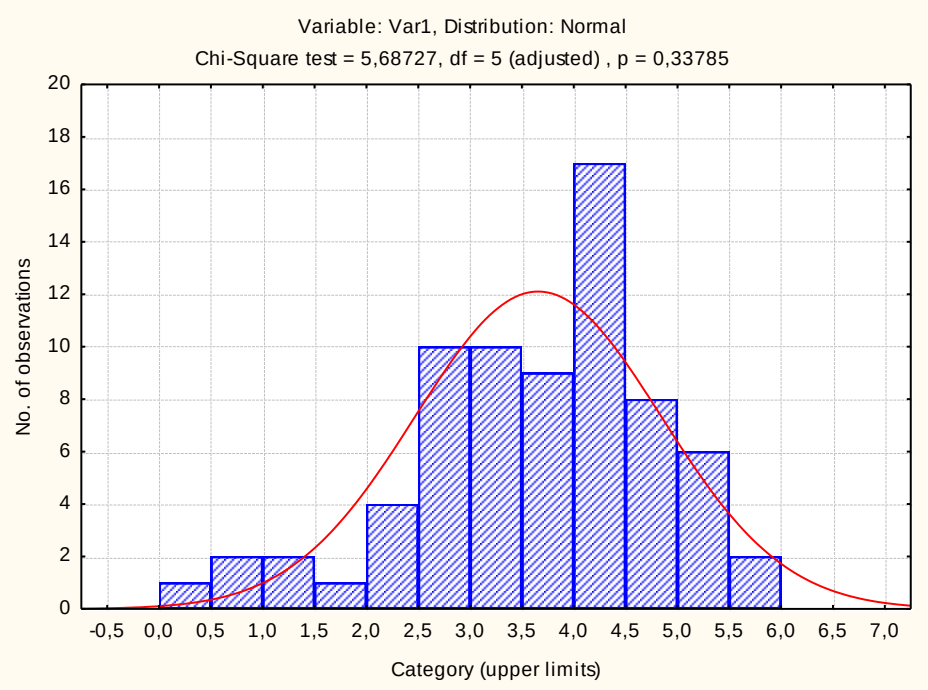
В первом столбце *Observed Frequency* (наблюдаемая частота) для каждого рассмотренного интервала указано количество значений, попавших в этот интервал.

Во втором столбце *Cumulative Observed* (совокупный наблюдаемый) для каждого интервала приведено количество значений, попавших в этот и все предшествующие интервалы (накопленные частоты).

В третьем и четвертом столбцах *Percent Observed* (процент наблюдаемый) и *Cumul %* (суммарный процент) указаны те же величины, что и в предыдущих двух, но исчисленные в процентах.

В пятом столбце *Frequency Expected* (ожидаемая частота) даны теоретические частоты, соответствующие нормальному распределению.

При нажатии на кнопку *Plot of Observed and expected distributions* будет построена кривая теоретического закона распределения и гистограмма эмпирического, построенного по имеющимся данным.



Над гистограммой выведен заголовок, в котором указана анализируемая переменная, предполагаемый закон распределения, а также три числовых параметра.

Первый параметр — это значение критерия χ^2. Чем меньше это значение, тем больше вероятность того, что проверяемая случайная величина имеет предполагаемый закон распределения.

Второй параметр *df*—число степеней свободы. Определяется как $df = n-l-1$, где n — число интервалов, на которые разбит диапазон изменения случайной величины; l — число оцениваемых параметров распределения.

Третий параметр *p*-уровень значимости критерия, который определяет вероятность ошибки при отклонении гипотезы о нормальности. Так как вероятность ошибки достаточно велика, примерно 0,33785 (что значительно больше 0,05), гипотезу о соответствии закона распределения нормальному принимаем.

Лабораторная работа №2
по дисциплине Компьютерные и информационные технологии
Пакет STATISTICA. Дисперсионный анализ.

Цель и задачи работы: провести дисперсионный анализ заданных наблюдений**.**

1. Задание

По заданному варианту произвести дисперсионный анализ.

ВАРИАНТ ЗАДАНИЯ №1

1. Результаты испытания трех сортов пшеницы

Сорт пшеницы	Урожайность по участкам, ц/га
1	18,3; 19,4; 20,8; 17,4; 20,3; 19,8; 20,3; 20,0; 19,3; 17,8
2	17,9; 22,4; 23,1; 20,5; 17,3; 19,7; 20,1; 19,4; 20,5
3	19,1; 23,6; 21,5; 22,7; 22,8; 23,4; 22,1; 20,8; 19,3; 23,1; 23,0

3 Выполнение

Сравнение средних является одним из способов выявления зависимостей между переменными. Наиболее общим методом сравнения средних является дисперсионный анализ — *ANOVA (Analysis of Variance)*.

Дисперсионный анализ можно определить как статистический метод, предназначенный для оценки влияния различных факторов на результат эксперимента, а также для последующего планирования экспериментов. В дисперсионном анализе можно исследовать зависимость количественного признака (зависимой переменной) от одного или нескольких качественных признаков (факторов).

Первоначально (1918 г.) дисперсионный анализ был разработан английским математиком — статистиком Фишером для обработки результатов агрономических опытов по выявлению условий получения максимального урожая различных сортов сельскохозяйственных культур. По числу факторов, влияние которых исследуется, различают однофакторный и многофакторный дисперсионный анализ.

В дисперсионном анализе анализируются не сами суммы квадратов отклонений, а *так называемые средние квадраты, являющиеся несмещенными оценками соответствующих дисперсий, которые получаются делением сумм квадратов отклонений на соответствующее число степеней свободы.* Число степеней свободы определяется как общее число наблюдений минус число связывающих их уравнений.

В однофакторном дисперсионном анализе на одну количественную переменную Y оказывает влияние один фактор (один качественный показатель), наблюдаемый на k уровнях, т.е. имеем k выборок для переменной Y.

Наблюдаемые данные обозначим y_{ij}, где i - индекс уровня (i = 1, ..., k), j - индекс наблюдения на i-м уровне, j = 1, ..., n_i (на каждом уровне может быть свое число наблюдений n_i).

Общее число опытов (наблюдений) $n = \sum_{i=1}^{k} n_i$.

По данным y_{ij} можно определить следующие характеристики:

$\bar{y}_i$ – среднее значение переменной Y на i-м уровне $\bar{y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}$;

$\bar{Y}$ – среднее значение переменной Y по всем значениям

$\bar{Y} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij} = \frac{1}{n}\sum_{i=1}^{k} \bar{y}_i n_i;$

S – сумму квадратов отклонений наблюдений от общего среднего

$S = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{Y})^2;$

S_F – сумму квадратов отклонений средних групповых значений от общего среднего значения $\bar{Y}$

$S_F = \sum_{i=1}^{k}(\bar{y}_i - \bar{Y})^2 n_i;$

S_{OCT} – остаточную сумму квадратов отклонений

$S_{OCT} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2.$

Справедливо соотношение: $S = S_F + S_{OCT}$.

Соответствующие суммы квадратов отклонений, отнесенные к числу степеней свободы, представляют собой несмещенные оценки соответствующих дисперсий.

Далее выполняются следующие действия.

Вычисление общей дисперсии $\sigma^2_{общ} = \frac{S}{n-1} = \frac{1}{n-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{Y})^2.$

Вычисление факторной дисперсии $\sigma^2_F = \frac{S_F}{k-1} = \frac{1}{k-1}\sum_{i=1}^{k}(\bar{y}_i - \bar{Y})^2 n_i.$

Вычисление остаточной дисперсии

$$\sigma^2_{OCT} = \frac{S_{OCT}}{n-1} = \frac{1}{n-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2.$$

В дисперсионном анализе проверяется гипотеза H_0 о равенстве средних групповых значений количественного показателя:

$$(H_0: \ \bar{y}_1 = \bar{y}_2 = ... = \bar{y}_k).$$

Критерием для проверки этой гипотезы является соотношение факторной дисперсии и остаточной дисперсии:

$$F_H = \frac{\sigma^2_F}{\sigma^2_{OCT}}.$$

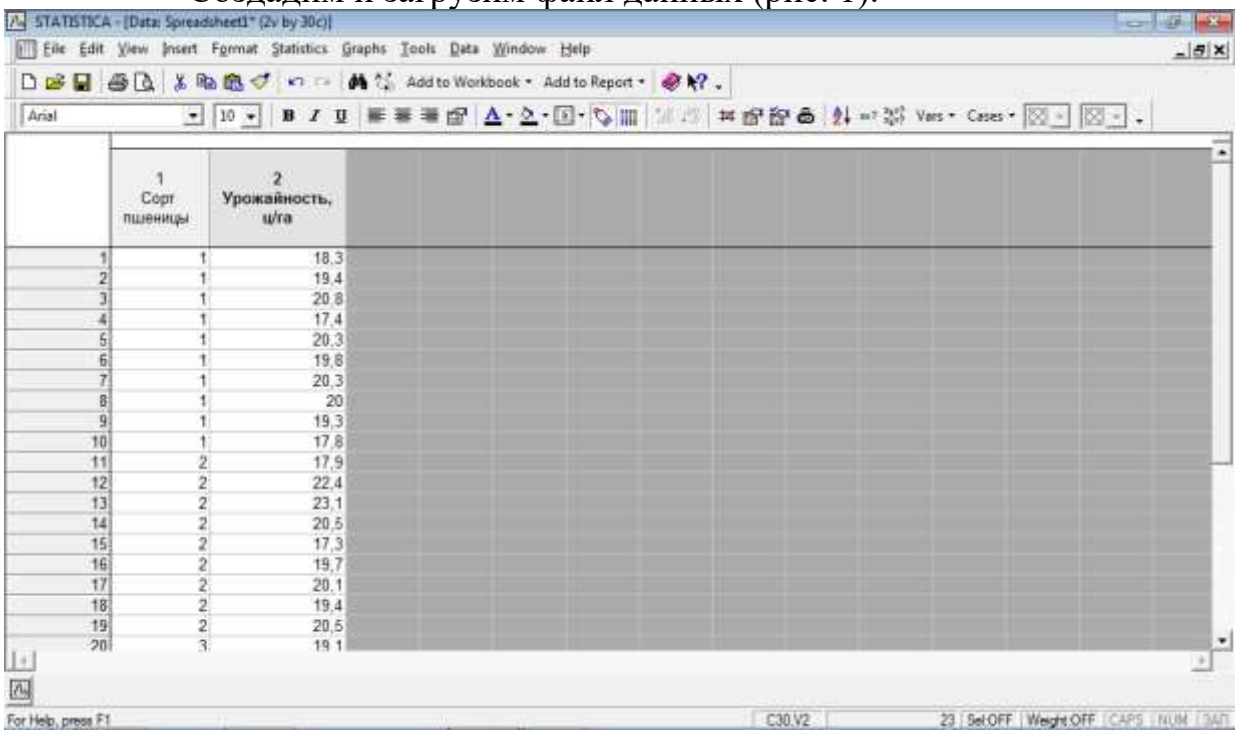
Проведем дисперсионный анализ по **традиционному «классическому» методу Фишера.** В рамках однофакторного дисперсионного анализа предусмотрена процедура *One–Way ANOVA*, которая заключается в анализе влияния одного качественного фактора на количественную переменную.

Качественным показателем, влияющим на <u>количественную переменную (урожайность по участкам)</u>, **является номер сорта пшеницы**.

Задача состоит в проверке гипотезы H_0: $\bar{y}_1 = \bar{y}_2 = \bar{y}_3$, где $\bar{y}_k$ – математическое ожидание случайной величины – урожайности пшеницы *k*-того сорта, *k = 1, 2, 3*. Число уровней фактора *l = 3*; общий объем выборки *n = 30*.

Выполним следующие действия:

- Создадим и загрузим файл данных (рис. 1).



	1 Сорт пшеницы	2 Урожайность, ц/га
1	1	18,3
2	1	19,4
3	1	20,8
4	1	17,4
5	1	20,3
6	1	19,8
7	1	20,3
8	1	20
9	1	19,3
10	1	17,8
11	2	17,9
12	2	22,4
13	2	23,1
14	2	20,5
15	2	17,3
16	2	19,7
17	2	20,1
18	2	19,4
19	2	20,5
20	3	19,1

Рис. 1. Фрагмент файла данных

- Для запуска программы в верхнем меню *Statistics* выберем команду *ANOVA*, что переводится как анализ вариаций или дисперсионный анализ. Появится стартовая панель *General ANOVA/MANOVA*. В списке *Type of analysis* (вид анализа) выберем *One-way ANOVA* (однофакторный дисперсионный анализ);
- Откроется диалоговое окно *ANOVA/MANOVA One–Way ANOVA*. На вкладке *Quick* нажмем кнопку *Variables*.
- Перенесем переменную *Урожайность* в список зависимых *(dependent)* переменных, а переменную *Сорт_пшеницы* – в поле *Categorical factor* (фактор).
- Щелкнем по кнопке ОК, откроется диалоговое окно *ANOVA Results 1* (результаты анализа) с набором вкладок, которые позволяют всесторонне отобразить результаты анализа в виде таблиц и графиков.
- На вкладке *Quick* нажмем кнопку *All effects/Graphs* (все эффекты/графики). Данный диалог *Table of All Effects* (таблицы всех эффектов) (рис.2) содержит результаты и используется для просмотра выбранных из данной таблицы эффектов в виде графиков средних или таблиц.

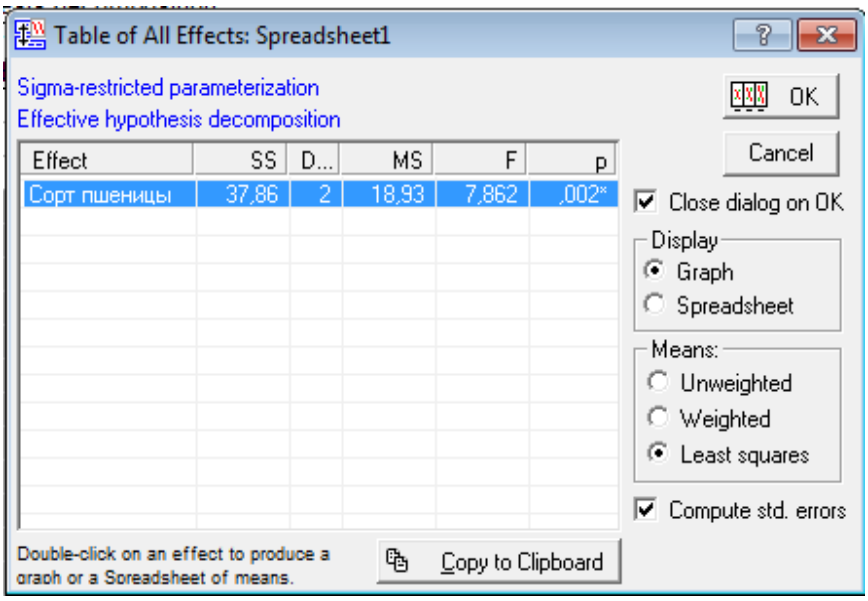


Рис. 2. Таблица эффектов

- Значимые эффекты (p<0,05) в таблице *Table of All Effects* помечены *. Видно, что гипотеза о равенстве средних для *Сорта пшеницы* неверна. Выделим опцию *Spreadsheet* (таблица) в рамке *Display* (отображать) и два раза щелкнем на эффекте *Сорт пшеницы* или, выделив эффект *Сорт пшеницы*, нажмем ОК. Появится таблица со значениями средних зависимой переменной и другими статистиками (рис.3).

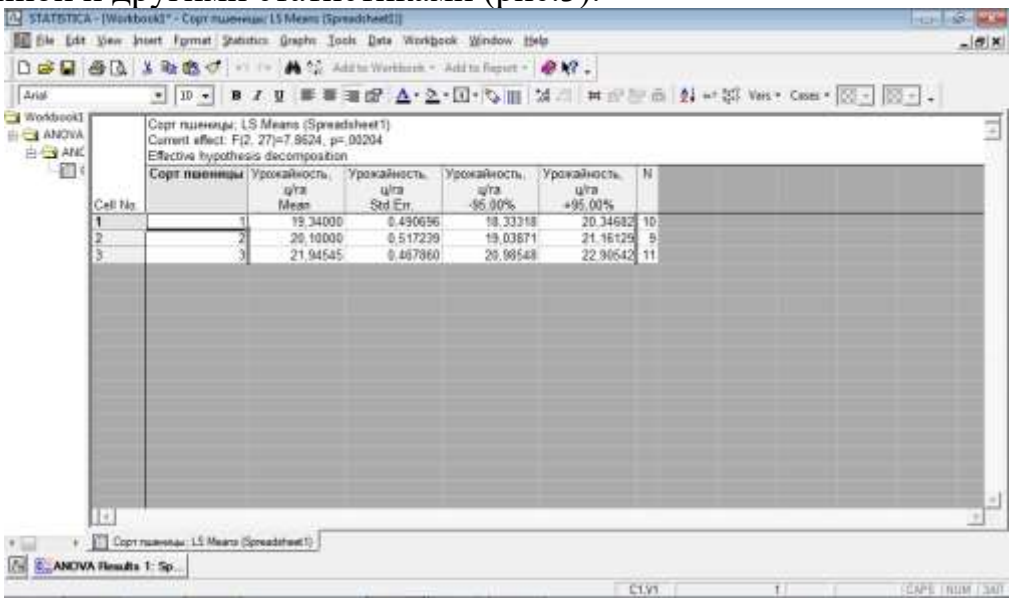


Рис. 3. Таблица значений средних.

- Вернемся в окно *Table of All Effects* и выделим опцию *Graph* (график) в рамке *Display*, нажмем ОК. Программа построит график средних переменной *Сорт пшеницы*.

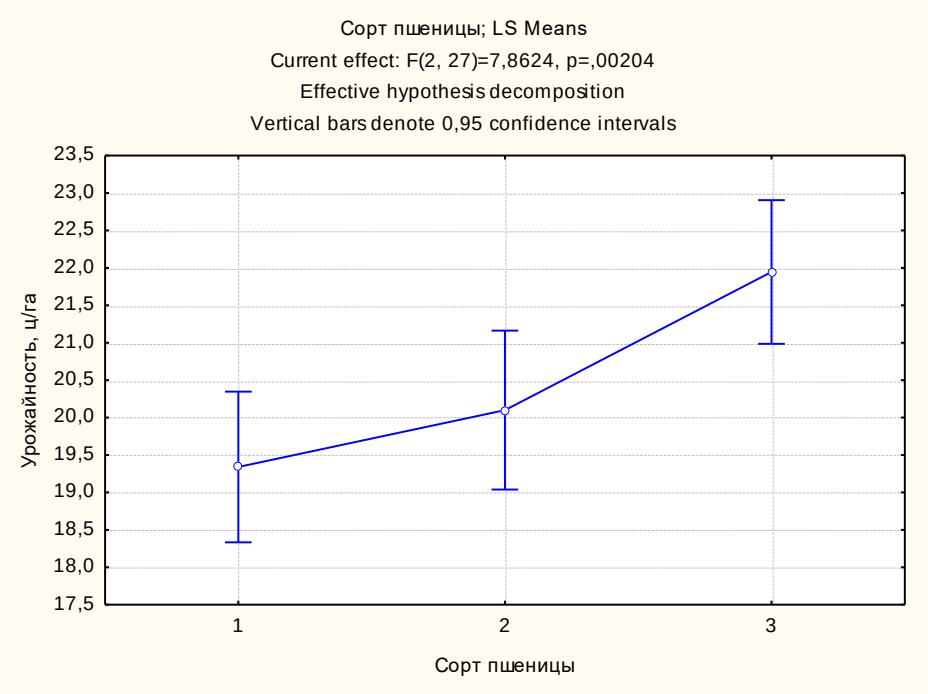


Рис. 4. Таблица эффектов

- Из графика видно, что средняя урожайность пшеницы сорта №3 превосходит урожайность пшеницы сорта №2 и значительно превосходит урожайность пшеницы сорта №1.
- Для проверки предположений, лежащих в основе метода дисперсионного анализа, воспользуемся вкладкой *Assumptions* (предположения) в окне *ANOVA Results 1*. На вкладке представлены различные критерии проверки гипотезы однородности дисперсий (критерий Кохрана, Хартли, Бартлетта, критерий Левена, М критерий Бокса), графические средства проверки соответствия закона распределения переменной нормальному закону (гистограммы, диаграммы рассеяния, нормальные вероятностные графики).
- Выберем эффект *Сорт пшеницы* и нажмем кнопку *Histograms*. В появившемся окне укажем группу, если нужно проанализировать распределение внутри каждой группы. Если выбрать *All* (все), то программа построит (рис. 5) гистограмму частот для всех групп. Видно, что общее распределение соответствует нормальному закону.

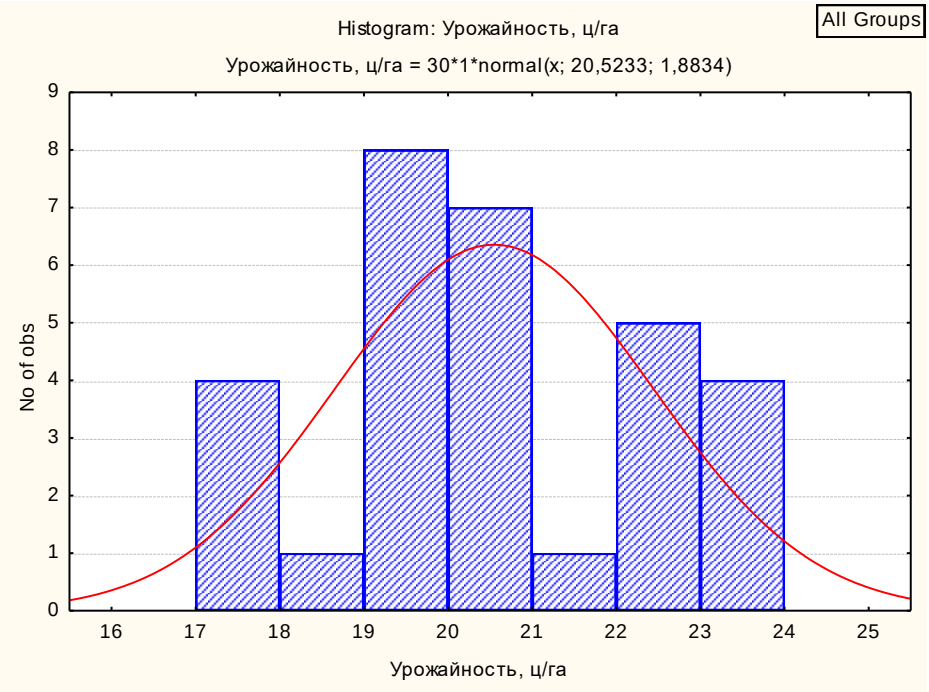


Рис. 5. Гистограмма частот.

- Дополнительным условием применимости дисперсионного анализа является отсутствие корреляции между средними и стандартными отклонениями. На вкладке *Assumptions* нажмем кнопку *Plot means vs.stdLdeviattons*.

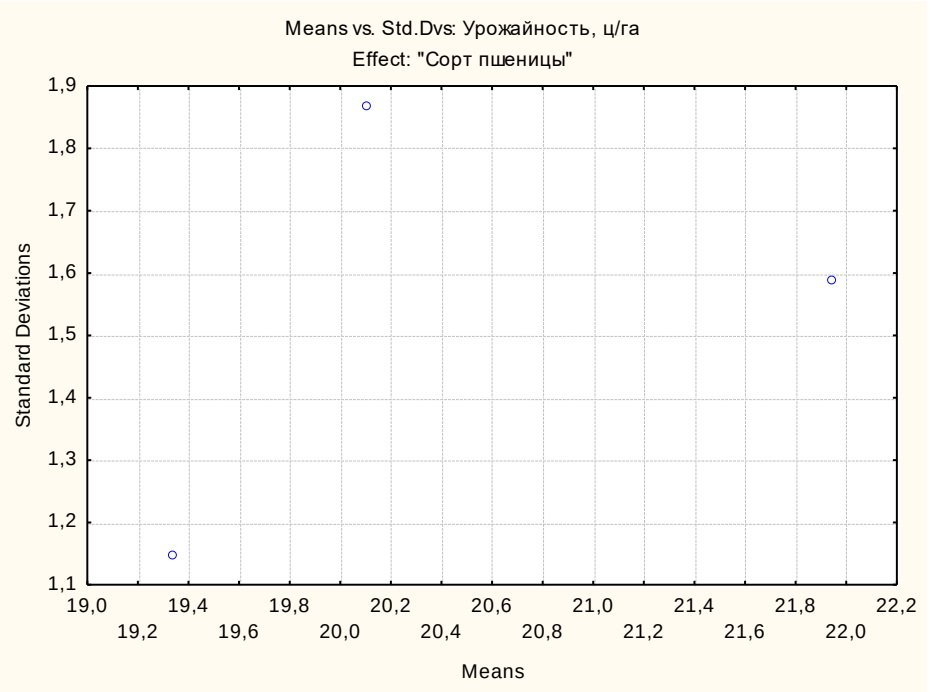
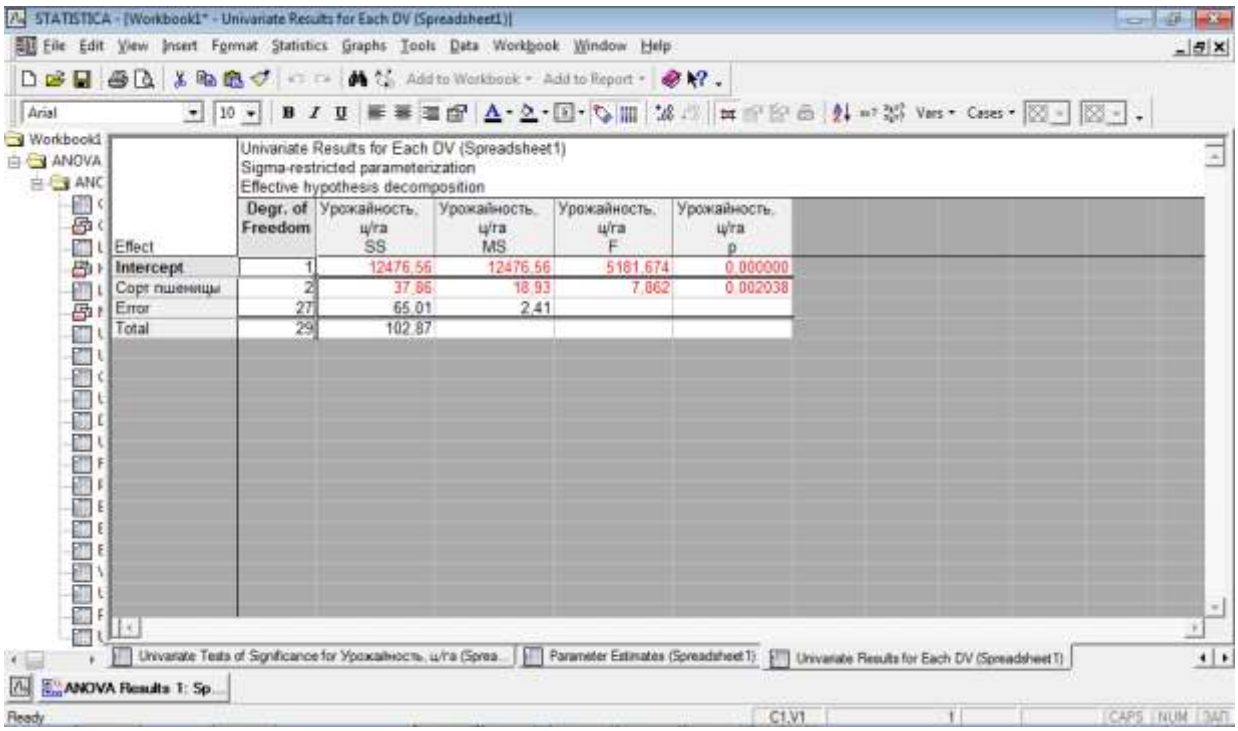


Рис. 6. Диаграмма рассеяния.

- Из диаграммы рассеяния, изображенной на рис. 6, видно, что средние и стандартные отклонения коррелируют незначительно.
- На вкладке *Summary* нажмем кнопку *Univariate Results* для получения данных о суммах квадратов отклонений (рис.7).



Univariate Results for Each DV (Spreadsheet1)
Sigma-restricted parameterization
Effective hypothesis decomposition

Effect	Degr. of Freedom	Урожайность, ц/га SS	Урожайность, ц/га MS	Урожайность, ц/га F	Урожайность, ц/га p
Intercept	1	12476,56	12476,56	5181,674	0,000000
Сорт пшеницы	2	37,86	18,93	7,862	0,002038
Error	27	65,01	2,41		
Total	29	102,87			

Рис. 7. Таблица *Univariate Results*.

Таким образом, имеем следующие характеристики:

- $n_1 = 10;\ n_2 = 9;\ n_3 = 11$.
- Средние значения переменной на *i*–м уровне $\bar{y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}$.
- $\bar{y}_1 = 19,34;\ \bar{y}_2 = 20,1;\ \bar{y}_3 = 21,9455$.
- Среднее значение переменной по всем значениям:
- $\bar{y} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij} = 20,5233$.
- Сумму квадратов отклонений всех наблюдений от общего среднего
- $S = \sum_{i=1}^{k}\sum_{j=1}^{n_i} \left[(y]_{ij} - \bar{y}\right)^2 = 102,874$. .
- Сумму квадратов отклонений средних групповых значений
- $S_F = \sum_{i=1}^{k} \left[(\bar{y}]_i - \bar{y}\right)^2 n_i = 37,862$.
- Остаточную сумму квадратов отклонений
- $S_{OCT} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} \left[(y]_{ij} - \bar{y}_i\right)^2 = 65,011$.

Отметим справедливость соотношения $S = S_F + S_{OCT}$.

Нулевая гипотеза принимается при. $F_H < F_{кр} = F(1-\alpha, k-1, n-k)$

Проверим гипотезу для уровня 0,05: $F(1-0,05, 2, 27)\ F_{кр} = 3,35;\ \ F_H = 7,862$.

Имеем, что $F_H > F_{кр}$; нулевую гипотезу отвергаем и делаем ***вывод, что различие между сортами пшеницы имеет существенное влияние на урожайность***. Кроме того, основные условия применимости дисперсионного анализа выполнены, что подтверждает достоверность полученных результатов.

Лабораторная работа №3
по дисциплине Компьютерные и информационные технологии
Пакет STATISTICA. Анализ временных рядов.

Цель и задачи работы: ознакомиться с методикой анализа временных рядов и научиться применять данную методику.

Задание

По заданному варианту проанализировать временной ряд, сделать прогноз.

Выполнение

Под временным (динамическим) рядом подразумевается последовательность наблюдений некоторого признака X (случайной величины) в последовательные равноотстоящие моменты t. Отдельные наблюдения называются уровнями ряда и обозначаются x_t, t= 1,..., n.

При исследовании временного ряда выделяются несколько составляющих:

$$x_t = u_t + \gamma_t + c_t + \varepsilon_t, t = 1,....,n, \qquad (1)$$

где u_t — тренд, плавно меняющаяся компонента, описывающая чистое влияние долговременных факторов (убыль населения, уменьшение доходов и т.д.);

γ_t — сезонная компонента, отражающая повторяемость процессов в течение не очень длительного периода (дня, недели, месяца и т.д.);

c_t — циклическая компонента, отражающая повторяемость процессов в течение длительных периодов времени свыше одного года;

ε_t — случайная компонента, отражающая влияние не поддающихся учету и регистрации случайных факторов.

Вариационный ряд рассматривается как одна из реализаций случайной величины X, а временной ряд — как одна из реализаций случайного процесса X(t) (t — время). Но в отличие от вариационного ряда, члены временного ряда, как правило, не являются статистически независимыми и одинаково распределенными.

Можно выделить следующие этапы анализа временных рядов:

1. Графическое представление и анализ поведения временного ряда.
2. Выделение и анализ детерминированных составляющих ряда.
3. Сглаживание и фильтрация (удаление низко- или высокочастотных составляющих) временного ряда.
4. Исследование случайной составляющей временного ряда, построение и проверка адекватности математической модели ее описания.
5. Прогнозирование поведения временного ряда на основе проведенных исследований.

Задача прогнозирования состоит в том, чтобы по значениям наблюдений, собранных к данному моменту, определить значения в следующие моменты.

В модуле *Time Series/Forecasting* (временные ряды и прогнозирование) представлен широкий набор процедур, реализующих все перечисленные этапы анализа. Для активизации модуля в меню *Statistica* выберем

последовательность команд *Advanced Linear/Nonlinear Models → Time Series/Forecasting*.

Кнопка *OK (transformations, autocorrelations, plots)* (преобразования, автокорреляции, графики) открывает специальное окно для преобразования ряда.

Анализ временного ряда начнем с графической иллюстрации ряда. Выберем вкладку *Review Series* (просмотреть серии). Верхняя часть окна — информационная, здесь записывается имя ряда и его преобразования.

С помощью кнопки *Plot* рядом с кнопкой *Review highlighted variables* (просмотр выделенной переменной) построим график ряда (рис. 1), который позволит произвести предварительную, визуальную оценку данных ряда.

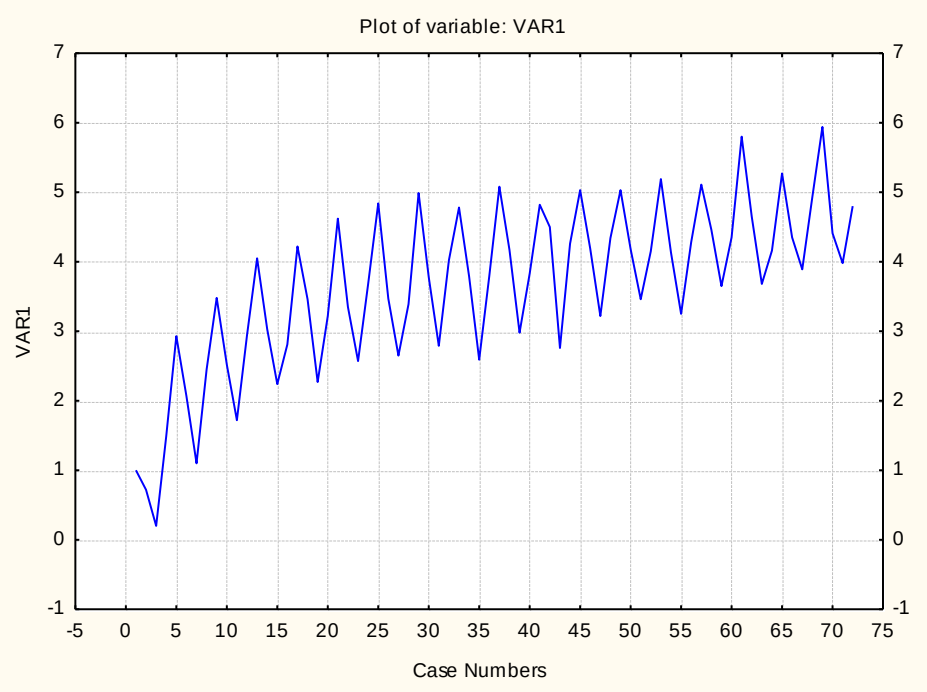


Рис. 1. График временного ряда.

Для построения «качественного» прогноза уровни исследуемого показателя обязательно должны быть сопоставимы и однородны, а для выявления тенденций, кроме этого, устойчивы и полны, т.е. количество наблюдений должно быть достаточно велико.

Из построенного графика видно, что ряд удовлетворяет всем этим требованиям, при этом нет резких скачков, просматривается тренд ряда, который выражается в плавном увеличении показателей, и сезонность, проявляемая в периодичности увеличения и уменьшения показателей.

Для нахождения такого рода периодичностей используем спектральный анализ. Для этого воспользуемся диалогом *Spectral (Fourier) analysis* (Фурье, спектральный анализ), нажав на кнопку *Cancel*. Нажмем кнопку *OK (одномерный анализ Фурье)*. Откроется диалоговое окно с результатами анализа (рис. 2).

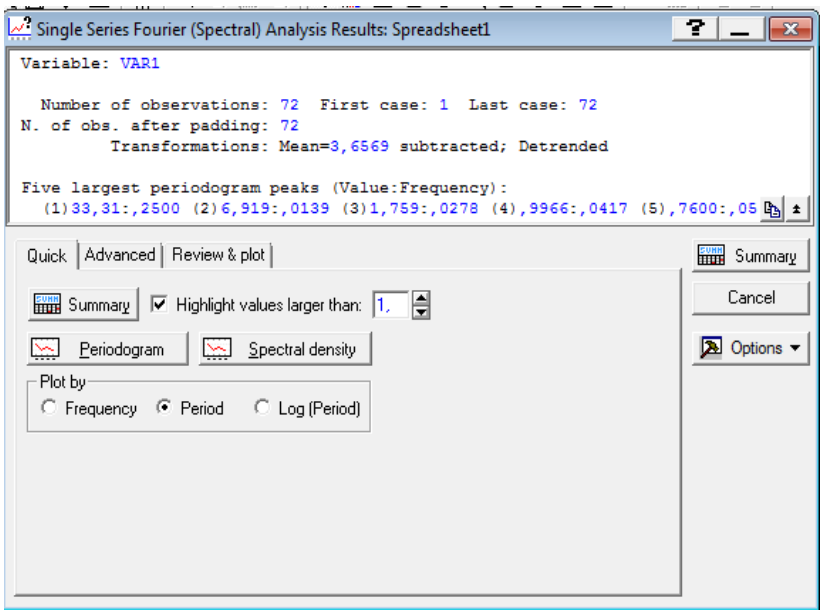


Рис. 2. Одномерный анализ Фурье.

Установим опцию *Plot by* (график) в положение *Period* (период) и нажмите кнопку *Periodogram* (периодограмма). Программа построит график, изображенный на рис. 3.

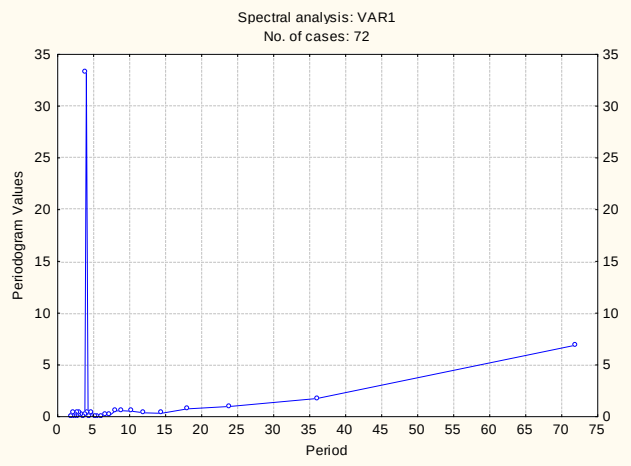


Рис. 3. Периодограмма.

График служит подтверждением того, что в ряду наблюдается тренд с переменным периодом. На графике имеется ярко выраженный пик, что дает основание предположить возможность существования тенденции к формированию устойчивого сезонного цикла с периодом в 4 месяца.

Два раза нажав на кнопку *Cancel*, вернемся к диалогу на вкладке *Advanced*. В рамке *Arima model parameters* (параметры АРГТСС) расположены опции, задающие число различных параметров модели:

• *p-Autoregressive* — параметр авторегрессии (регулярный);
• *P-Seasonal* — сезонный параметр авторегрессии;
• *q-Moving average* — параметр скользящего среднего (регулярный);
• *Q-Seasonal* — сезонный параметр скользящего среднего.

В поле возле каждого параметра задается число параметров данного типа.

Нажмем кнопку *Other transformation & plots* (другие преобразования и графики), откроется окно *Transformation of variables* (преобразование переменных). Воспользуемся преобразованием *Natural log* (взять

натуральный логарифм) на вкладке *x = f(x)*. Щелкнем кнопкой *OK* (*Transform selected series*). Программа построит график (рис. 4).

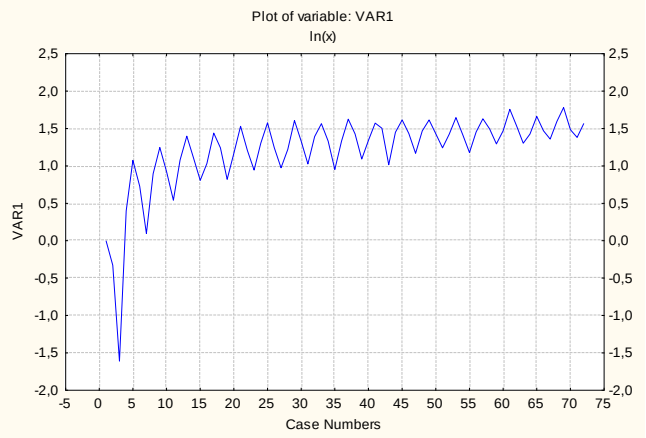


Рис. 4. График прологарифмированного ряда

После уменьшения разброса надо идентифицировать параметры модели для возможности дальнейшего оценивания параметров модели АРПСС. В модели АРПСС имеются следующие типы параметров: p — порядок авторегрессии, d — порядок разности, q — порядок скользящего среднего, сокращенно — модель АРПСС (p, d, q). Идентифицировать модель АРПСС — значит определить эти параметры. Основной критерий идентификации — поведение автокорреляционной и частной автокорреляционной функций ряда.

Пусть *d* — неизвестный порядок модели, который нужно оценить. Прежде всего, визуализируем ряд и определим, является ряд стационарным или нет.

Из графиков на рис. 1 и 4 видно, что ряд *Var1* имеет некоторую тенденцию к возрастанию значений при увеличении номера наблюдения, т.е. наблюдается монотонный тренд. Наличие тренда, который хорошо виден, — первое свидетельство о нестационарности анализируемого ряда.

Критерий нестационарности выражается в отсутствии тенденции к затуханию у выборочной автокорреляционной функции ряда, т.е. тренд может быть не виден отчетливо на графике, однако нестационарность проявится с помощью данного критерия. На вкладке *Autocorrs* при помощи опции *Number of lag* увеличим число лагов до 30 и нажмем кнопку *Autocorelations*. Появится график автокорреляционной функции (рис. 5). Из построенного графика видно, что автокорреляционная функция имеет слабую тенденцию затухания.

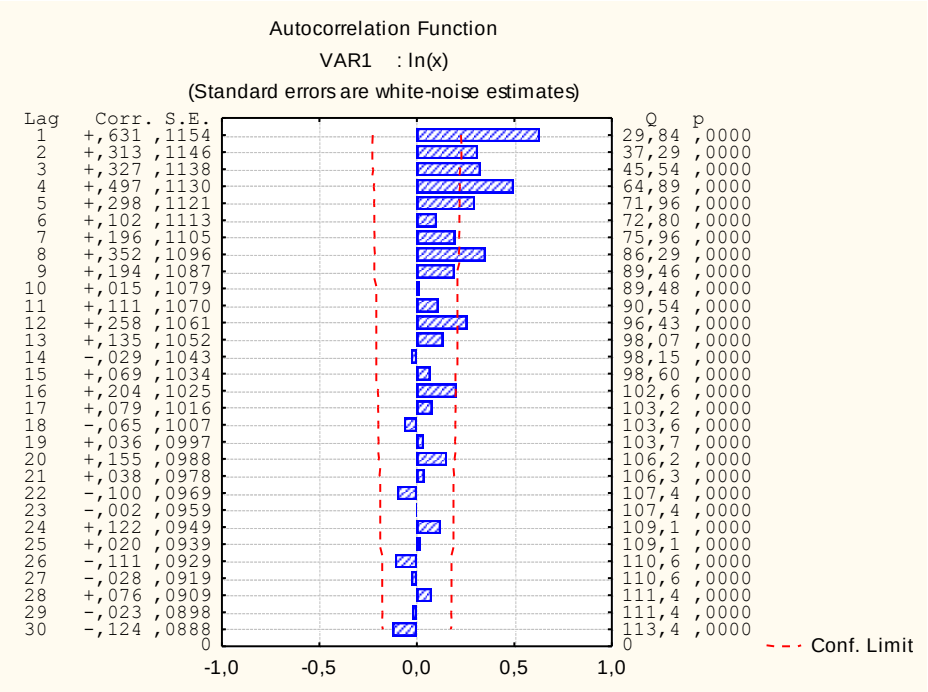


Рис. 5. График автокорреляционной функции

Итак, рассмотрев график автокорреляционной функции, приходим к выводу: ряд нестационарен. Теперь нужно рассмотреть разность первого порядка наблюдаемого ряда, предполагая, что ряд первых разностей будет стационарным. Определим для ряда *Var1 ln(x) d = k =1*. На вкладке *Differencing, Integrate* возьмем разность первого порядка, выделив опцию *Differencing (x=x- x(lag))* и указав значение *lag=1*. Щелкнем кнопкой *OK (Transform selected series)*. Из графика (рис. 6) видно, что ряд стал стационарным, так как нет тренда.

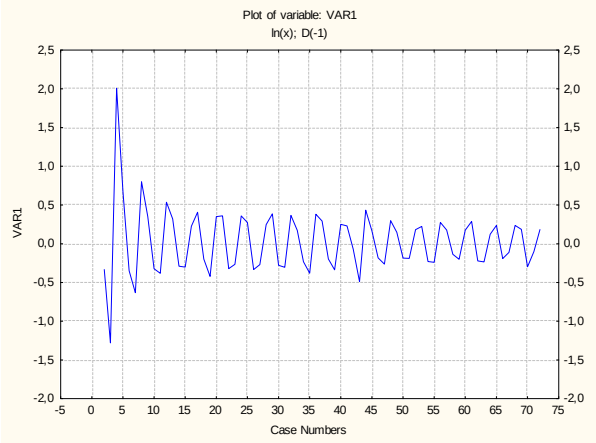


Рис. 6. График ряда после преобразования

Автокорреляционная функция (рис. 7) имеет незначительный выброс на лаге 1 и слабую тенденцию к затуханию, если не считать пиков устойчивого сезонного цикла с периодом в 4 месяца.

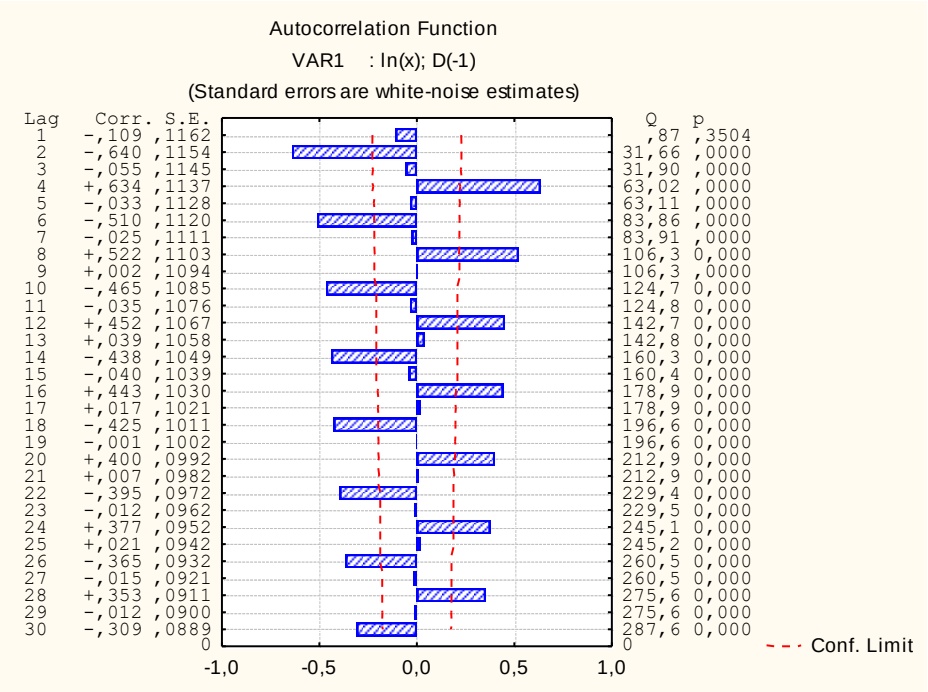


Рис. 7. График автокорреляционной функции

Таким образом, после двух преобразований построена стационарная модель.

Для определения параметров *p*, *q* рассматривают поведение выборочных автокорреляционной и частной автокорреляционной функций ряда.

Модель может быть отнесена с достаточной степенью точности к одному из следующих пяти классов:

• модели авторегрессии с одним параметром: $p = 1, q = 0$;

• модели авторегрессии с двумя параметрами: $p = 2, q = 0$;

• модели скользящего среднего с одним параметром: $p = 0, q = 1$;

• модели скользящего среднего с двумя параметрами: $p = 0, q = 2$;

• модели авторегрессии с одним параметром и скользящего среднего с одним параметром: $p = q = 1$.

Анализируя поведение автокорреляционной и частной автокорреляционной функций и учитывая приведенные критерии, можно сделать вывод, что наиболее подходящей моделью для ряда *Var1 ln(x) D(-1)* будет модель — один параметр скользящего среднего $p = 0, q = 1$.

• один параметр скользящего среднего: автокорреляционная функция имеет выброс на лаге 1 (нет корреляций для других задержек); частная автокорреляционная функция экспоненциально затухает — либо монотонно, либо осциллируя, т.е. меняя знак;

Учитывая, что d = 1, имеем несезонную модель АРПСС (0,1,1).

Так как ряд имеет ярко выраженную сезонную составляющую с периодом в 4 месяца, надо внести в модель сезонную корректировку. Полная мультипликативная сезонная модель может быть представлена в виде АРПСС (p, d, q) (Ps, Ds, Qs), где к параметрам модели АРПСС p, d, q добавлены сезонные параметры: сезонный параметр авторегрессии — Ps, сезонная разность — Ds, сезонный параметр скользящего среднего — Qs.

Для того чтобы учесть сезонные колебания с периодом в 4 месяца, необходимо взять сезонную разность с лагом 4 ряда *Var1 ln(x) D(-1)*.

Вернемся в окно *Transformations of variables*. На вкладке *Differencing, Integrate* возьмем разность четвертого порядка, выделив опцию *Differencing (x = x - x(lag))* и указав значение *lag =4*. Щелкните кнопкой *OK (Transform selected series)*. На рисунках 8 и 9 приведены графики построенного временного ряда и его автокорреляционной функции.

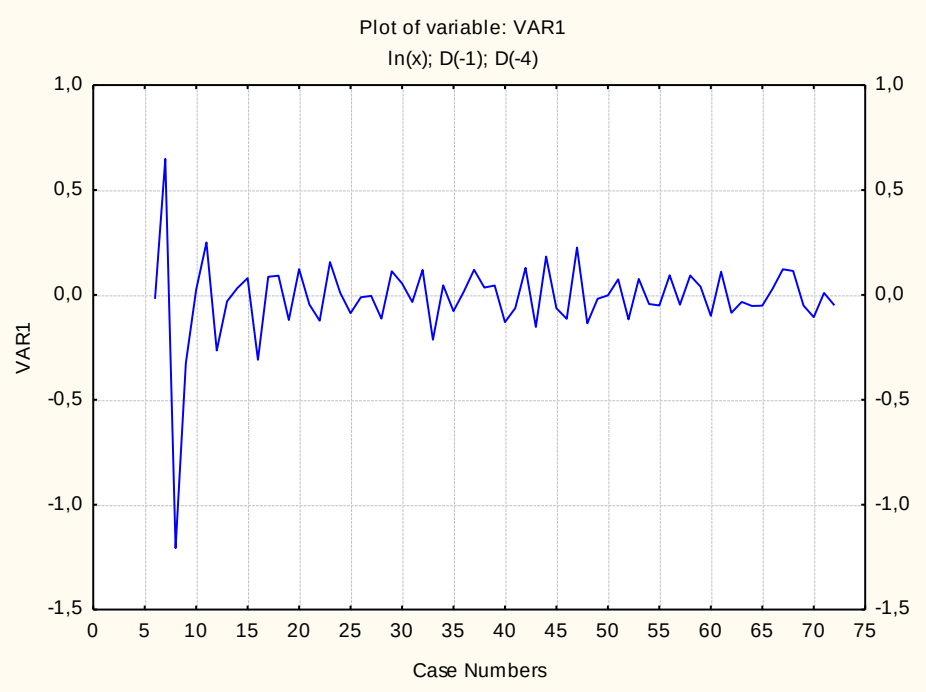


Рис. 8. График ряда после сезонной корректировки

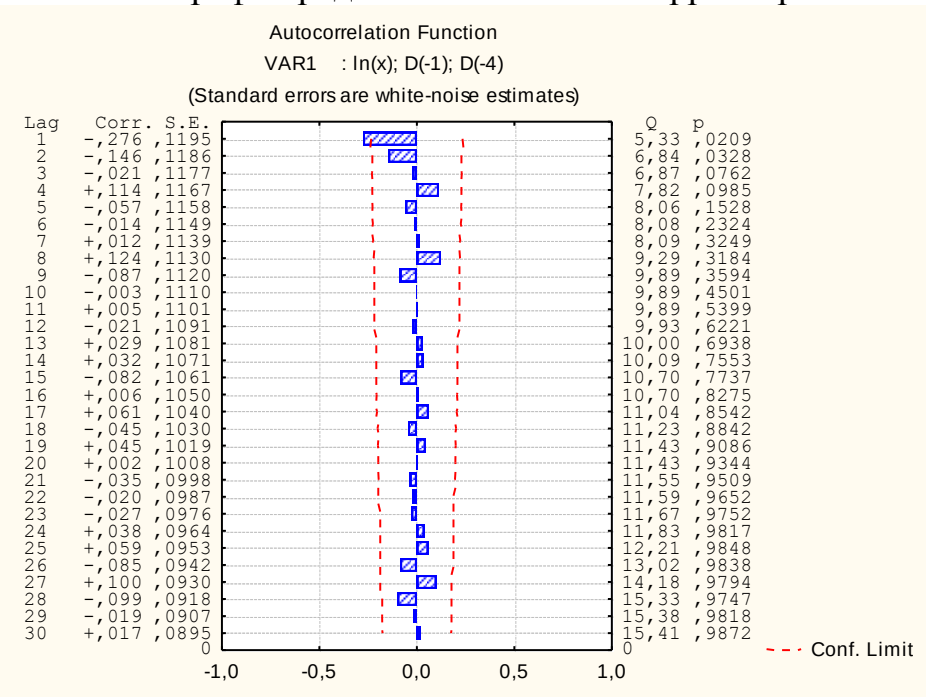


Рис. 9. Автокорреляционная функция

Из графиков видно, что ряд является стационарным, автокорреляционная функция экспоненциально затухает.

Значит, можно определить сезонный параметр авторегрессии Ps = 0, сезонную разность Ds = 1, сезонный параметр скользящего среднего Qs=0. Таким образом, полную модель определим как АРПСС (0,1,1), (0,1,0).

Преобразованный ряд можно теперь непосредственно использовать в АРПСС. Будем анализировать исходный ряд и зададим необходимые преобразования внутри АРПСС.

Теперь снова вернемся в диалоговое окно *Single Series ARIMA*, нажав *Cancel* в окне *Transformation of variables*. Нажмем *ARIMA & autocorrelation functions* (модель авторефессии и проинтегрированного скользящего среднего — сокращенно АРПСС). В диалоговом окне высвечена исходная переменная *Var1*. В рамке *Arima model parameters* установим значения соответствующих параметров. Нажмем *OK* (начать оценивание параметров) и запустим итеративную процедуру оценивания.

После того как процедура оценивания сойдется, откроется диалоговое окно *Single Series ARIMA Results* (результаты одномерной АРПСС).

Нажмем кнопку *Summary: Parameter estimates* (оценки параметров), чтобы увидеть таблицу (рис. 10) результатов с оценками, стандартными ошибками, асимптотическими значениями t-статистик и т.д.

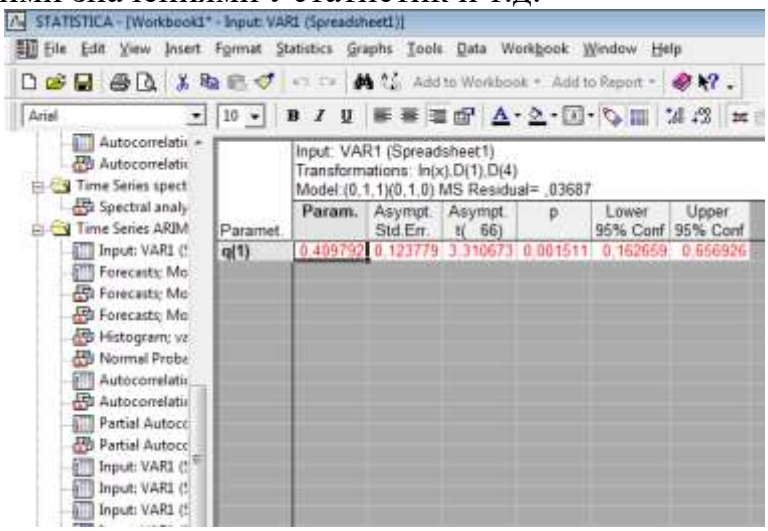


Рис. 10. Оценка параметров

Из таблицы (рис. 10) видно, что оценки параметра высоко значимы (p значительно меньше 0,05).

Сделаем прогноз. Нажмем кнопку *Forecast cases* (прогноз) на вкладке Advanced. Таблица результатов (рис. 11) содержит прогнозы и их доверительные интервалы для наблюдений, начиная с 73-го (т.е. строит прогноз на 4 наблюдения).

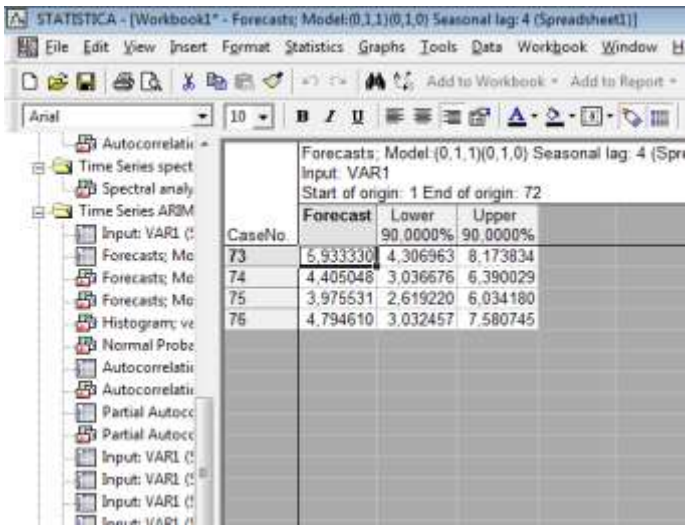


Рис. 11. Таблица результатов прогноза

Нажмем кнопку *Plot series & forecasts* (график ряда и прогнозов). Программа построит прогноз на 4 месяца (рис. 12).

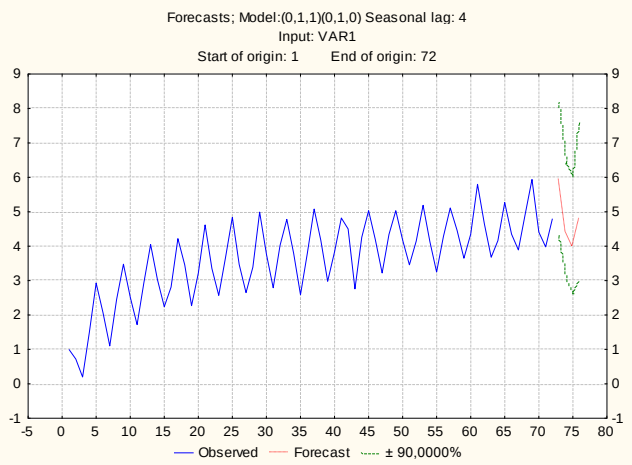


Рис. 12. График прогноза

Чтобы проверить, насколько хорошо построенная модель АРПСС прогнозирует последние 4 наблюдения, установим в поле *Start at case* (начать с наблюдения) значение 60 и снова нажмем кнопку *Plot series & forecasts*. Из графика (рис. 13) видно, что прогнозная кривая практически повторяет фрагмент кривой исходного ряда, причем все наблюдаемые значения ряда попадают в доверительный интервал.

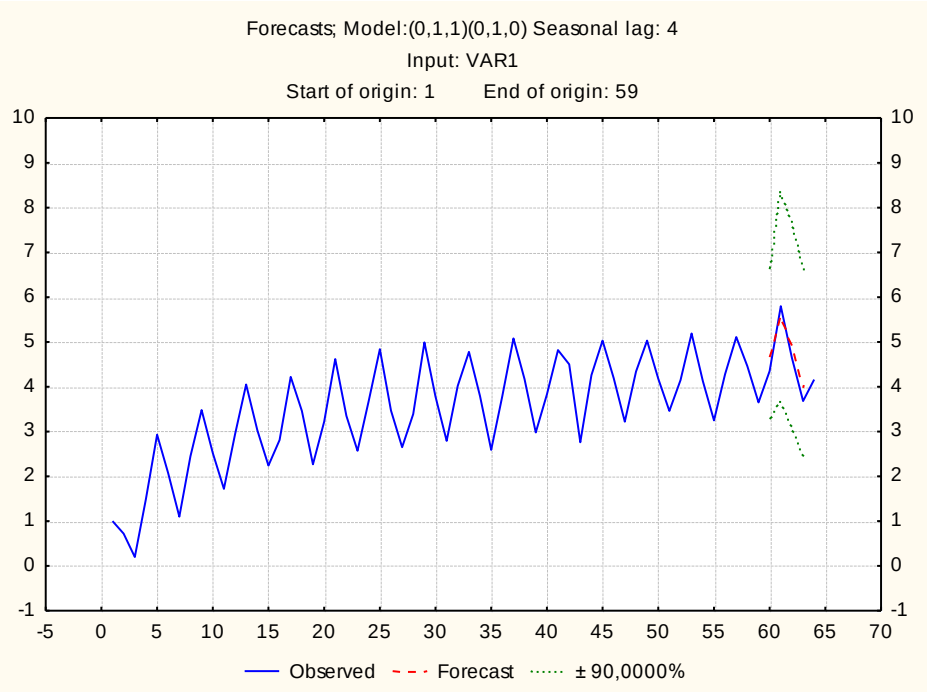


Рис. 13. График проверки качества прогноза

Для анализа адекватности модели исследуют остатки, представляющие собой разности наблюдаемых значений и значений, предсказанных с помощью модели.

На вкладке *Distribution of residuals* нажмем кнопку *Histogram*. Из графика (рис. 14) видно, что выборочная плотность распределения остатков успешно аппроксимируется нормальным законом распределения, что является признаком адекватности построенной модели прогноза.

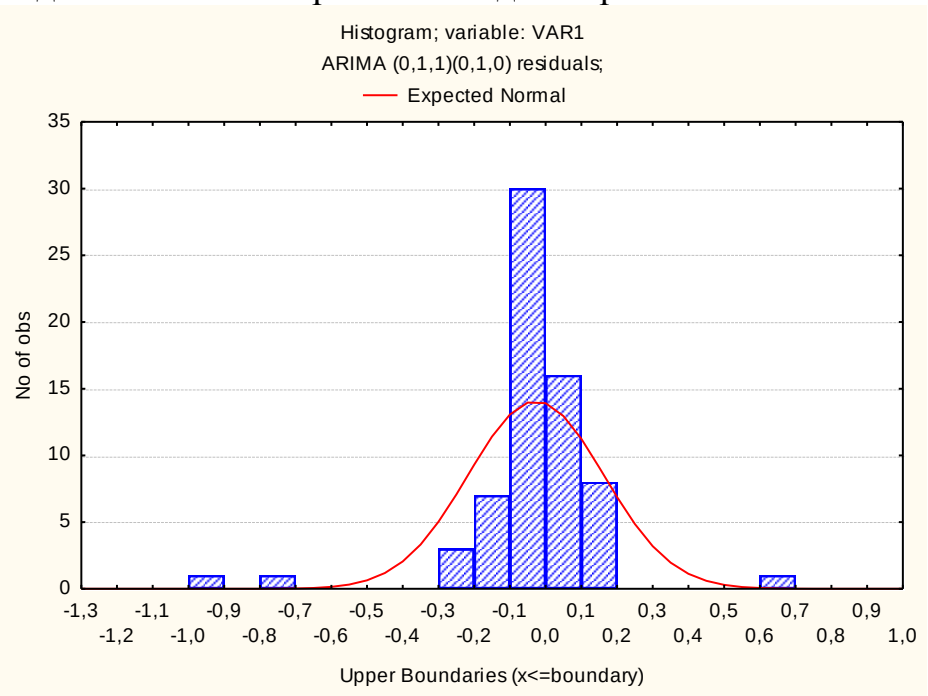


Рис. 14. Гистограмма

Откроем вкладку *Review & residuals*. В опции *Review & plot variables* (просмотреть переменные и построить график) нажмем кнопку *Plot(3)* (график остатков) напротив кнопки *Review residuals*. Программа построит

график остатков (рис. 15), из которого следует, что остатки имеют примерно равную вариацию на всем протяжении ряда, и нет очевидного тренда или сдвига в них.

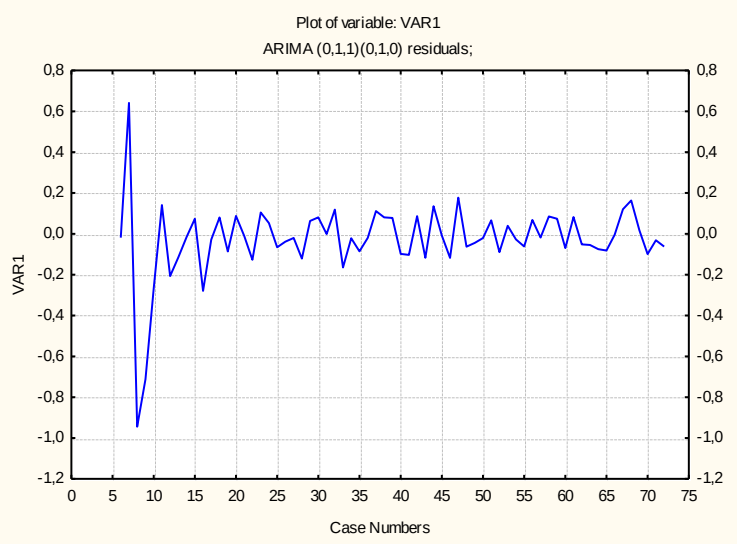
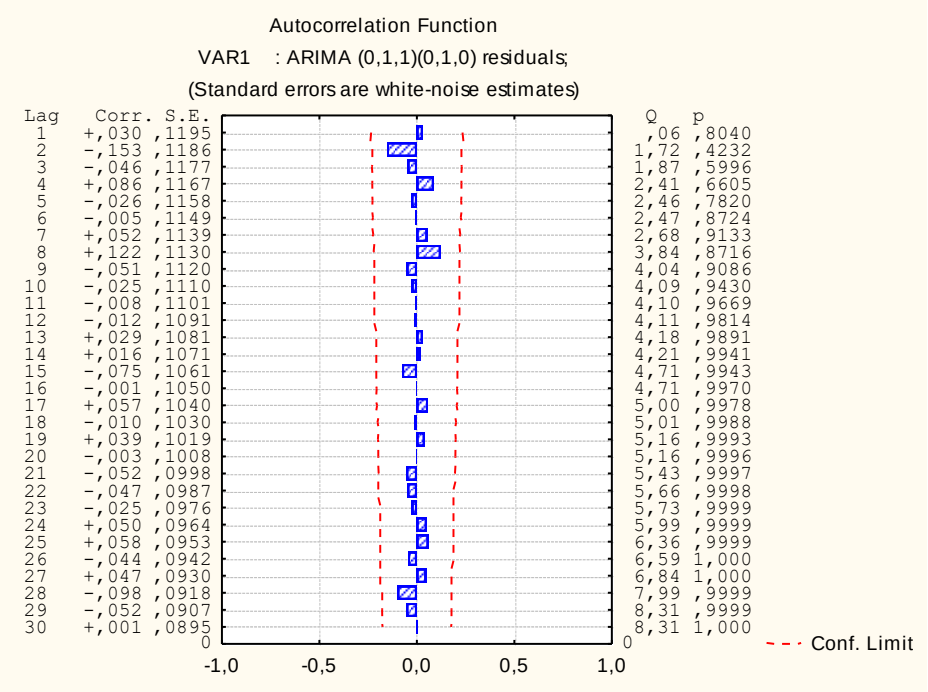


Рис. 15. График остатков

На вкладке *Autocorrelations* нажмем кнопки *Autocorrelations* и *Partial Autocorrelations*. Из графиков (рис. 16-17) видно, что остатки практически являются белым шумом.



Lag	Corr.	S.E.	Q	p
1	+,030	,1195	,06	,8040
2	-,153	,1186	1,72	,4232
3	-,046	,1177	1,87	,5996
4	+,086	,1167	2,41	,6605
5	-,026	,1158	2,46	,7820
6	-,005	,1149	2,47	,8724
7	+,052	,1139	2,68	,9133
8	+,122	,1130	3,84	,8716
9	-,051	,1120	4,04	,9086
10	-,025	,1110	4,09	,9430
11	-,008	,1101	4,10	,9669
12	-,012	,1091	4,11	,9814
13	+,029	,1081	4,18	,9891
14	+,016	,1071	4,21	,9941
15	-,075	,1061	4,71	,9943
16	-,001	,1050	4,71	,9970
17	+,057	,1040	5,00	,9978
18	-,010	,1030	5,01	,9988
19	+,039	,1019	5,16	,9993
20	-,003	,1008	5,16	,9996
21	-,052	,0998	5,43	,9997
22	-,047	,0987	5,66	,9998
23	-,025	,0976	5,73	,9999
24	+,050	,0964	5,99	,9999
25	+,058	,0953	6,36	,9999
26	-,044	,0942	6,59	1,000
27	+,047	,0930	6,84	1,000
28	-,098	,0918	7,99	,9999
29	-,052	,0907	8,31	,9999
30	+,001	,0895	8,31	1,000

Рис. 16. Автокорреляционная функция

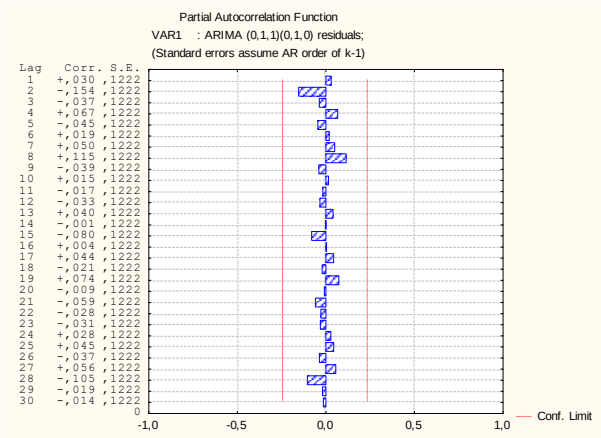


Рис. 17. Частная автокорреляционная функция

Таким образом, всесторонний анализ остатков показал, что при помощи процедуры *ARIMA & autocorrelation functions* построена вполне адекватная модель прогноза.

Лабораторная работа №4
по дисциплине Компьютерные и информационные технологии
Пакет MatLab. Работа с матрицами

Цель Изучить работу с матрицами в пакете MatLab.

Задание.

1. Ввести две матрицы из текстовых файлов
2. Провести над ними операции: сложение, перемножение, транспонирование, нахождение собственных значений и векторов
3. Решить систему линейных алгебраических уравнений (Ax = b), используя одну из матриц и случайный вектор b. (использовать обратную матрицу).

Выполнение

1. Записываем 2 матрицы и сохраням в формате .dat. Для вывода записанных матриц в Matlab записываем «load «путь»»название».dat»:

```
>> load R:\A.dat
>> load R:\B.dat
>> A

A =

     2     3    10    99
     3     5     7     1
     1     1     1    11
   100   101   102    59

>> B

B =

    1.1000    1.2000    5.6000   10.0000
   10.0000    5.6000    1.2000  101.9000
   52.0000   53.7000   11.7000   19.0000
   18.0000   17.0000   20.5000   34.8000
```

2. Первые операции элементарны:

```
>> A+B

ans =

    3.1000    4.2000   15.6000  109.0000
   13.0000   10.6000    8.2000  102.9000
   53.0000   54.7000   12.7000   30.0000
  118.0000  118.0000  122.5000   93.8000

>> A*B

ans =

   1.0e+04 *

    0.2334    0.2239    0.2161    0.3961
    0.0435    0.0425    0.0125    0.0707
    0.0261    0.0248    0.0244    0.0514
    0.7486    0.7166    0.3084    1.5283

>> A'

ans =

     2     3     1   100
     3     5     1   101
    10     7     1   102
    99     1    11    59

>> B'

ans =

    1.1000   10.0000   52.0000   18.0000
    1.2000    5.6000   53.7000   17.0000
    5.6000    1.2000   11.7000   20.5000
   10.0000  101.9000   19.0000   34.8000
```

Для нахождения собственных значений и векторов существует функция “eig”.

```
[R,D]=eig(A)

R =

   -0.5801   -0.7809   -0.2760    0.6630
   -0.0221    0.0283    0.7950   -0.7442
   -0.0677   -0.0791   -0.5388    0.0805
   -0.8114    0.6190    0.0380    0.0114


D =

  141.7561         0         0         0
         0  -75.5728         0         0
         0         0   -0.7383         0
         0         0         0    1.5551

>> [R,D]=eig(B)

R =

   0.1038 + 0.0000i  -0.7249 + 0.0000i   0.0854 + 0.0599i   0.0854 - 0.0599i
   0.6123 + 0.0000i   0.6881 + 0.0000i   0.2994 - 0.5189i   0.2994 + 0.5189i
   0.6235 + 0.0000i   0.0153 + 0.0000i  -0.7672 + 0.0000i  -0.7672 + 0.0000i
   0.4749 + 0.0000i   0.0294 + 0.0000i   0.0749 + 0.1897i   0.0749 - 0.1897i


D =

  87.5624 + 0.0000i   0.0000 + 0.0000i   0.0000 + 0.0000i   0.0000 + 0.0000i
   0.0000 + 0.0000i  -0.5619 + 0.0000i   0.0000 + 0.0000i   0.0000 + 0.0000i
   0.0000 + 0.0000i   0.0000 + 0.0000i -16.9003 +27.5622i   0.0000 + 0.0000i
   0.0000 + 0.0000i   0.0000 + 0.0000i   0.0000 + 0.0000i -16.9003 -27.5622i
```

В нашем случае R — это собственные векторы соответствующих собственных значений D.

3. Случайный вектор-столбец от 0 до 100 зададим с помощью встроенной функции rand:

```
b=(100).*rand(4,1)

b =

   67.8735
   75.7740
```

```
   74.3132
   39.2227
```

По правилу, для нахождение x домножаем вектор b слева на обратную матрицу A:

```
x=inv(A)*b

x =

 -156.7497
  264.3881
 -111.8653
    7.1400
```

Можно провести проверку:

```
A*x

ans =

   67.8735
   75.7740
   74.3132
   39.2227
```

Всё верно

Лабораторная работа №5
по дисциплине Компьютерные и информационные технологии
Пакет MatLab. Функции и графики.

Цель: Освоить правила построения функций и графиков в MATLAB.

Задание.

1 Выполнить арифметические и логические операции, в том числе с комплексными числами.

2 Построить графики:

2.1 функции в общем виде (использовать разные стили и масштабы, подписи, легенды, сетки)

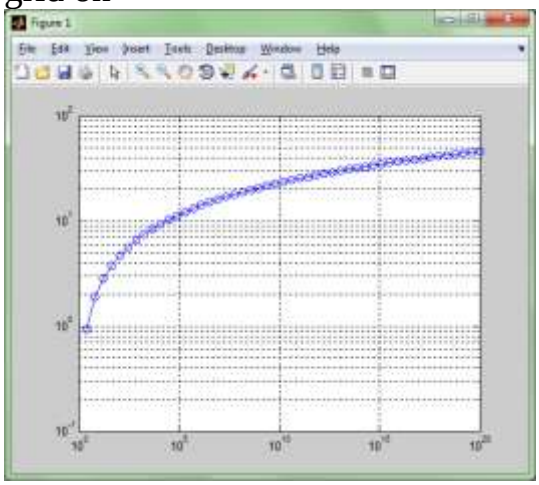
2.2 графики функций в полярных координатах,

2.3 гистограммы

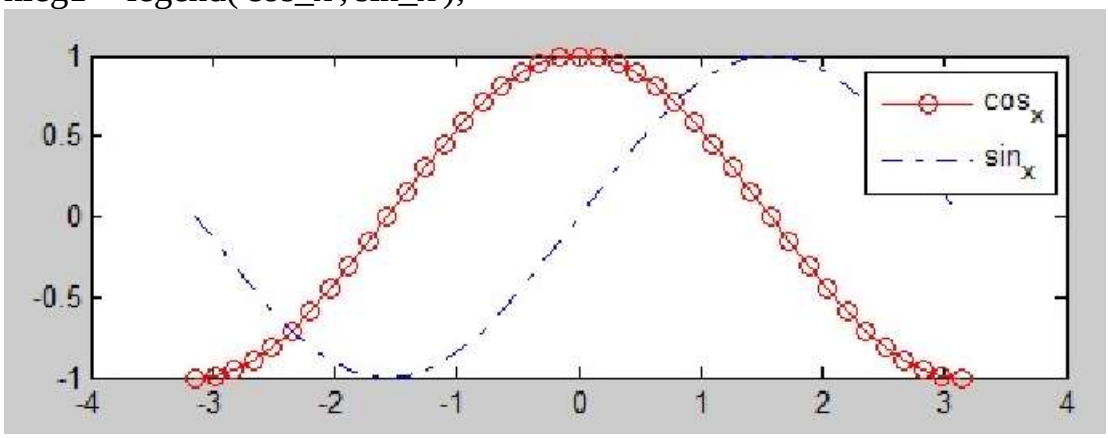
3 Изучить средства управления в графическом окне, редактирование графиков.

Выполнение

```
x = logspace(0,20);
loglog(x,log(x),'-o')
grid on
```

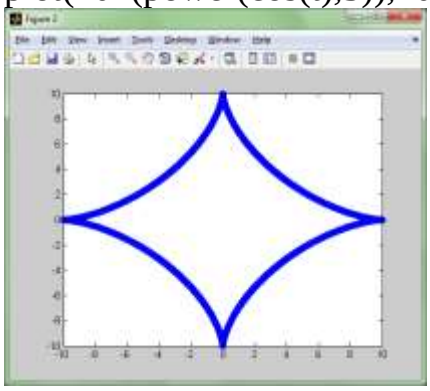


```
figure
x = -pi:pi/20:pi;
plot(x,cos(x),'-ro',x,sin(x),'-.b')
hleg1 = legend('cos_x','sin_x');
```



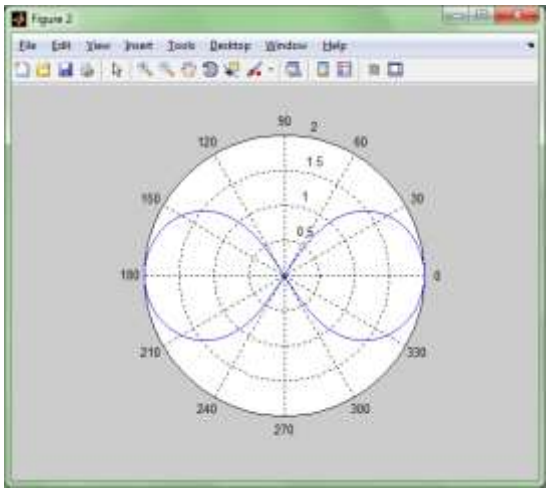
Астроида

```
figure
t = 0:.000001:2*pi;
plot(10*(power(cos(t),3)),10*(power(sin(t),3)),'-o')
```

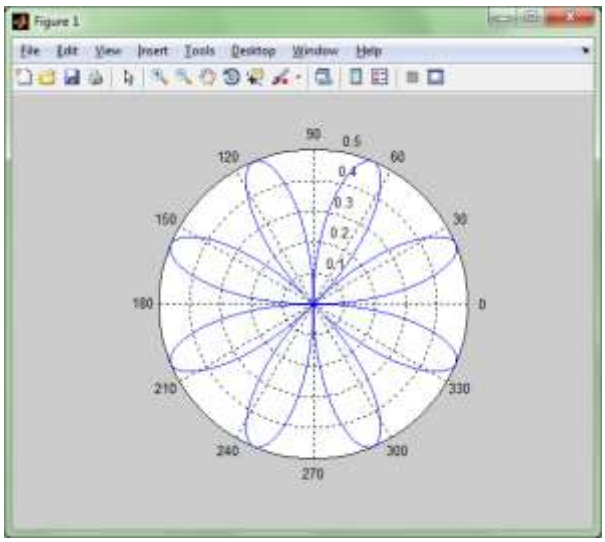


Лемниската Бернулли

```
figure
t = -pi:.001:pi;
polar(sin(t),2*cos(t))
```

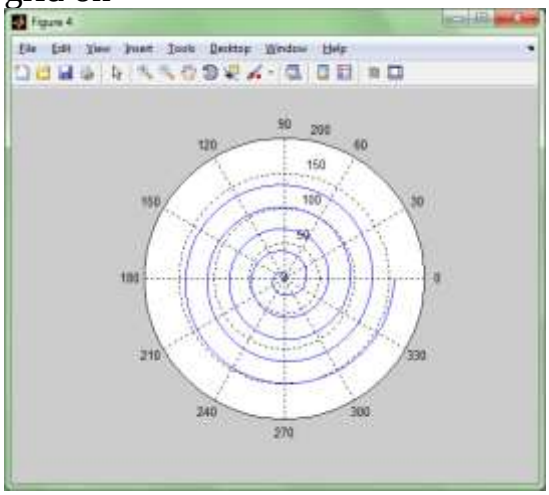



```
figure
t = 0:.001:2*pi;
polar(t,sin(2*t).*cos(2*t))
axis equal
grid on
```

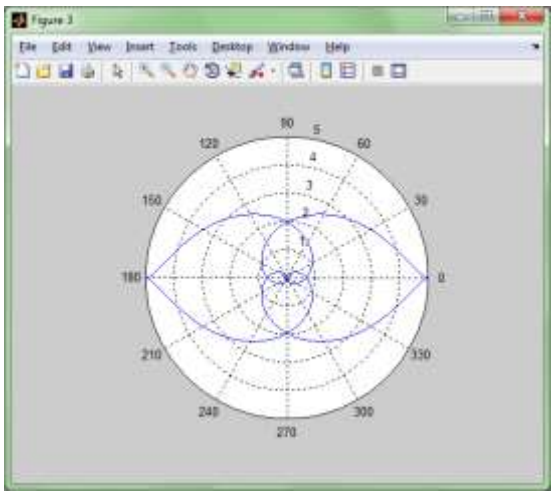


Архимедова спираль

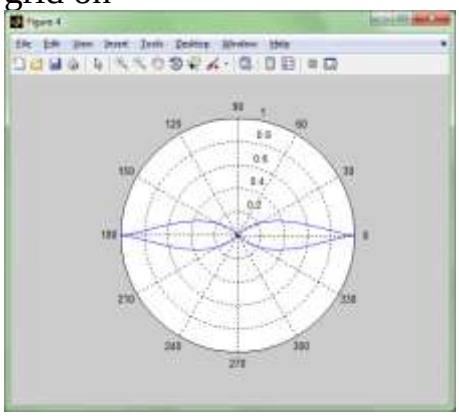
```
figure
t = 0:.001:10*pi;
polar(t,5*t)
axis equal
grid on
```



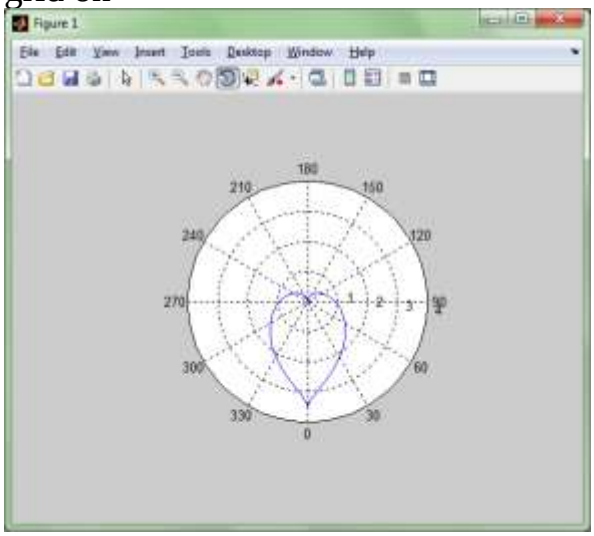

```
figure
t = 0:.001:2*pi;
polar(5*power(cos(t),3),5*power(sin(t),3))
axis equal
grid on
```




```
figure
t = 0:.001:2*pi;
polar(power(cos(t),3),power(sin(t),3))
axis equal
grid on
```



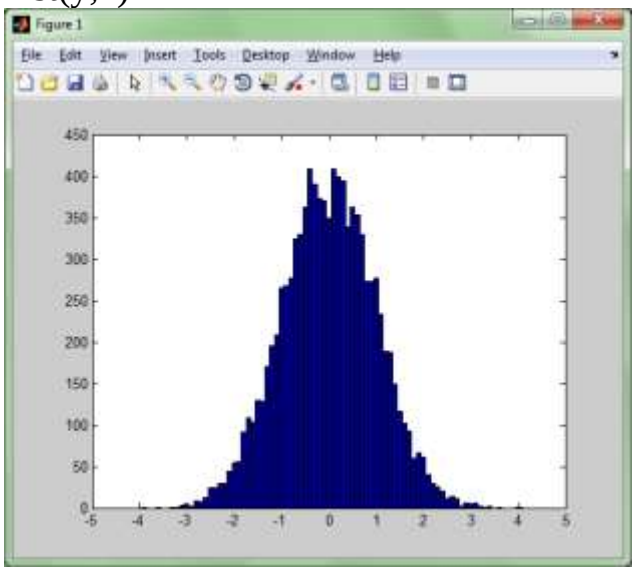

```
figure
t = 0:.001:pi;
polar(3.5*power(cos(t),3),3.5*power(sin(t),3))
axis equal
grid on
```



Нормальное распределение

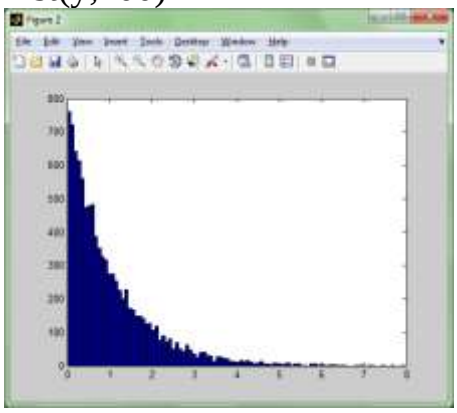
```
figure
```

```
x = -4:0.1:4;
y = randn(10000,1);
hist(y,x)
```

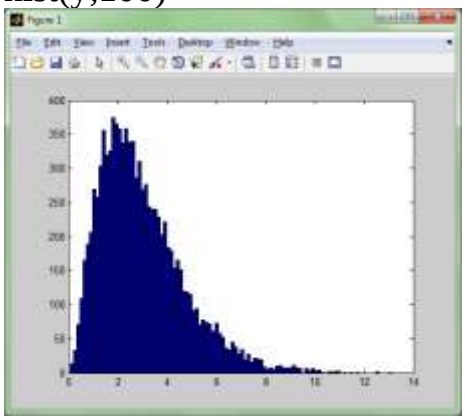


Гамма распределения

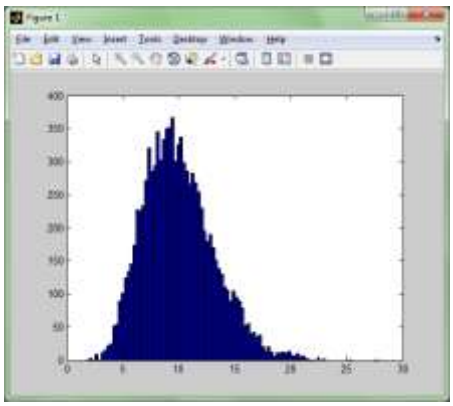
```
figure
y = randg(1,10000,1);
hist(y,100)
```



```
figure
y = randg(3,10000,1);
hist(y,100)
```



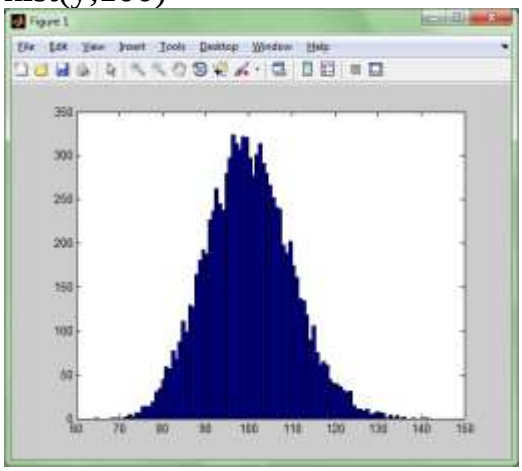
```
figure
y = randg(10,10000,1);
hist(y,100)
```



figure

y = randg(100,10000,1);

hist(y,100)



Лабораторная работа №6
по дисциплине Компьютерные и информационные технологии
Пакет MatLab. Решение систем дифференциальных уравнений.

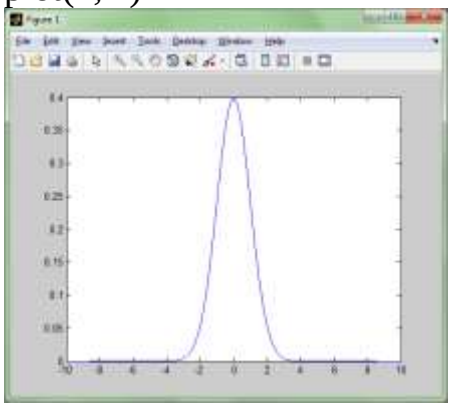
Цель: Освоить правила построения функций и графиков в MATLAB.
Задание.
Задание.

1. Генерация случайных чисел
2. Решение дифференциального уравнения с начальными условиями (первого и второго порядков)
3. Решение уравнений в частных производных

Выполнение

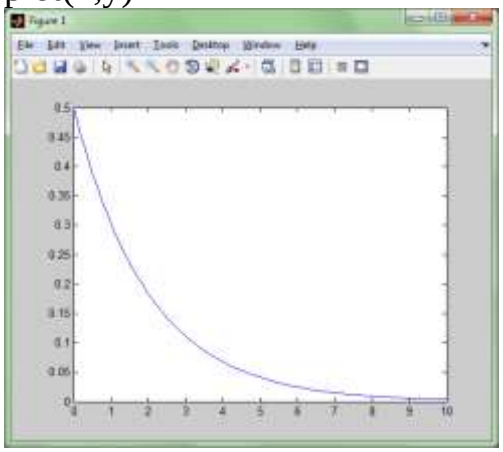
1. Нормальное распределение:

```
randn(1,10)
ans =
   0.0294    0.0815   -1.0833   -1.2780   -0.7340   -0.3688   -0.0233   -1.3001   -0.7068   -0.4749
x=-10:.01:10;
Y = pdf('normal',x,0,1);
plot(x,Y)
```



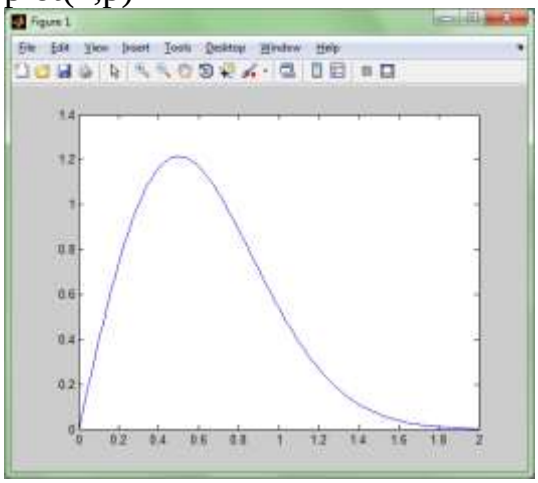
Экспоненциальное распределение:

```
x = 0:0.1:10;
y = exppdf(x,2);
plot(x,y)
```



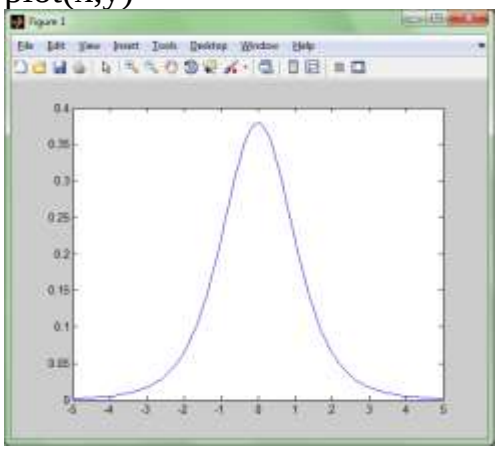
Распределение Рэлея

```
x = [0:0.01:2];
p = raylpdf(x,0.5);
plot(x,p)
```



Распределение Стьюдента:

```
x = -5:0.1:5;
y = tpdf(x,5);
plot(x,y)
```



2. Найдём решение дифференциальных уровнений:

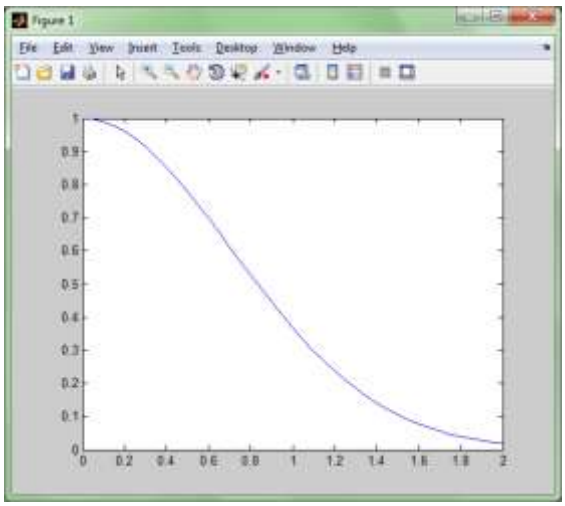
1) $\frac{dy}{dx} = -2xy$ на [0;2] при $y(0) = 1$.

Аналитическое решение:

```
dsolve('Dy=-2*x*y','y(0)=1')
ans =
exp(-2*t*x)
```

График решения построим с помощью функции

```
function [ dd ] = func( x, y )
dd=-2*x*y;
end
[T,Y]=ode45(@func,[0,2],1);
plot(T,Y)
```



2) $y''' - 2y'' - y' + 2y = 0.$

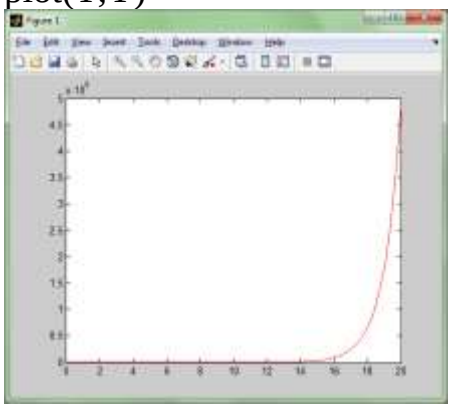
```
dsolve('D3y-2*D2y-Dy+2*y=0')

ans =

C114*exp(t) + C113*exp(2*t) + C115*exp(-t)

function [ dd ] = fu( x, y )
dd=zeros(3,1);
dd(1)=y(2);
dd(2)=y(3);
dd(3)=2*y(3)+y(2)-2*y(1);
end

[T,Y]=ode45('fu',[0 20],[1 1 1 ]);
plot(T,Y)
```



3. Решать будем следующее уравнение:

$\pi^2 \dfrac{\partial u}{\partial t} = \dfrac{\partial}{\partial x}\left(\dfrac{\partial u}{\partial x}\right)$ при $0 \le x \ge 1$ и $t \ge 0$.

Начальное условие:

$u(x,0) = \sin \pi x$

Краевые условия:

$u(0,t) \equiv 0;$

$\pi e^{-t} + \frac{\partial u}{\partial x}(1,t) = 0.$

```
function [c,f,s] = pdex1pde(x,t,u,DuDx)
c = pi^2;
f = DuDx;
s = 0;
end

function u0 = pdex1ic(x)
u0 = sin(pi*x);
end

function [pl,ql,pr,qr] = pdex1bc(xl,ul,xr,ur,t)
pl = ul;
ql = 0;
pr = pi * exp(-t);
qr = 1;
end

m = 0;
x = linspace(0,1,20);
t = linspace(0,2,5);

sol = pdepe(m,@pdex1pde,@pdex1ic,@pdex1bc,x,t);
% извлекаем первое решение
u = sol(:,:,1);

% поверхность решений
surf(x,t,u)
title('Numerical solution computed with 20 mesh points.')
xlabel('Distance x')
ylabel('Time t')
% контур решения
figure
plot(x,u(end,:))
title('Solution at t = 2')
xlabel('Distance x')
ylabel('u(x,2)')
```

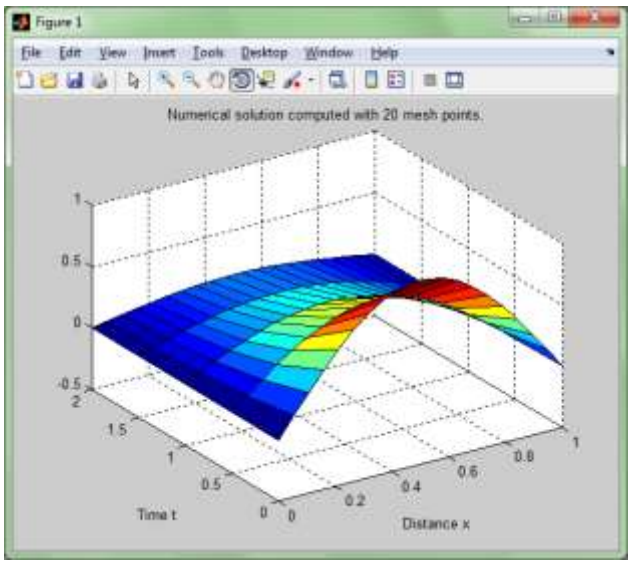


Figure 1
File Edit View Insert Tools Desktop Window Help
Numerical solution computed with 20 mesh points.
1
0.5
0
-0.5
2
1.5
1
0.5
0
Time t
0
0.2
0.4
0.6
0.8
1
Distance x

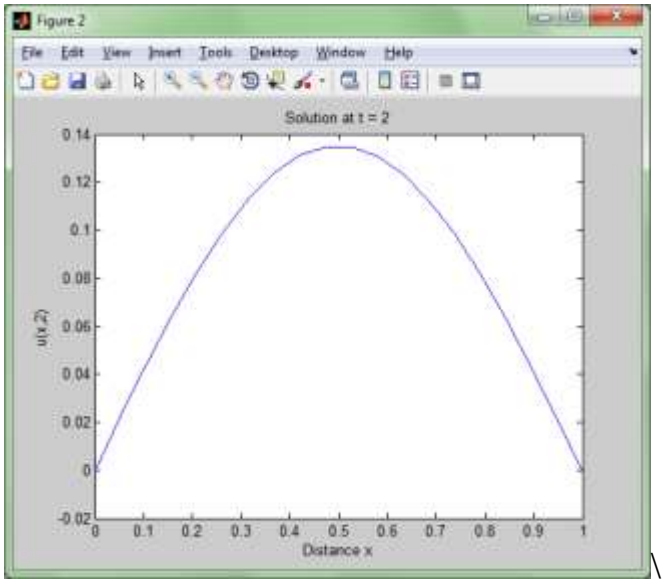


Figure 2
File Edit View Insert Tools Desktop Window Help
Solution at t = 2
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
-0.02
u(x,2)
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Distance x