

Министерство образования Республики Беларусь
УО «Полесский государственный университет»

Л.Н. БАЗАКА, А.И. РАЗИНКОВ

**Статистическая обработка данных в среде пакетов
Statistica, EViews и MS Excel**

Методические указания по выполнению лабораторных работ
для студентов экономических специальностей и слушателей факультета
повышения квалификации и переподготовки кадров

Пинск
ПолесГУ
2015

УДК 519.86(075.8)
ББК 65в6я73
Б17

Р е ц е н з е н т ы:

кандидат физико-математических наук, доцент П.А. Павлов,
кандидат физико-математических наук, доцент М.А. Романова

У т в е р ж д е н о
научно-методическим советом ПолесГУ

Базака, Л.Н.

Б17 Статистическая обработка данных в среде пакетов Statistica, EViews и MS Excel: методические указания по выполнению лабораторных работ / Л.Н. Базака, А.И. Разинков. – Пинск: ПолесГУ, 2015. – 138 с.

ISBN 978-985-516-375-7

Методические указания осуществляют методическую поддержку лабораторных занятий по эконометрике в рамках курсов «Корпоративные информационные системы», «Эконометрика» (продвинутый уровень). Не требуют основательной математической подготовки. Содержат краткие указания, примеры решения типовых задач, описание реализации на компьютере с помощью пакетов прикладных программ Excel, Statistica, EViews, а также задания для самостоятельного выполнения.

Предназначены для студентов экономических специальностей и слушателей факультета повышения квалификации и переподготовки кадров.

УДК 519.86(075.8)
ББК 65в6я73

ISBN 978-985-516-375-7

©УО «Полесский государственный университет», 2015

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	6
1 Основные понятия.....	7
1.1 Методы описательной статистики	7
1.2 Диаграмма разброса (рассеяния).....	10
1.3 Модели временного ряда.....	14
1.4 Коэффициенты регрессии, корреляции и детерминации.....	14
2 Лабораторная работа №1	16
2.1 Задания.....	16
2.2 Методические указания.....	18
2.2.1 Понятие выборки, временного ряда	18
2.2.2 Настройка Пакета анализа данных	18
2.2.3 Инструмент Гистограмма Пакета анализа данных	18
2.2.4 Инструмент Генерация случайных чисел	19
2.2.5 Нормальный закон распределения	20
2.2.6 Правило трех сигм.....	22
2.2.7 Анализ однородности выборки.....	23
2.2.8 Технология вычислений в MS Excel	23
2.2.9 Технология вычислений в EViews.....	33
2.2.10 Технология вычислений в пакете Statistica	36
2.2.11 Анализ результатов вычисления.....	43
2.3 Задания для самостоятельной работы.....	48
3 Лабораторная работа №2.....	49
3.1 Задания.....	49
3.2 Методические указания.....	50
3.2.1 Парная регрессия.....	50
3.2.2 Множественная регрессия.....	51
3.2.3 Корреляционный анализ.....	51
3.2.4 Задачи регрессионного анализа	53
3.2.5 Оценка параметров уравнения регрессии.....	53
3.2.6 Технологии вычислений в MS Excel	54

3.2.7	Технологии вычислений в <i>EViews</i>	63
3.2.8	Технологии вычислений в <i>Statistica</i>	68
4.	Лабораторная работа №3.....	86
4.1	Задания.....	86
4.2	Методические указания.....	86
4.2.2	F-статистика. Критерий Фишера.....	87
4.2.3	Оценка адекватности модели.....	88
4.2.4	Нормальность распределения остатков.....	89
4.2.5	Гомоскедастичность (гетероскедастичность) остатков.....	89
4.2.6	Результаты регрессионной статистики и их оценка.....	90
4.2.7	Технологии вычислений в MS Excel.....	90
4.2.8	Технологии вычислений в <i>EViews</i>	100
4.2.9	Анализ результатов вычисления.....	106
4.3	Задания для самостоятельной работы.....	109
5.	Лабораторная работа №4.....	109
5.1	Задания.....	109
5.2	Методические указания.....	110
5.2.1	Показатели оценок точности прогноза.....	110
5.2.2	Доверительный интервал прогноза.....	111
5.2.3	Интерпретация полученного уравнения регрессии.....	113
5.2.4	Технологии вычислений в MS Excel.....	113
5.2.5	Технологии вычислений в <i>EViews</i>	114
5.2.6	Анализ результатов вычисления.....	118
5.3	Задания для самостоятельной работы.....	120
	ПРИЛОЖЕНИЕ А.....	121
	ПРИЛОЖЕНИЕ Б.....	122
	ПРИЛОЖЕНИЕ В.....	123
	ПРИЛОЖЕНИЕ Г.....	124
	ПРИЛОЖЕНИЕ Д.....	125
	ПРИЛОЖЕНИЕ Е.....	126
	ПРИЛОЖЕНИЕ Ж.....	127
	ПРИЛОЖЕНИЕ З.....	128
	ПРИЛОЖЕНИЕ И.....	129

ПРИЛОЖЕНИЕ Л	131
ПРИЛОЖЕНИЕ М	132
ПРИЛОЖЕНИЕ Н	133
ПРИЛОЖЕНИЕ О	134
ПРИЛОЖЕНИЕ П	135
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	136

ВВЕДЕНИЕ

Постоянно усложняющиеся экономические процессы требуют повышения уровня образования современных специалистов в области экономики и управления. Для исследования конкретных количественных и качественных взаимосвязей экономических объектов и процессов можно использовать математические и статистические методы и моделирование процессов. Методы статистического анализа, которые используются для исследования конкретных экономических данных, называются эконометрические методы.

Оценка результатов эконометрического моделирования достигается посредством решения качественной и количественной проблемы. Качественная составляющая заключается в установлении соответствия между построенной моделью и основополагающей экономической концепцией, а количественная – в точности аппроксимации имеющейся информации данными расчётов.

К основным задачам эконометрики можно отнести: построение эконометрических моделей – представление экономических моделей в математической форме, удобной для проведения эмпирического анализа; оценка параметров построенной модели; оценка адекватности модели реальными данными; проверка качества полученной модели; использование построенной модели для прогнозирования.

Модель парной линейной регрессии является частным случаем модели многомерной регрессии. Её исследование представляет самостоятельный интерес, так как она имеет многие характерные свойства общих многомерных моделей, но более наглядна и проста для изучения.

1 Основные понятия

1.1 Методы описательной статистики

Методы предназначены для первичного анализа большой выборки значений одного признака.

Пусть из генеральной совокупности X извлечена выборка

$$\{(x_i, m_i)\}, i = \overline{1, n} \quad (1)$$

где n – объем выборки, n_i – число появлений значения x_i .

Наблюдаемые значения называют *вариантами*. Число m_i появлений значения x_i называют *частотой*, а частное m_i/n от деления частоты на объем выборки – *относительной частотой*. Последовательность вариант и соответствующих им частот, упорядоченная в возрастающем порядке, называется *дискретным вариационным рядом*.

Если объем выборки значителен, то дискретный вариационный ряд теряет наглядность. В этом случае выполняют группировку данных – построение *непрерывного вариационного ряда*.

При выполнении группировки весь диапазон изменения $[x_{min}; x_{max}]$ величины x делится на несколько интервалов – *разрядов*, число которых выбирают по *правилу Стерджесса (Herbert Sturges, 1926)*:

$$k = 1 + 3,322 \lg n, k = 1 + \log_2 n \quad (2)$$

Результат вычисления округляется до целого значения в большую сторону.

Частоты, соответствующие каждому разряду, находятся как суммы частот всех вариант, попавших в этот разряд (если в исходной выборке каждая варианта встречается только один раз, то частота находится как количество вариант, попавших в интервал).

Для графического представления непрерывного вариационного ряда выполняют построение *гистограммы* – ступенчатой фигуры, состоящей из прямоугольников, основания которых построены на соответствующих разрядах, а высоты h_j равны частному от деления относительной частоты на длину разряда:

$$h_j = \frac{m_j}{n(x_{j+1} - x_j)} = \frac{m_j}{n\Delta x}, j = \overline{1, k} \quad (3)$$

Гистограмма позволяет сделать предварительное суждение о плотности распределения генеральной совокупности.

Статистическими (выборочными, эмпирическими, опытными) оценками называют функции от наблюдаемых значений. *Точечными оценками* называют оценки, выражаемые одним числом.

При анализе данных (набора данных, называемых выборками) происходит замена теоретической функции $F(x)$ (генеральной совокупности) на ее выборочный аналог $F_n(x)$.

Это приводит к тому, что выборочные характеристики являются оценками соответствующих характеристик генеральной совокупности. Эти оценки должны удовлетворять определенным требованиям. В соответствии с важнейшими требованиями, оценки должны быть:

- несмещенными, то есть стремиться к истинному значению характеристики генеральной совокупности при неограниченном увеличении количества испытаний;
- состоятельными, то есть с ростом размера выборки оценка должна стремиться к значению соответствующего параметра генеральной совокупности с вероятностью, приближающейся к 1;
- эффективными, то есть для выборок равного объема используемая оценка должна иметь минимальную дисперсию.

Среди выборочных характеристик выделяют показатели, относящиеся к центру распределения (меры положения), показатели рассеяния вариант (меры рассеяния) и меры формы распределения.

К показателям, характеризующим центр распределения, относят различные виды средних (арифметическое, геометрическое и т.п.), а также моду, медиану и математическое ожидание.

Мода (Mo) — это элемент выборки с наиболее часто встречающимся значением (наиболее вероятная величина).

Оценку моды обычно находят графически. Для этого на гистограмме находят прямоугольник с наибольшей высотой и проводят из противоположных вершин его верхнего основания два отрезка к противоположным вершинам верхних оснований соседних прямоугольников. В качестве оценки моды принимается абсцисса точки пересечения этих отрезков.

Средним значением (\bar{x}) выборки, или выборочным аналогом математического ожидания, называется величина

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (4)$$

Иначе говоря, среднее значение - это центр выборки, вокруг которого группируются элементы выборки. При увеличении числа наблюдений среднее приближается к математическому ожиданию.

Оценкой *математического ожидания (M, μ, a , в англоязычной литературе принято обозначение E)* является *выборочное среднее* - среднее арифметическое вариант.

Выборочная медиана (Me) - это число, которое является серединой выборки, то есть половина чисел имеет значения большие, чем медиана, а половина чисел имеет значения меньшие, чем медиана. Для нахождения медианы обычно выборку ранжируют — располагают элементы в порядке возрастания. Если количество членов ранжированного ряда нечетное, медианой является значение ряда, которое расположено посередине, то есть элемент с номером $(n + 1)/2$. Если число членов ряда четное, то медиана равна среднему членов ряда с номерами $n/2$ и $n/2 + 1$.

$$Me = \begin{cases} x_{(n+1)/2}, & n = 2k + 1 \\ \frac{x_{n/2} + x_{n/2+1}}{2}, & n = 2k \end{cases} \quad (5)$$

Основными показателями рассеяния вариант являются интервал, дисперсия выборки, стандартное отклонение и стандартная ошибка.

Интервал (амплитуда, вариационный размах) - это разница между максимальным и минимальным значениями элементов выборки. Интервал является простейшей и наименее надежной мерой вариации или рассеяния элементов в выборке.

Более точно отражают рассеяние показатели, учитывающие не только крайние, но и все значения элементов выборки.

Дисперсией выборки, или выборочным аналогом дисперсии, называется величина

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (6)$$

Дисперсия выборки (D, s^2, σ^2) — это параметр, характеризующий степень разброса элементов выборки относительно среднего значения. Чем больше дисперсия, тем дальше отклоняются значения элементов выборки от среднего значения.

Среднее квадратичное отклонение определяется как обобщающая характеристика размеров вариации признака в совокупности. Среднее квадратичное отклонение — это квадратный корень из среднего арифметического всех квадратов разностей между данными величинами и их средним арифметическим. Среднее квадратичное отклонение принято обозначать греческой буквой сигма σ .

Синонимы: среднеквадратическое отклонение, квадратичное отклонение.

Близкие термины: стандартное отклонение, стандартный разброс.

В теории вероятности и статистике этот параметр (среднеквадратическое отклонение), является показателем степени разброса элементов выборки (рассеяния значений случайной величины) относительно математического ожидания. При ограниченных массивах выборок значений вместо математического ожидания используется среднее арифметическое значение совокупности выборок. Чем больше среднее квадратичное отклонение, тем дальше отклоняются значения элементов выборки от среднего значения.

Среднеквадратическое отклонение измеряется в единицах измерения самой случайной величины и используется при расчете стандартной ошибки среднего арифметического, при построении доверительных интервалов, при статистической проверке гипотез, при измерении линейной взаимосвязи между случайными величинами. Определяется как квадратный корень из дисперсии случайной величины.

Среднеквадратическое отклонение:

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_i)^2} \quad (7)$$

Стандартное отклонение, s (*standard deviation*) - оценка среднеквадратического отклонения случайной величины относительно ее математического ожидания на основе несмещенной ее дисперсии.

$$s = \sqrt{\frac{n}{n-1} \sigma^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_i)^2} \quad (8)$$

где σ^2 — дисперсия; x_i — i -й элемент выборки; n — объём выборки; \bar{x}_i — среднее арифметическое выборки.

Следует отметить, что обе оценки являются смещёнными. В общем случае несмещённую оценку построить невозможно. Однако оценка на основе оценки несмещённой дисперсии является состоятельной.

Следует отметить отличие стандарта (в знаменателе $n - 1$) от корня из среднеквадратического отклонения (в знаменателе n), при малом объёме выборки оценка дисперсии через последнюю величину является несколько смещённой, при бесконечно большом объёме выборки разница между указанными величинами исчезает.

Стандартная ошибка, S (standard error), или ошибка среднего — это параметр, характеризующий стандартное отклонение выборочного среднего, рассчитанное по выборке размера n из генеральной совокупности.

Стандартная ошибка среднего вычисляется как частное от деления среднеквадратического отклонения на квадратный корень из объема выборки (или как корень из частного от деления дисперсии на объем выборки).

Поскольку дисперсия генеральной совокупности, как правило, неизвестна, то оценка стандартной ошибки вычисляется как частное от деления стандартного отклонения на квадратный корень из объема выборки.

Показателями, характеризующими форму распределения, являются выборочные эксцесс и асимметрия.

Эксцесс (Ek, Kurt) — это степень выраженности «хвостов» распределения, то есть частоты появления удаленных от среднего значений.

Оценка эксцесса

$$Kurt = \left(\frac{1}{ns^4} \sum_{i=1}^n (x_i - \bar{x})^4 \right) - 3 \quad (9)$$

характеризует «островершинность» (при $Kurt > 0$) или «плосковершинность» (при $Kurt < 0$) распределения по сравнению с нормальным.

Асимметрия (As, Skew) — величина, характеризующая несимметричность распределения элементов выборки относительно среднего значения. Принимает значения от -1 до 1. В случае симметричного распределения асимметрия равна 0.

Оценка асимметрии

$$Skew = \frac{1}{ns^3} \sum_{i=1}^n (x_i - \bar{x})^3 \quad (10)$$

характеризует «скос» распределения относительно его «центра» в положительном или отрицательном направлениях, соответственно.

Часто значения асимметрии и эксцесса используют для проверки гипотезы о том, что данные (выборка) принадлежат к определенному теоретическому распределению, в частности, нормальному распределению.

Для нормального распределения асимметрия и эксцесс равны нулю.

В результате наблюдений или эксперимента получают наборы данных, называемые выборками. Для проведения их анализа данные подвергаются статистической обработке. Первое, что всегда делается при обработке данных, это вычисление элементарных статистических характеристик выборок (как минимум: среднего, среднеквадратичного отклонения, ошибки среднего) по каждому параметру и по каждой группе. Полезно также вычислить эти характеристики для объединения родственных групп и суммарно по всем данным.

1.2 Диаграмма разброса (рассеяния)

Диаграмма рассеяния (разброса, поле корреляции) — инструмент, позволяющий выявить вид и степень зависимости (корреляцию) между парами переменных x , y , которые могут представлять:

- характеристику качества и воздействующий на нее фактор;
- две характеристики качества;
- два фактора, воздействующие на одну и ту же характеристику качества.

Сама диаграмма представляет собой множество (совокупность) точек, координаты которых равны значениям параметров x и y .

Алгоритм построения:

1. Сведите полученные значения пар данных x , y в таблицу для удобства дальнейшего использования. Для получения достоверного результата рекомендуется использовать не менее 30 пар данных.
2. Постройте горизонтальную и вертикальную оси. Для удобства прочтения графика рекомендуется выбрать масштаб и диапазон шкал для осей таким образом, чтобы размер

рабочих частей осей для отображения полученных значений пар x , y примерно совпадал. В случае если одна из переменных – фактор, а другая – характеристика качества, рекомендуется для фактора выбрать ось x , а для характеристики качества – ось y .

3. Нанесите точки полученных пар значений x , y на график. Если в результате разных наблюдений получены одинаковые пары значений x , y , то отметьте эти точки каким-либо знаком или поставьте рядом вторую точку.

4. Вычислите коэффициент корреляции (он позволяет количественно определить силу линейной связи между x и y) по формуле:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (11)$$

где n – количество пар данных,
 \bar{x} – среднее арифметическое значение параметра x ,
 \bar{y} – среднее арифметическое значение параметра y .

Проверьте, что значение полученного коэффициента корреляции не выходит за пределы $-1 < r < +1$. Если при подсчете получено абсолютное значение r больше 1, значит, в вычислениях произошла ошибка и коэффициент корреляции необходимо пересчитать.

5. Определите вид связи между x и y , проведя анализ формы построенного графика и вычисленного коэффициента корреляции.

Типичные разновидности диаграммы рассеяния:

1. Ярко выраженная тенденция увеличения y с увеличением x соответствует сильной положительной корреляции.

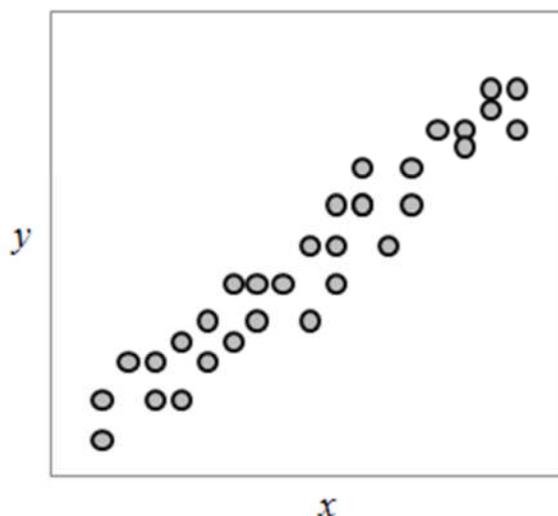


Рисунок 1.1 – Положительная корреляция

2. Ярко выраженная тенденция уменьшения y с увеличением x соответствует сильной отрицательной корреляции.

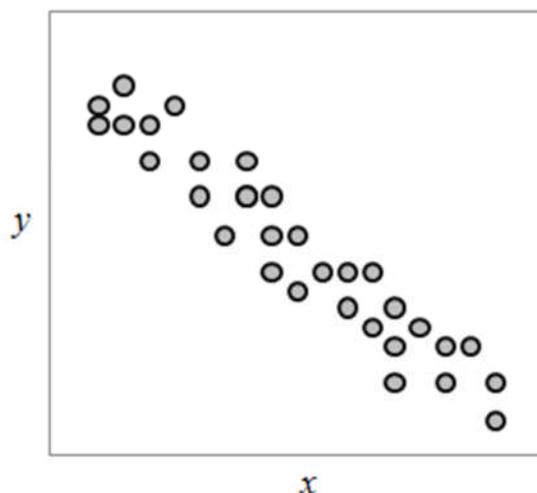


Рисунок 1.2 – Отрицательная корреляция

3. Слабо выраженная тенденция увеличения y с увеличением x свидетельствует о слабой положительной корреляции.

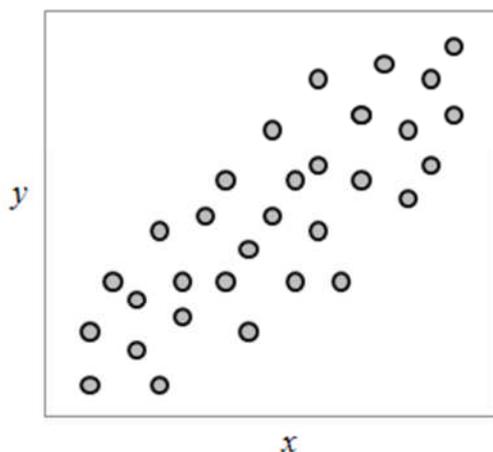


Рисунок 1.3 – Слабая положительная корреляция

4. Слабо выраженная тенденция уменьшения y с увеличением x свидетельствует о слабой отрицательной корреляции.

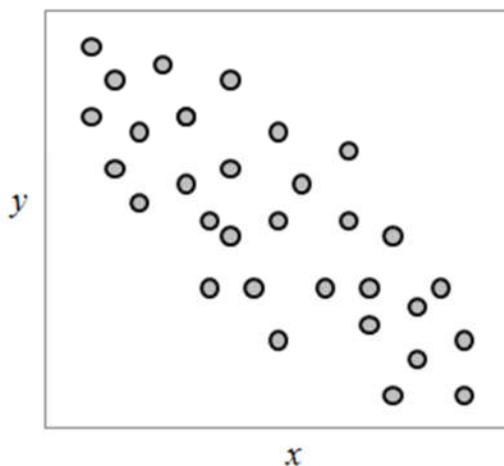


Рисунок 1.4 – Слабая отрицательная корреляция

5. Наблюдаемая тенденция нелинейного изменения y с увеличением x соответствует криволинейной корреляции.

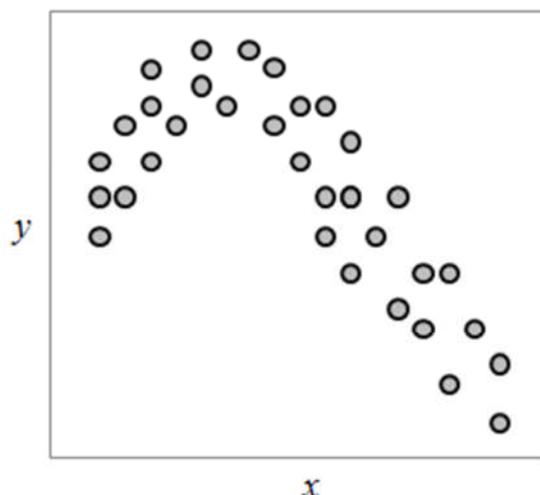


Рисунок 1.5 – Криволинейная корреляция

6. Отсутствие наблюдаемой на графике зависимости между x и y свидетельствует об отсутствии корреляции.

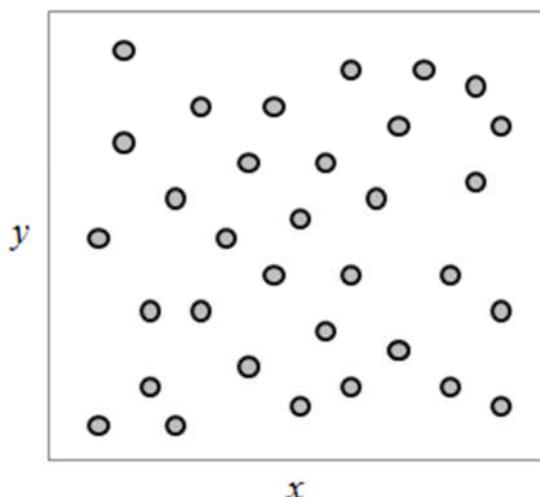


Рисунок 1.6 – Отсутствие корреляции

Тип связи между x и y по значению коэффициента корреляции оценивается следующим образом: Значение $r > 0$ соответствует положительной корреляции, $r < 0$ – отрицательной корреляции. Чем больше абсолютное значение r , тем сильнее корреляция, а $|r| = 1$ соответствует точной линейной зависимости между парами значений наблюдаемых переменных. Чем меньше абсолютное значение r , тем слабее корреляция, а $|r| = 0$ свидетельствует об отсутствии корреляции. Абсолютное значение r близкое к 0 может быть также получено при определенном виде криволинейной корреляции.

Для увеличения результативности следует строить и производить сравнение графиков рассеяния, полученных в разные моменты времени. Также рекомендуется проводить стратификацию диаграмм разброса для различных средств и условий производства продукции.

Источник: <http://www.tools-quality.ru/index.php/q7/dispersion-diagramme>

1.3 Модели временного ряда

Модели временного ряда включают несколько составляющих: тенденция (тренд), сезонность, циклы, нерегулярные изменения, случайная составляющая.

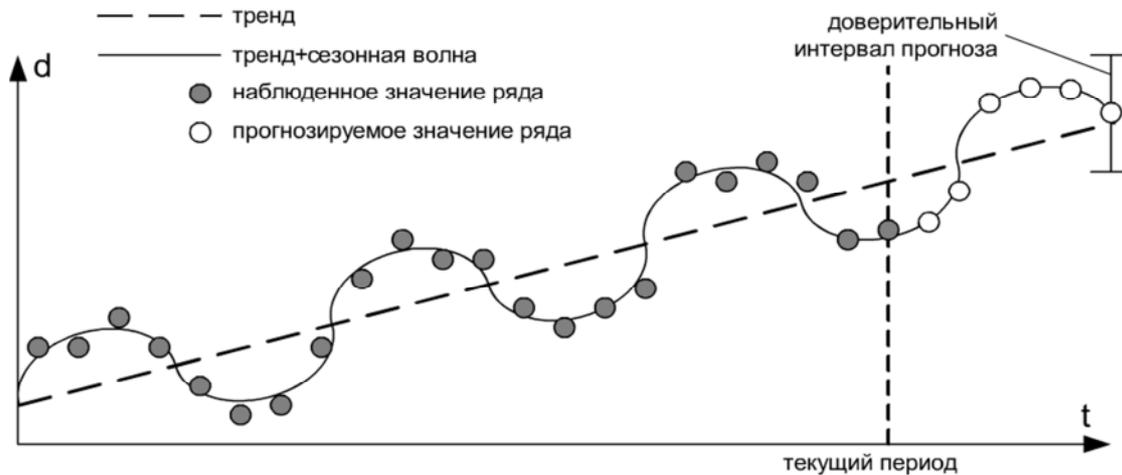


Рисунок 1.7 - Составляющие временного ряда

Временной ряд можно считать состоящим из двух частей: детерминированная и случайная.

Таблица 1.1 - Временной ряд

Детерминированная составляющая			Случайная составляющая $e(t)$ (ε)			
Тренд $f(t)$	Циклическая компонента $u(t)$	Сезонная компонента $s(t)$	«Белый шум»	Авторегрессия	Скользящее среднее	Смешанная

1.4 Коэффициенты регрессии, корреляции и детерминации

При исследовании зависимости между двумя переменными чаще всего используется линейная форма связи. Это связано с двумя обстоятельствами:

- 1) чёткая экономическая интерпретация параметров линейной модели регрессии;
- 2) в большинстве случаев нелинейные модели регрессии преобразуются к линейному виду.

Общий вид модели парной регрессии зависимости переменной y от переменной x

$$y_i = b_0 + b_1 x_i + \varepsilon_i \quad (12)$$

где

y_i – результативные переменные, $i = \overline{1, n}$;

x_i – факторные переменные, $i = \overline{1, n}$;

b_0, b_1 – параметры модели регрессии, подлежащие оцениванию;

ε_i – случайная ошибка модели регрессии.

Данная величина является случайной, она характеризует отклонения реальных значений результативных переменных от теоретических, рассчитанных по уравнению регрессии.

Присутствие случайной ошибки в модели регрессии порождено следующими источниками:

- 1) нерепрезентативность выборки. Модель парной регрессии в большинстве случаев является большим упрощением истинной зависимости между переменными, потому что в

модель входит только одна факторная переменная, не способная полностью объяснить вариацию результативной переменной. При этом результативная переменная может быть подвержена влиянию множества других факторных переменных в гораздо большей степени;

- 2) ошибки, возникающие при измерении данных;
- 3) неправильная функциональная спецификация модели.

Коэффициент b_1 , входящий в модель парной регрессии, называется коэффициентом регрессии. Он характеризует, на сколько в среднем изменится результативная переменная y при условии изменения факторной переменной x на единицу своего измерения.

Знак коэффициента регрессии указывает на направление связи между переменными:

- 1) если $b_1 > 0$, то связь между изучаемыми переменными прямая (с уменьшением факторной переменной x уменьшается и результативная переменная y , и наоборот);
- 2) если $b_1 < 0$, то связь между изучаемыми переменными обратная (с увеличением факторной переменной x результативная переменная y уменьшается, и наоборот).

Коэффициент b_0 , входящий в модель парной регрессии, трактуется как среднее значение результативной переменной y при условии, что факторная переменная x равна нулю. Но если факторная переменная не имеет и не может иметь нулевого значения, то подобная трактовка коэффициента b_0 не имеет смысла.

Оценка тесноты связи с помощью показателей корреляции (выборочный коэффициент корреляции) и детерминации.

Уравнение регрессии всегда дополняется показателем тесноты связи. При использовании линейной регрессии в качестве такого показателя выступает *линейный коэффициент корреляции* r_{xy} . Существует несколько видов формулы линейного коэффициента корреляции, основные из них:

$$r_{xy} = b_1 \frac{\sigma_x}{\sigma_y} = \frac{cov(x, y)}{\sigma_x * \sigma_y} = \frac{\overline{x * y} - \bar{y} * \bar{x}}{\sigma_x * \sigma_y} \quad (13)$$

где b_1 — коэффициент регрессии, σ_x, σ_y — среднеквадратическое отклонение соответствующего факторной переменной (факторного признака).

Ковариация $cov(x, y)$ определяется как математическое ожидание произведения отклонений случайных величин.

Корреляционная связь между переменными называется прямой, если $r_{xy} > 0$, и обратной, если $r_{xy} < 0$.

Для практических расчетов наиболее удобна формула

$$r_{xy} = \frac{n * \sum_{i=1}^n y_i x_i - \sum_{i=1}^n y_i * \sum_{i=1}^n x_i}{\sqrt{(n * \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2) * (n * \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2)}} \quad (14)$$

так как по ней коэффициент корреляции находится из данных наблюдений, и на значение r_{xy} не оказывает влияния погрешность округления.

Коэффициент корреляции принимает значения от -1 до +1.

При значении коэффициента корреляции равном ± 1 связь представлена линейной функциональной зависимостью. При этом все наблюдаемые значения располагаются на линии регрессии.

При $r_{xy} = 0$ корреляционная связь между признаками в линейной форме отсутствует. При этом линия регрессии параллельна оси Ox .

При $r_{xy} > 0$ – корреляционная связь между переменными называется прямой, а при $r_{xy} < 0$ – обратной.

Для характеристики силы связи можно использовать шкалу Чеддока.

Таблица 1.2 - Шкала Чеддока

Показатель тесноты связи	0,1 – 0,3	0,3 – 0,5	0,5 – 0,7	0,7 – 0,9	0,9 – 0,99
Характеристика силы связи	Слабая	Умеренная	Заметная	Высокая	Весьма высокая

Для оценки качества подбора линейной функции рассчитывается квадрат линейного коэффициента корреляции r_{xy}^2 , называемый *коэффициентом детерминации*. Коэффициент детерминации обозначим R^2 , таким образом, имеем

$$R^2 = r_{xy}^2 \quad (15)$$

Коэффициент детерминации характеризует долю дисперсии результативного признака y , объясняемую регрессией, в общей дисперсии результативного признака. Соответственно, величина $1-R^2$ характеризует долю дисперсии y , вызванную влиянием остальных, не учтенных в модели факторов.

Замечание. Вычисление R^2 корректно, если константа включена в уравнение регрессии.

2 Лабораторная работа №1

Тема: Описательная статистика. Проверка гипотез о виде распределения случайной выборки.

Цель: Освоить основные принципы работы с приложениями MS Excel, EViews и Statistica при проверке гипотез о виде распределения случайной выборки.

2.1 Задания

Задание 1.1 Скопировать исходные данные (файлы, которые созданы в Блокноте) из материалов для студентов (\\...ForStudent\КИС...).

Задание 1.2 Построить графики значений временных рядов. Произвести расчет элементарных статистических характеристик. Определить, является ли совокупность однородной. Осуществить проверку гипотезы о соответствии реального распределения значений временных рядов нормальному закону распределения с помощью тестов: χ^2 -Пирсона, Колмогорова-Смирнова, Жака-Бера.

Рассмотрим примеры.

Пример 1.1 Пусть имеются следующие наблюдения (временной ряд): темпы роста ВВП Беларуси, (%), представленные в табл.2.1.

Таблица 2.1 - Темпы роста ВВП Беларуси, %

Год	Значение	Год	Значение
1993	-7.6	2003	7.0
1994	-11.7	2004	11.5
1995	-11.3	2005	9.4
1996	2.8	2006	10.0
1997	11.4	2007	8.6
1998	8.4	2008	10.2
1999	3.4	2009	0.2
2000	5.8	2010	7.7
2001	4.7	2011	5.5
2002	5.0	2012	1.5

Источник - CIA World Factbook

Введем обозначение: пусть x – темпы роста ВВП Беларуси, (%). Построить график данных и произвести его анализ. Произвести расчет и анализ описательной статистики. Определить, является ли выборка количественно однородной по данному признаку. Произвести проверку соответствия выборки нормальному закону распределения при уровне значимости $\varepsilon = 5\%$.

<http://www.ereport.ru/stat.php?razdel=country&count=belarus&table=ggecia&time=1>

Пример 1.2 Формирование последовательностей случайных чисел по нормальному закону распределения.

Пример 1.3 Сформировать выборку из 100 случайных величин, лежащих в диапазоне от 5 до 30. Построить точечную диаграмму, отображающую значения выборки из 100 случайных величин. Произвести расчет и анализ описательной статистики. Определить, является ли выборка количественно однородной по данному признаку. Произвести проверку соответствия выборки нормальному закону распределения при уровне значимости $\varepsilon = 5\%$.

2.2 Методические указания

2.2.1 Понятие выборки, временного ряда

Выборкой называется подмножество элементов, выделяемых из множества, т.е. из генеральной совокупности и отличается от подмножеств, выделенных аналогичным образом.

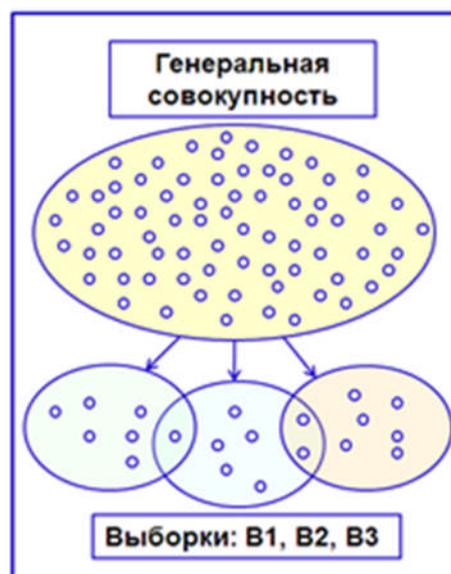
Классификация выборок по их объему:

$n=6\div 12$ — малые выборки,

$n=20\div 60$ — средние выборки,

$n=200\div 600$ — представительные выборки.

Малые выборки ввиду их слабой информативности используются очень редко.



Временной ряд - это совокупность значений какого-либо показателя за несколько последовательных моментов или периодов.

Использование критериев согласия является частным случаем более широкого круга задач статистического анализа, который обобщается названием «Проверка статистических гипотез». Под анализом нулевой гипотезы понимается соответствие реального закона распределения конкретной выборки непрерывных данных теоретическому — нормальному закону распределения.

2.2.2 Настройка Пакета анализа данных

Вкладка *Файл* \ *Параметры* \ *Надстройки*. Появится окно *Параметры Excel*. В нижней части окна *Параметры Excel* нажать кнопку *Перейти...* .

Появится окно *Надстройки*. Установить флажок *Пакет анализа* и нажать *ОК*. На вкладке *Данные* появится группа *Анализ*. При выборе инструмента *Анализ данных* появляется окно диалога *Анализ данных*. В окне диалога *Анализ данных* отображается список инструментов.

2.2.3 Инструмент Гистограмма Пакета анализа данных

Выполним команду вкладка *Данные* \ группа *Анализ* \ *Анализ данных*. Появится окно диалога *Анализ данных*, в котором выберем опцию *Гистограмма*. На экране появится окно диалога *Гистограмма*.

Инструмент *Гистограмма* вычисляет частоты появления данных. Числовой промежуток между наименьшим и наибольшим значениями данных делится на интервалы равной ширины. Под частотой понимается количество чисел, попавших в такой интервал.

Параметр *Интервал карманов (отрезков)* является необязательным. Это блок ячеек, содержащих граничные значения отрезков, для которых определяются частоты (вероятности) попадания случайной величины. Если интервал карманов не задается, автоматически создается набор равных отрезков в диапазоне от минимума до максимума.

Excel вычисляет число данных между текущим началом отрезка и соседним большим по порядку, если такой есть. При этом включаются значения на нижней границе

отрезка и не включаются значения на верхней. Например, отрезки могут быть: меньше либо равно 2, от 2 до 3 включительно, от 3 до 4 включительно, от 4 до 5 включительно, и больше 5.

Парето (Отсортированная Гистограмма) – значения частот в выходной таблице и на гистограмме отображаются в порядке их убывания.

Интегральный процент – вычисление интегральной функции распределения. В выходной таблице выводит столбец для интегральных процентных отношений, при этом в гистограмме появится график интегрального процентного отношения.

Установить флажок *Вывод Графика*, если необходимо, чтобы для выходной таблицы была автоматически построена гистограмма. Диаграмма будет расположена на том листе, что и выходная таблица.

При статистическом моделировании и первичной обработке данных используются следующие инструменты: *Генерация случайных чисел*, *Гистограмма*.

2.2.4 Инструмент Генерация случайных чисел

Инструмент *Генерация случайных чисел* заполняет интервал независимыми случайными числами.

При помощи параметра *Число переменных* можно получить многомерную выборку. Для этого введите число столбцов в выходной таблице.

Параметром *Число случайных чисел* определяется число точек данных, которое вы хотите генерировать для каждой переменной.

Выбор закона распределения случайных чисел задаётся параметром *Распределение*.

1. *Равномерное распределение* характеризуется верхней и нижней границами. Вероятность попадания переменной в отрезок фиксированной длины зависит только от ширины отрезка и не зависит от его расположения на интервале. Как правило, в приложениях используют равномерное распределение в интервале $[0,1]$.

2. *Нормальное распределение* характеризуется средним значением и стандартным отклонением. Обычно приложения для этого распределения используют среднее значение 0 и стандартное отклонение 1.

3. *Распределение Бернулли* характеризуется вероятностью успеха в данном испытании. Случайная величина принимает значение 0 или 1. Например, при бросании игральной кости или выпадет 6 очков с вероятностью $1/6$, или выпадет не 6 очков с вероятностью $5/6$, то есть случайная величина принимает значение 1 с вероятностью $1/6$ или 0 с вероятностью $5/6$.

4. *Биноминальное распределение* характеризуется вероятностью успеха для некоторого числа испытаний. Например, можно сгенерировать случайные числа, моделирующие процесс бросания монеты с вероятностью успеха ровно в “k” случаях из “n” испытаний.

5. *Распределение Пуассона* характеризуется значением *Лямбда*, равным $1/\text{среднее}$. Распределение Пуассона часто используется для характеристики числа событий, случающихся в единицу времени, например, число телефонных соединений в минуту.

6. *Модельное распределение* характеризуется нижней и верхней границей, шагом, числом повторений значений и числом повторений последовательности.

7. *Дискретное распределение* характеризуется значением и связанным с ним интервалом вероятности. Интервал должен содержать два столбца: левый содержит значения, правый – вероятности, связанные со значением в данной строке. Сумма вероятностей должна быть равна 1.

При помощи параметра *Случайное рассеяние* можно зафиксировать последовательность выводимых случайных чисел. При повторных запусках генератора можно использовать это значение для получения тех же самых случайных чисел.

2.2.5 Нормальный закон распределения

Нормальное распределение (распределение Гаусса) является предельным случаем почти всех реальных распределений вероятности. Этому закону подчиняется, при соблюдении определенных условий, распределение суммы достаточно большого числа случайных величин, каждая из которых может иметь произвольное распределение. Нормальное распределение задается функцией плотности вероятности, совпадающей с функцией Гаусса:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (16)$$

где параметр $\mu=M(\xi)=a$ - математическое ожидание (либо среднее значение), медиана и мода распределения, а параметр $\sigma=\sigma(\xi)$ - среднеквадратическое отклонение ($\sigma^2=D(\xi)$ - дисперсия) распределения, ξ - случайная величина.

Таким образом, распределение Гаусса зависит от двух параметров: математическое ожидание (μ) и среднеквадратичное отклонение (σ). Кривая плотности вероятности нормального распределения имеет вид симметричной колоколообразной кривой, распространяющейся по всей числовой оси.

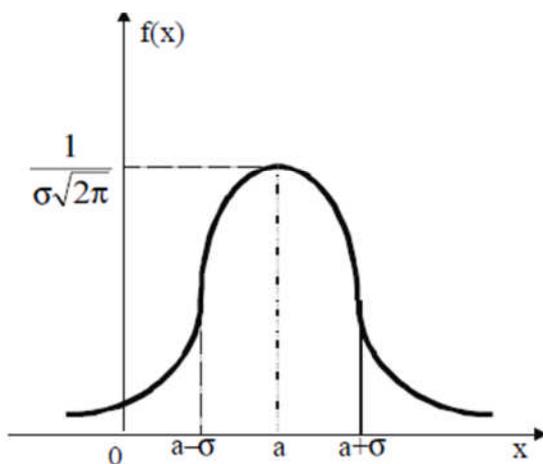


Рисунок 2.1 - График плотности нормального распределения

Стандартным нормальным распределением называется нормальное распределение с математическим ожиданием μ (или M) =0 и стандартным отклонением σ (или s)=1.

Показателями, характеризующими форму распределения, являются выборочные эксцесс (**Ek**) и асимметрия (**As**).

Коэффициент *асимметрия* (Skewness, **As**, γ) характеризует степень несимметричности распределения относительно среднего вправо (>0) и влево (<0). Принимает значения от -1 до 1. В случае симметричного распределения асимметрия равна 0. Чем больше величина $|As|$, тем более асимметрично распределение. Оценочная шкала асимметрии: $|As| \leq 0,25$ - асимметрия незначительная; $0,25 < |As| \leq 0,5$ - асимметрия заметная (умеренная); $|As| > 0,5$ - асимметрия существенная.

Коэффициент *эксцесс* (Kurtosis, **Ek**, ϵ) характеризует степень остроконечности или сглаженности «хвостов» распределения (это степень выраженности «хвостов» распределения, то есть частоты появления удаленных от среднего значений).

Свойство 1. Нормальное распределение является симметричным. Кривая плотности вероятности симметрична относительно математического ожидания.

Следствия симметрии:

$$M(X) = Mo(X) = Me(X),$$

Отсутствие скошенности: $Ek = 0$.

Свойство 2. «Идеальная» форма вершины, взята за эталон: $As = 0$.

Статистический пакет Statistica		EViews	
Кривая плотности вероятности распределения			
Экцесс >0	островершинная	Экцесс >3	островершинная
Экцесс <0	плосковершинная	Экцесс <3	плосковершинная

Значения коэффициентов асимметрии (*Skewness*), эксцесса (*Kurtosis*) не превышают двукратных среднеквадратических ошибок (*Standard error of skewness, Standard error of kurtosis*). Это указывает на отсутствие значимой скошенности и остро-(плоско)вершинности фактического распределения значений случайных величин выборки по сравнению с нормальным распределением.

Зрительно оценить закон распределения случайных величин можно при визуальном анализе гистограммы.

Предположим, что в нашем распоряжении результаты наблюдений над непрерывной случайной величиной X , оформленные в виде простой статистической совокупности.

Алгоритм построения гистограммы:

1. Разделить весь диапазон эмпирических значений x на интервалы (группы, разряды): вычислить

- диапазон данных (выборочного размаха) $R = x_{max} - x_{min}$
- число интервалов k с помощью формулы Стерджесса
- размер или ширину интервалов (карман) $d = R/k$.

2. Определить границы интервалов.

Сначала определяют нижнюю границу первого интервала и прибавляют к ней ширину этого интервала, чтобы получить границу между первым и вторым интервалами. Далее продолжают прибавлять найденную ширину интервала d к предыдущему значению для получения второй границы, затем третьей и т. д. После завершения такой работы можно удостовериться, что верхняя граница последнего интервала совпадает с максимальным значением x_{max} .

3. Подсчитать частоты, т.е. количество значений m_i , приходящееся на каждый i -й разряд ($\sum_{i=1}^k m_i = n$).

4. Вычислить относительные частоты, т.е. разделить значения m_i на общее число наблюдений n . Сумма частот всех разрядов, очевидно, должна быть равна единице.

$$\tilde{p}_i = \frac{m_i}{n}, \sum_{i=1}^k \tilde{p}_i = 1$$

5. Построить таблицу, в которой приведены разряды в порядке их расположения вдоль оси абсцисс и соответствующие частоты. Эта таблица называется статистическим рядом.

Таблица 2.2 - Статистический ряд

k_i	$x_1; x_2$	$x_2; x_3$...	$x_i; x_{i+1}$...	$x_k; x_{k+1}$
\tilde{p}_i	\tilde{p}_1	\tilde{p}_2	...	\tilde{p}_i	...	\tilde{p}_k

Здесь k_i - обозначение i -го разряда; $x_i; x_{i+1}$ - его границы; \tilde{p}_i - соответствующая частота; k - число разрядов.

6. Для построения полигона по данным вариационного ряда с равными интервалами на оси абсцисс откладывают значения аргумента, а на оси ординат - значения частот или относительных частот. Гистограммой называют диаграмму, состоящую из вертикальных прямоугольников, основаниями которых являются интервалы шириной $d = \Delta x = R/k$, а высоты равны частоте m_i , относительной частоте m_i/n , $m_i/n\%$, отношению $m_i/\Delta x$ или $m_i/\Delta x n$ для соответствующих интервалов.

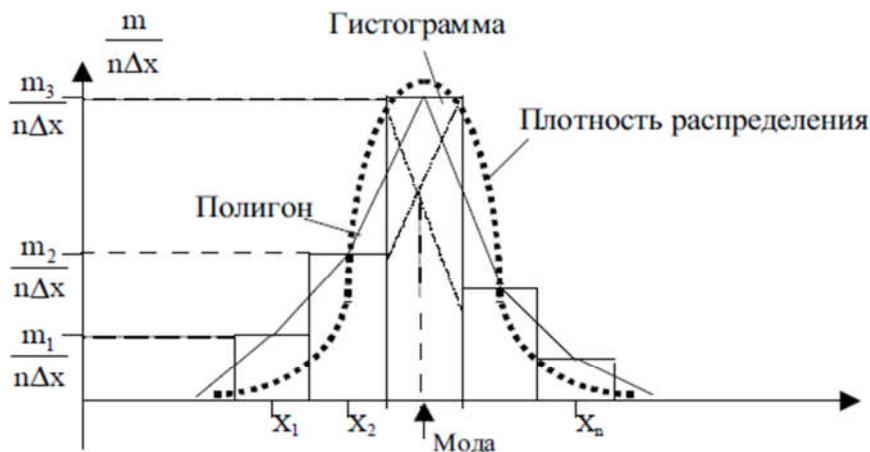


Рисунок 2.2 - Гистограмма

Источник: http://www.bsmu.by/downloads/kafedri/k_fiziki/2013-2/20120917110323vexcel.pdf

Имеется калькулятор: <http://planetcalc.ru/936/>

2.2.6 Правило трех сигм

Вероятность того, что случайная величина отклонится от своего математического ожидания на величину, большую, чем утроенное среднее квадратичное отклонение, практически равна нулю.

Считается, что если для какой-либо случайной величины выполняется правило трех сигм, то эта случайная величина имеет нормальное распределение.

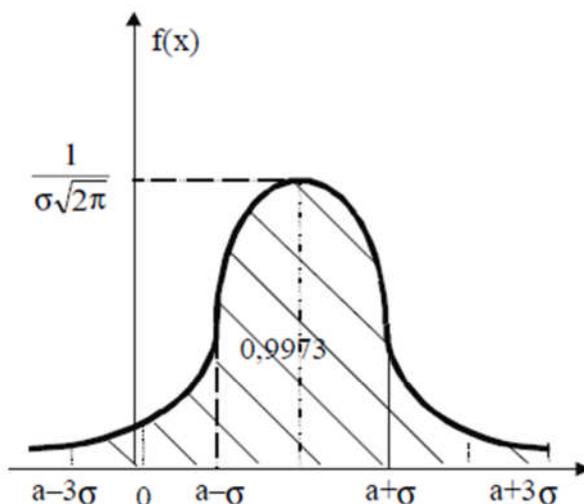


Рисунок 2.3 - Правило трех сигм

2.2.7 Анализ однородности выборки

Одним из важных вопросов, возникающих при анализе выборки, является вопрос: относится та или иная варианта (случайная величина) к данной статистической совокупности. Решение вопроса не представляет сложности, если распределение в этой совокупности является нормальным. Для этого достаточно использовать правило трех сигм. Согласно этому правилу, в пределах $\mu \pm 3\sigma$ находится 99,7% всех вариантов (μ – математическое ожидание, σ – среднее квадратичное отклонение). Поэтому, если варианта попадает в этот интервал, то она считается принадлежащей к данной совокупности. Если не попадает, то она может быть отброшена. Хотя этот метод и предполагает нормальность исходного распределения, на практике в большинстве случаев может быть использован.

Для нормальных и близких к нормальным распределениям показатель V (коэффициент вариации) служит индикатором однородности выборочных наблюдений: принято считать, что при выполнении неравенства $V_x \leq 33\%$ (x –признак) выборка является количественно однородной по данному признаку. Чем меньше его значение, тем лучше.

V_x	Вывод об однородности выборки
<17%	абсолютно однородная
17–33%	достаточно однородная
35–40%	недостаточно однородная
40–60%	большая колеблемость выборки

$$\text{Коэффициент Вариации} = \frac{\text{StandartDeviation}}{\text{Mean}} 100\%$$

2.2.8 Технология вычислений в MS Excel

Пример 1.2 Генерация последовательности случайных чисел (по нормальному закону распределения).

- Выполните команду: вкладка *Данные* \ группа *Анализ* \ *Анализ данных*. Появится окно диалога *Анализ данных*.

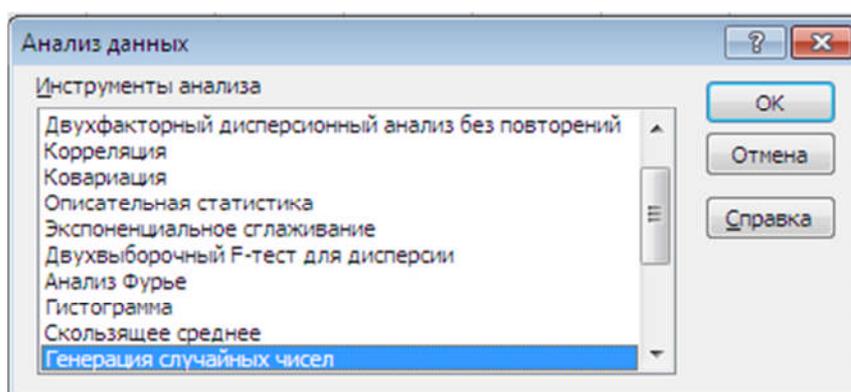


Рисунок 2.4 - Генерация случайных чисел

- Выберите опцию *Генерация случайных чисел* из списка инструментов *Анализа данных* и нажмите кнопку *ОК*. На экране появится окно диалога *Генерация случайных чисел*.
- Значения параметров в диалоговом окне установите следующим образом:
Число переменных – введите 1.
Число случайных чисел – введите 100.

Распределение – выберите *Нормальное*. Появится окно диалога для ввода параметров нормального распределения. *Среднее* – введите 0. *Стандартное отклонение* - введите 1.

Параметры вывода - установите флажок *Выходной интервал* и выберите ячейку A2, начиная с которой будут выводиться на экран случайные числа.

Нажмите кнопку *ОК*. На экране появятся генерируемые инструментом случайные числа.

Примечание. Имеется калькулятор (<http://www.stathelp.ru/ots/g5p6.html>), который помимо построения гистограммы с использованием числа классов, полученных по формуле Стерджесса, может построить гистограммы с числом классов по Скотту и Фридману/Диаконису, а также с числом классов, произвольно заданных пользователем.

Пример 1.3 Графическое представление данных. Произвести расчет и анализ описательной статистики. Определить, является ли совокупность однородной. Проверка соответствия выборки нормальному закону распределения при уровне значимости 5%. Выполнить на новом листе.

1. Сгенерировать выборку случайных чисел на новом листе. (Можно использовать рекомендации реализации примера 1.2).

- Ввести в ячейку A1 формулу: =СЛУЧМЕЖДУ(5;30). Нажатие *Enter* переводит на ячейку A2.

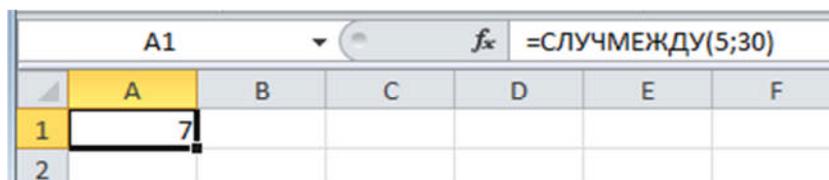


Рисунок 2.5 - Функция СЛУЧМЕЖДУ()

- Заполнить остальные ячейки удобно с помощью автозаполнения. Необходимо выделить ячейку A1, подвести курсор к маркеру автозаполнения и, удерживая левую клавишу мыши, протянуть маркер до ячейки A100.

Полученные значения скопировать как значения в новый диапазон.

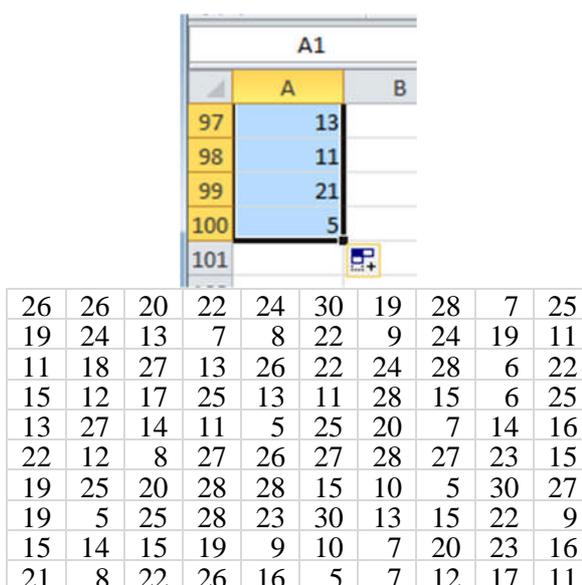


Рисунок 2.6 - Автозаполнение

2. Графически представить числовые данные (построить точечную диаграмму).

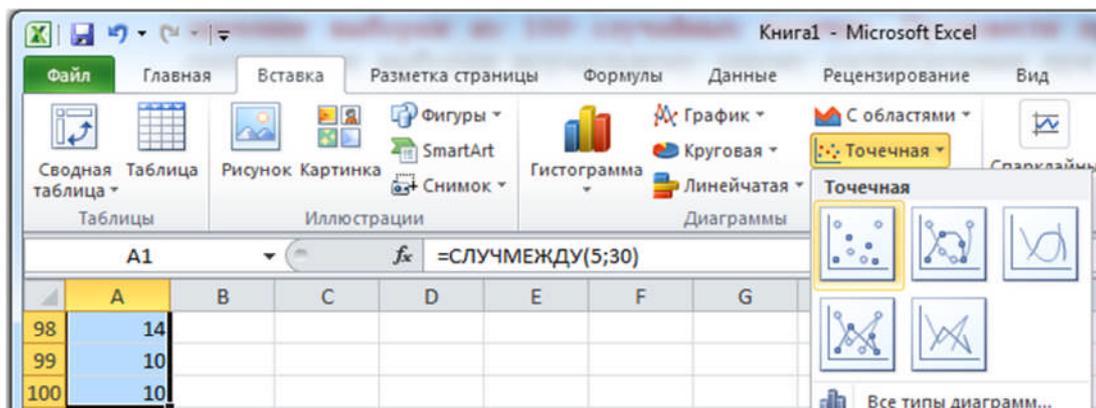


Рисунок 2.7 - Выбор типа диаграммы Точечная

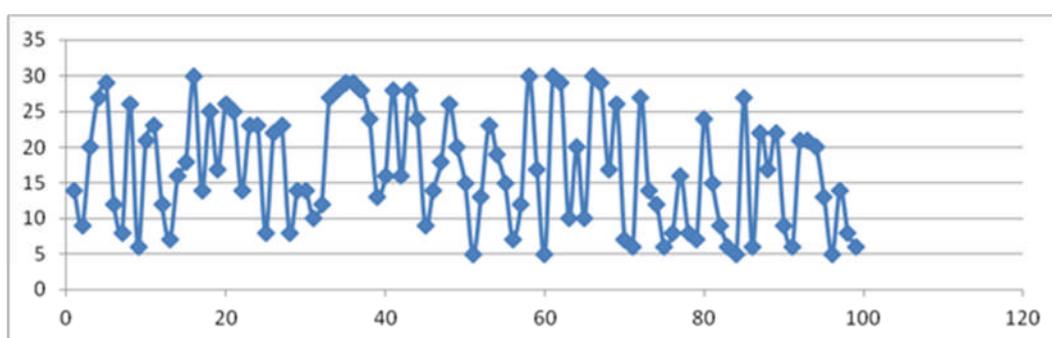


Рисунок 2.8 - Графическое представление числовых данных

Для построения выборочной функции распределения в MS Excel используется инструмент ГИСТОГРАММА из Пакета анализа или функция ЧАСТОТА. При этом весь диапазон изменения случайной величины разбивают на интервалы равной ширины, называемые карманами. Число карманов обычно 5-15. Вычисляется число попаданий значений случайной величины в каждый карман. По ним вычисляются статистические частоты – отношение числа попаданий в карман m к общему числу испытаний n (m/n), по которым и строится гистограмма выборочной функции распределения статистических вероятностей.

3. Произвести расчет элементарных статистических характеристик.

1 способ: использование Пакета анализа данных:

Вкладка *Данные* \ группа *Анализ* \ Анализ данных \ (Описательная статистика и ОК. Появится диалоговое окно Описательная статистика.

Значения параметров в диалоговом окне установите следующим образом:

Входной интервал – указать диапазон рабочего листа, содержащий выборку (например, \$A\$1:\$A\$100).

Группирование – выбрать по столбцам.

Параметры вывода - *Выходной интервал* и выберите ячейку B1.

Итоговая статистика – установите флажок.

Уровень надежности - установите флажок.

Нажмите кнопку ОК. На экране появятся вычисленная инструментом случайные числа описательная статистика выборки случайных чисел.

Результат расчета элементарных статистических характеристик представлен в таблице 2.3.

Таблица 2.3 - Значения, вычисляемые средством Описательная статистика

Значение	Описание	Результат
Среднее	Выборочное среднее	18,03
Стандартная ошибка	Оценка среднеквадратического отклонения выборочного среднего	0,741068707
Медиана	Значение медианы, т.е. квантиля порядка 0,05	19
Мода	Значение моды – если нет одинаковых выборочных значений, то возвращается значение ошибки одинаковых выборочных значений, то возвращается значение ошибки #Н/Д.	15
Стандартное отклонение	Оценка среднеквадратического (или стандартного) отклонения генеральной совокупности	7,410687069
Дисперсия выборки	Оценка дисперсии генеральной совокупности	54,91828283
Эксцесс	Выборочный коэффициент эксцесса	-1,230777562
Асимметричность	Выборочный коэффициент асимметрии	-0,153634571
Интервал	Размах выборки. Выявляется как разность между максимальным и минимальным выборочными значениями	25
Минимум	Минимальное выборочное значение	5
Максимум	Максимальное выборочное значение	30
Сумма	Сумма выборочных значений	1803
Счет	Объем выборки	100
Уровень надежности (X%)	Граница доверительного интервала для неизвестного математического ожидания с доверительным уровнем X%. При малых выборках к данному показателю следует относиться осторожно.	1,47044109

2 способ: использование специальных функций

Таблица 2.4 - Функции

Статистические функции	Обозн.	Пояснение
=ОКРВВЕРХ(1+3,32*log10(n);1)	k	Количество диапазонов по формуле Стерджесса
=ОКРУГЛВВЕРХ(($x_{min} - x_{max}$)/M;1)	$\Delta x/k$	Округленная ширина диапазона
=СРЗНАЧ(X)	\bar{x}	Вычисляет выборочное (или генеральное) среднее, то есть среднее арифметическое значение признака выборочной (или генеральной) совокупности
=СТАНДОТКЛОН.В(X), =КОРЕНЬ(ДИСП.В(X))	s	Вычисляет стандартное отклонение по выборке
=ДИСП()	D, σ^2, s^2	Вычисляет дисперсию

Для оценки разброса данных используются такие статистические характеристики, как среднее квадратическое (или стандартное) отклонение и дисперсия. Стандартное отклонение есть квадратный корень из дисперсии: Большое стандартное отклонение указывает на то, что значения измерения сильно разбросаны относительно среднего, а малое – на то, что значения сосредоточены около среднего.

=МОДА(массив данных)	M_o	Это значение признака, которое чаще других встречается в совокупности данных
=МЕДИАНА()	M_e	Это значение признака, которое разделяет совокупность на две равные по числу элементов части
=СКОС()	A_s, A	Вычисляется асимметрия
=ЭКЦЕСС()	E_k, E	Вычисляется эксцесс
=СЧЁТ(массив данных)	n	Определяется объем выборки
=СЧЁТЕСЛИ(диапазон; критерий)		Используется для построения атрибутивного или вариационного ряда. Аргументами являются диапазон массива выборочных значений признака и критерий – числовое или текстовое значение признака или номер ячейки, в которой оно находится. Результатом является частота появления этого значения в выборке
=ЧАСТОТА(массив данных; массив интервалов)		Построение вариационного ряда. Аргументами являются диапазон массива выборочных данных и столбец интервалов. Если требуется построить дискретный ряд, то здесь указываются значения варианты, если интервальный – то верхние границы интервалов (их еще называют «карманами»). Поскольку результатом является столбец частот, введение функции следует завершить нажатием сочетания клавиш CTRL+SHIFT+ENTER. Заметим, что задавая массив интервалов при введении функции, последнее значение в нем можно и не указывать – в соответствующий «карман» будут помещены все значения, не попавшие в предыдущие «карманы». Иногда это помогает избежать ошибки, состоящей в том, что наибольшее выборочное значение не помещается автоматически в последний «карман»

Примечание. Средние величины могут быть вычислены как по выборке (простая средняя), так и по вариационному ряду (взвешенная средняя) в зависимости от вида исходной информации

Таблица 2.5 - Виды средних величин

Формула средней	Наименование средней			
	Арифметическая	Гармоническая	Геометрическая	Квадратическая
Простая (по выборке)	$\tilde{x} = \frac{\sum x_i}{n}$	$\tilde{x} = \frac{n}{\sum \frac{1}{x_i}}$	$\tilde{x} = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$	$\tilde{x} = \sqrt{\frac{\sum x_i^2}{n}}$
в Excel	СРЗНАЧ(<i>массив</i>)	СРГАРМ(<i>массив</i>)	СРГЕОМ(<i>массив</i>)	КОРЕНЬ((СУММКВ(<i>массив</i>)/n)
Взвешенная (по ВР)	$\tilde{x} = \frac{\sum x_i n_i}{n}$	$\tilde{x} = \frac{n}{\sum \frac{1}{x_i} n_i}$	$\tilde{x} = \sqrt[n]{x_1^{n_1} \cdot x_2^{n_2} \cdot \dots \cdot x_k^{n_k}}$	$\tilde{x} = \sqrt{\frac{\sum x_i^2 n_i}{n}}$
в Excel	Вычисляются с применением функции СУММПРОИЗВ(<i>массив1</i> ; <i>массив2</i>)			

4. Произвести анализ описательной статистики.

Примечание. Для оценки отклонения распределения данных эксперимента от нормального распределения используются такие характеристики как асимметрия A и эксцесс E . Для нормального распределения $As=0$ и $Ek=0$.

Асимметрия показывает, на сколько распределение данных несимметрично относительно нормального распределения: если $As>0$, то большая часть данных имеет значения, превышающие среднее; если $As<0$, то большая часть данных имеет значения, меньшие среднего.

Эксцесс оценивает «крутизну», т.е. величину большего или меньшего подъема максимума распределения экспериментальных данных по сравнению с максимумом нормального распределения. Если $Ek>0$, то максимум экспериментального распределения выше нормального; если $Ek<0$, то максимум экспериментального распределения ниже нормального.

5. Определить, является ли совокупность однородной. Необходимо вычислить коэффициент вариации. Выборка считается однородной, если коэффициент вариации не превышает 33%. В нашем примере значение коэффициента вариации равно 0,41102 %. Следовательно, выборка является однородной по признаку x . Смотрите также анализ результатов вычисления (с.43).

6. Провести непараметрический тест χ -квадрат-Пирсона (рис. 2.11).

Одной из частых задач научного исследования является определение соответствия (согласия или различия) эмпирического и теоретического распределений или нескольких эмпирических распределений. В современной статистике для этого широко используется критерий χ^2 (хи-квадрат), предложенный Пирсоном.

В Excel критерий хи-квадрат реализован в функции ХИ2ТЕСТ (фактический_интервал; ожидаемый_интервал), аргументами которой являются диапазон экспериментальных частот и диапазон теоретических частот для соответствующих интервалов. Функция ХИ2ТЕСТ вычисляет вероятность совпадения наблюдаемых (фактических) значений и теоретических (гипотетических) значений. Если вычисленная вероятность ниже уровня значимости (0,05), то нулевая гипотеза отвергается и

утверждается, что наблюдаемые значения не соответствуют нормальному закону распределения. Если вычисленная вероятность близка к 1, то можно говорить о высокой степени соответствия экспериментальных данных нормальному закону распределения.

Теоретические частоты вычисляются при помощи функции НОРМРАСП (x ; среднее; станд_откл; интегральная). Здесь среднее – математическое ожидание теоретического распределения, в данном случае совпадает с выборочным средним; станд_откл – стандартное отклонение распределения, в данном случае берется оценка по выборочным данным; интегральная – логическое значение, следует поставить 1 чтобы получить интегральную функцию распределения. Для получения вероятности попадания гипотетического значения из нормально распределенной совокупности в интервал $[x_1; x_2]$, следует вычислить разность между значением функции при $x=x_1$ и $x=x_2$. Для получения теоретических частот надо умножить вероятности на объем выборки.

Нулевая гипотеза: значения случайных величин полученной выборки подчиняются нормальному закону.

- Необходимо провести анализ нулевой гипотезы соответствия конкретной выборки непрерывных значений случайных величин нормальному закону распределения. Вычислите точечные оценки распределения (то есть, произвести расчет элементарных статистических характеристик непрерывных значений случайной величины).

Функции	Пакет анализа данных
Выборочное среднее значение при помощи функции СРЗНАЧ; оценку генерального среднего квадратического отклонения (стандартного отклонения), используя функцию СТАНДОТКЛОН.В(X)	Описательная статистика
См. Таблицы 2.3 и 2.4, Рисунок 2.11	

- Построить группированный статистический (непрерывный вариационный или равноинтервальный) ряд.

1 способ (F2:G10)

Примечание. Интервальный ряд может строиться как с интервалами равной ширины (равноинтервальный ряд) так и с неодинаковыми интервалами, если это диктуется условиями статистического исследования.

Если цель исследования не определяет способ построения интервального ряда, то строится равноинтервальный ряд, число интервалов в котором определяется по формуле Стерджесса: $k=1+3,322\lg(n)$.

где k – число интервалов, n – объем выборки. (Формула обычно дает число дробное, а в качестве числа интервалов выбирается ближайшее целое к полученному число.) Ширина интервала в таком случае определяется по формуле:

$$d = \frac{x_{max} - x_{min}}{k} \quad (17)$$

Графически вариационные ряды могут быть представлены в виде гистограммы (над каждым интервалом интервального ряда выстраивается «столбик» высоты, соответствующей частоте в этом интервале), полигона распределения (ломаная линия, соединяющая точки $(x_i; m_i)$ либо кумуляты (строится по накопленным частотам, т.е. для каждого значения признака берется частота появления в совокупности объектов со значением признака, меньшим данного).

- Определить количество интервалов по формуле Стерджесса: $=1+3,322*\log_{10}(100)$.
- Например, значение, вычисленное по формуле Стерджесса, приблизительно равно 7,64. Значит, количество интервалов можно взять равным $k=8$.
- Определить длину каждого интервала: $=(30-5)/8$
- Получится значение, равное 3,125. Следовательно, ближайшее целое к полученному – это число 3; $d=3$.
- Подготовить таблицу для построения вариационного ряда, введя названия для столбца интервалов (значений вариантов) и столбца частот (абсолютных частот).
 - Задать массив интервалов, указывая для каждого из 8 интервалов верхнюю границу. Для этого в соответствующей ячейке вычислить верхнюю границу первого интервала, введя формулу $=D12+D18$ ($=$ (минимальное в выборке) + (ширина интервала)); в следующей ячейке верхнюю границу второго интервала, введя формулу $=$ (верхняя граница первого интервала)+(ширина интервала). Для вычисления оставшихся значений верхних границ интервалов зафиксируем номер ячейки (в которой содержится ширина интервала) во введенной формуле при помощи знака \$ перед (F4): $=F3+D$18. Скопировать содержимое ячейки (содержит верхнюю границу второго интервала) в диапазон из шести ячеек. Значение верхней границы последнего интервала равно вычисленному ранее максимальному значению в выборке.$
 - Вычислить абсолютные частоты: выделить столбец частот, ввести формулу $=\text{ЧАСТОТА}(\text{массив значений выборки}; \text{массив интервалов})$ и нажать сочетание клавиш CTRL+SHIFT+ENTER.

Примечание. В дальнейшем фраза «введите формулу массива» предполагает выполнение четырех действий: 1) выделить заполняемый диапазон ячеек; 2) нажать на клавиатуре кнопку F2; 3) ввести формулу $=\text{ЧАСТОТА}(A1:A100;F3:F10)$; 4) нажать комбинацию клавиш Ctrl+Shift+Enter.

Важно, чтобы каждый элемент выборки был отнесен к одному и только к одному интервалу, а если значение элемента попадает на границу интервала, то будем относить его к интервалу с младшим номером. Минимальный элемент всегда относим к первому интервалу, максимальный к последнему. Убедитесь, что сумма всех частот наблюдения (абсолютных частот) равна объему выборки ($n=100$).

2 способ (F14:H23)

- Построить вариационный ряд можно, используя инструмент Гистограмма из Пакета анализа.

В настройке MS Excel инструмент Гистограмма Пакета анализа данных используется для генерации интервального вариационного ряда с равными по величине интервалами, а также для построения гистограммы и кумуляты сформированного ряда распределения.

Инструмент Гистограмма производит следующие действия:

- Рассчитывает число интервалов по формуле Стерджесса
- Определяет ширину интервала d по формуле $d=R/(k-1)$, где R – размах вариации в выборке и k - количество интервалов
- Определяет нижние границы интервалов
- Формирует интервальный вариационный ряд в соответствии с величинами d и k
- Рассчитывает частоты и накопленные частоты интервалов, определяя число попаданий данных в сформированные интервалы

- Строит столбиковую диаграмму частот (которая может быть преобразована в гистограмму) и кумуляту накопленных частот для полученного ряда распределения
- Генерирует для вариационного ряда выходную таблицу.

Между терминологией, генерируемой в режиме Гистограмма выходной таблицы, и терминами, принятыми для вариационных рядов, имеются расхождения.

Таблица 2.6 - Статистическая интерпретация терминологии инструмента Гистограмма

Термин инструмента Гистограмма	Термин, принятый в статистике
Карманы	Интервалы вариационного ряда
Интервал карманов	Диапазон ячеек, содержащий в возрастающем порядке верхние границы интервалов
Интервальный процент	Накопленная частота, выраженная в процентах

Инструмент Гистограмма имеет два режима работы:

- Режим автоматического формирования интервалов вариационного ряда, имеющих равную величину d
- Режим формирования интервального ряда в соответствии с границами, заданными пользователем. Если при этом заданные интервалы будут не равны между собой, то в сгенерированной столбиковой диаграмме частоты попадания в интервал не будут связаны с размером интервала, что не позволит правильно оценить характер распределения значений изучаемой выборки.

Запуск инструмента Гистограмма осуществляется аналогично инструменту Описательная статистика надстройки Пакета анализа данных.

В появившемся диалоговом окне инструмента Гистограмма задаются следующие параметры:

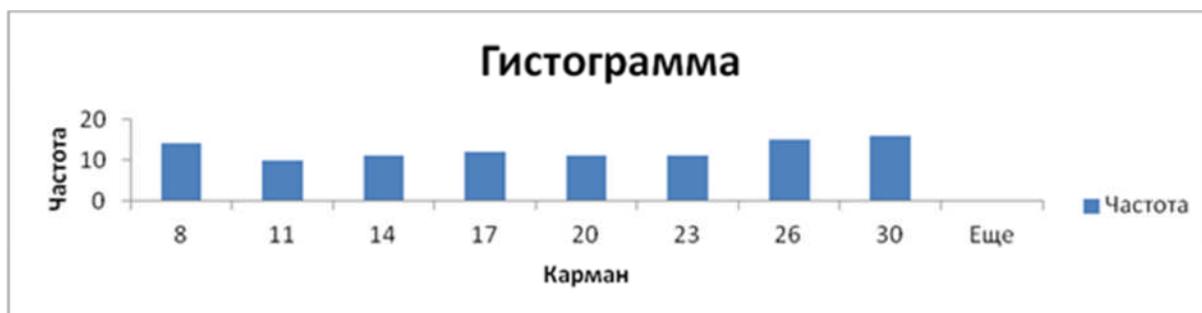


Рисунок 2.9 - Столбиковая диаграмма (пример 1.3)

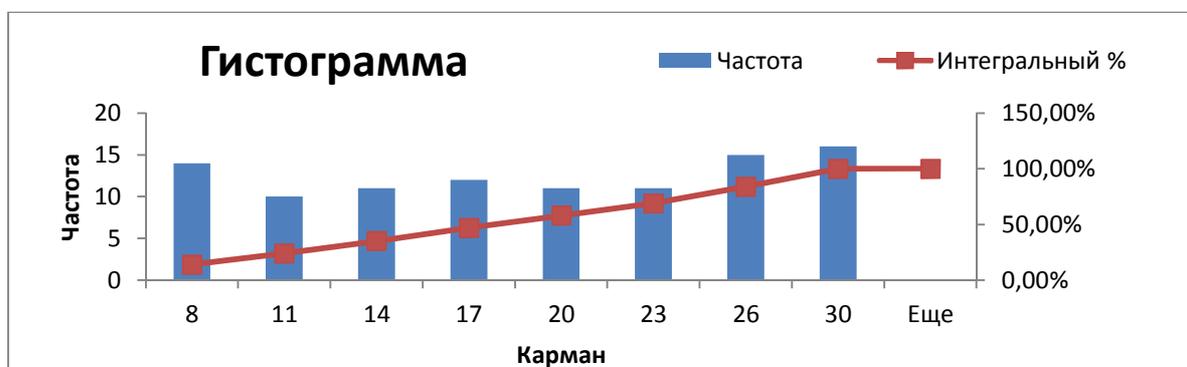


Рисунок 2.10 - Столбиковая диаграмма и график кумуляты (пример 1.3)

Входной интервал – вводится ссылка на диапазон ячеек, содержащих значения анализируемого признака.

Интервал карманов (необязательный параметр) – вводится ссылка на диапазон ячеек, в которых задаются верхние границы интервалов (ВВЕСТИ диапазон ячеек F3:F10). Если такой диапазон не указан, MS Excel осуществляет расчет нижних границ автоматически.

Метки – флажок не активизируется.

Выходной интервал – вводится ссылка на ячейку заголовка первого столбца формируемой таблицы интервального вариационного ряда.

Новый рабочий лист/Новая рабочая книга – переключатель открывает новый рабочий лист/новую рабочую книгу.

Парето (отсортированная гистограмма) флажок устанавливается в активное состояние при необходимости представить данные в порядке убывания частоты. Если флажок снят, то данные в выходном диапазоне будут приведены в порядке следования интервалов.

Интегральный процент – флажок устанавливается в активное состояние, если необходимо рассчитать накопленные частоты (выраженные в процентах) и построить график кумуляты.

Вывод графика – флажок устанавливается в активное состояние при необходимости автоматического построения столбиковой диаграммы.

	A	B	C	D	E	F	G	H	I	J
1	26	<i>Столбец1</i>				Интервалы, хправ	Абсолютные частоты	Вероятности (попадания в интервал)		
2	19			Функции	№ инт.	Карманы	Частоты наблюдений	Теоретические частоты	Теоретические частоты	
3	11	Среднее	18,03	18,03	1	8	14	0,048605175	4,860517549	
4	15	Стандартная ошибка	0,741068707		2	11	10	0,08344858	8,344857999	
5	13	Медиана	19		3	14	11	0,121881963	12,18819634	
6	22	Мода	15		4	17	12	0,151443254	15,14432535	
7	19	Стандартное отклонение	7,410687069	7,41069	5	20	11	0,160086184	16,00861844	
8	19	Дисперсия выборки	54,91828283		6	23	11	0,143963398	14,3963398	
9	15	Эксцесс	-1,230777562		7	26	15	0,110139463	11,01394629	
10	21	Асимметричность	-0,153634571		8	30	16	0,087951398	8,795139815	
11	26	Интервал	25	25		Сумма	100	0,907519416	90,75194159	
12	24	Минимум	5	5		вероятность совпадения экспериментальных и теоретических данных				
13	18	Максимум	30	30					0,000220313	p-значение
14	12	Сумма	1803			Карман	Частота	Интегральный %	0,05	Уровень значимости
15	27	Счет	100	100		8	14	14,00%		
16	12	Уровень надежности(95,0%)	1,47044109			11	10	24,00%	STATISTICA	
17	25	Количество интервалов	7,64	8		14	11	35,00%	0,00012	p-значение
18	5	Ширина интервала	3,125	3		17	12	47,00%	Eviews	
19	14					20	11	58,00%	0,035432	p-значение
20	8					23	11	69,00%		
21	20					26	15	84,00%		
22	13					30	16	100,00%		
23	27					Еще	0	100,00%		

Рисунок 2.11 - Результаты вычисления (критерий согласия хи-квадрат)

- Вычислить теоретические вероятности (частоты) попадания нормально распределенной величины в карманы построенного интервального ряда. Для первого интервала вычислим разность значений функции нормального распределения в верхней границе интервала, указанной в ячейке F3, и в нижней, равной минимальному значению в

выборке и указанному в ячейке С12. Таким образом, в ячейке Н3 надо ввести формулу: =НОРМРАСП(F3;С3;С7;1)-НОРМРАСП(С12;С3;С7;1).

Аналогично, для следующего интервала в ячейке Н4 введем формулу: =НОРМРАСП(F4;\$С\$3;\$С\$7;1)-НОРМРАСП(F3;\$С\$3;\$С\$7;1).

Затем размножить ее на диапазон Н4:Н10.

Или функция НОРМ.РАСП().

- Вычислить теоретические частоты: в ячейку I3 записать формулу =100*Н3 (100 – количество наблюдений).

Вычислить значение функции ХИ2ТЕСТ (вероятность соответствия экспериментальных данных): в ячейку I13 ввести функцию ХИ2ТЕСТ.

Или функция ХИ2.ТЕСТ().

Параметры функции указаны на рисунке ниже.

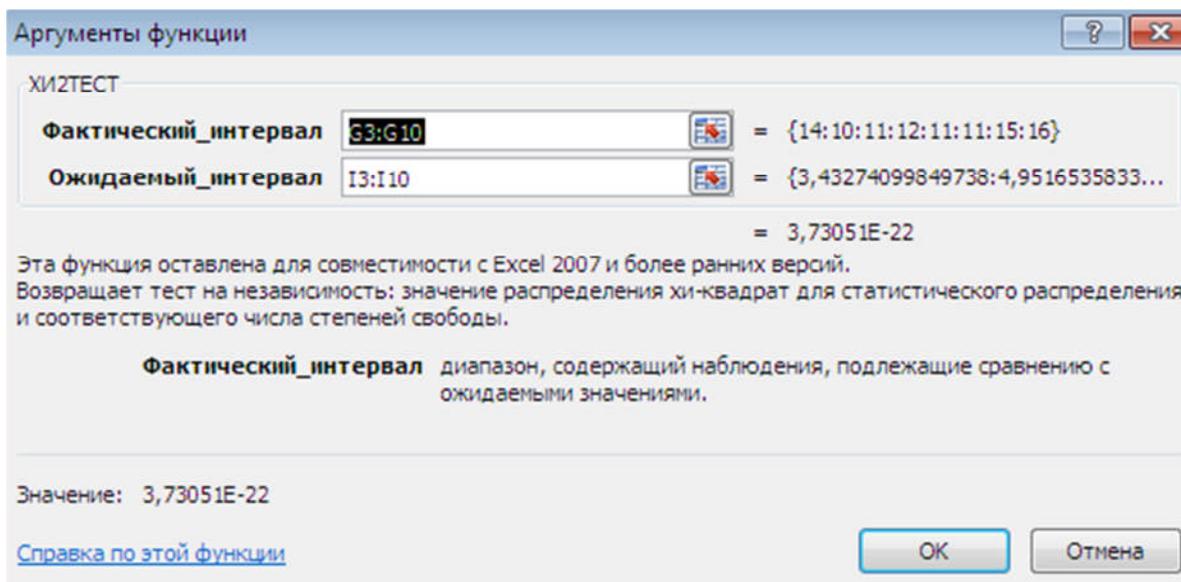


Рисунок 2.12 - ХИ2ТЕСТ()

- Произвести анализ нулевой гипотезы посредством непараметрического теста хи-квадрат-Пирсона.

Например, полученное значение 0,000220313 означает, что основания отвергнуть гипотезу о нормальности распределения есть, поскольку значение меньше уровня значимости 0,05.

Источники: www.pdfactory.com

Смирнов, В.А. Прикладная статистика в пакете анализа MS Excel [текст]: учебное пособие / В.А. Смирнов. – Пенза: ПГУАС, 2008. - 88 с.

<http://piter-melnikov.narod.ru/part3/7.3.3.2.htm>

<http://pmvt.ru/archive/proverka-gipotez/view.html>

http://www.bntu.by/images/stories/ftk/Kaf/VM_3/stat_1.pdf

2.2.9 Технология вычислений в EViews

Пример 1.1 Имеются наблюдения: темпы роста ВВП Беларуси (%), представленные в таблице 2.1. Произвести проверку соответствия выборки нормальному закону распределения при уровне значимости 5%. Произвести расчет и анализ описательной статистики. Определить, является ли выборка количественно однородной по данному признаку x .

1. Необходимо создать файл (Workfile) со своим именем и импортировать исходные данные из файла. Смотрите приложения Ж, З, И.
2. Необходимо построить график временного ряда (представить данные в графическом виде):
 - Открыть серию: двойным щелчком открываем серию значений x .

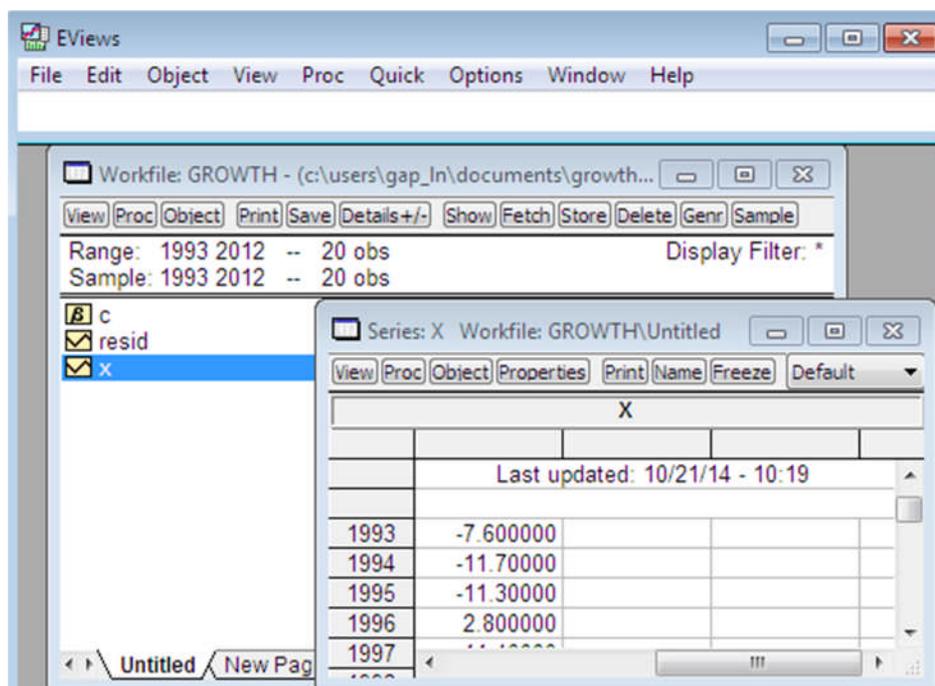


Рисунок 2.13 - Числовые значения ряда (EViews)

Growth - рост, прирост, развитие, возрастание, увеличение.

- Построить график временного ряда (View\Graph\Line).

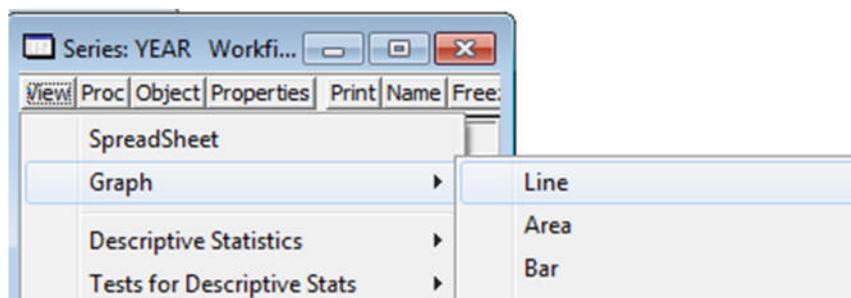


Рисунок 2.14 - Действия по построению графика

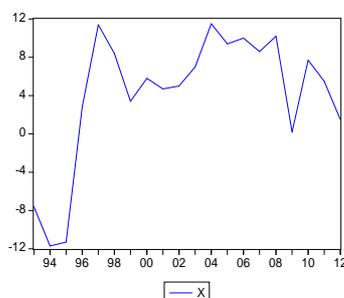


Рисунок 2.15 - Графическое представление значений ряда x (EViews)

3. Необходимо произвести анализ графика наблюдений. Указать отрезки времени, где возрастает и убывает показатель (темпы роста ВВП Беларуси, %).
4. Необходимо вычислить статистику Жака-Бера (критерий согласия Жака-Бера):

1 способ (статистика Жака-Бера)

- Открыть ряд (серию) x .
- Вычислить статистику Жака-Бера для тестирования ряда. (View\Descriptive\Histogram and Stats). Появится рисунок, отображающий результат.

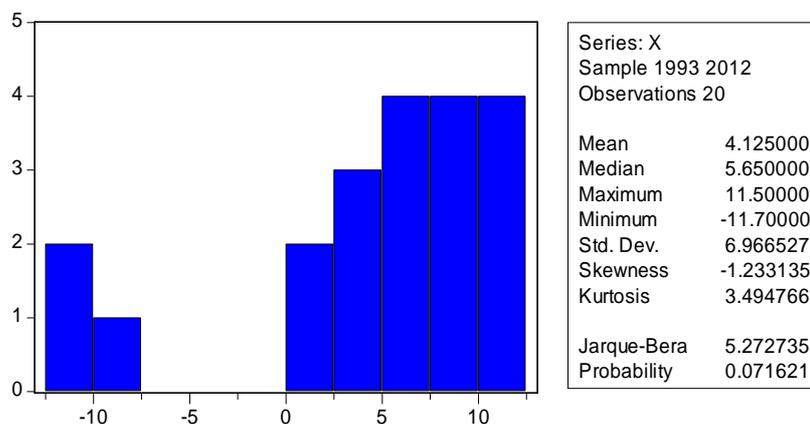


Рисунок 2.16 - Гистограмма и описательная статистика (EViews)

5. Сделать вывод о принятии или отклонении нулевой гипотезы по эквивалентной форме критерия 3 (смотри с. 48).

2 способ

Описательная статистика

- Для возврата к отображению данных в виде таблицы значений выберите закладку View\SpreadSheet.

Таблица 2.6 - Характеристики нормального закона распределения

<p>асимметрии ($=0$)</p> <p>Skewness</p>	<p>характеризует степень смещения вариационного ряда относительно среднего значения по величине и направлению.</p>	<p>В симметричной кривой коэффициент асимметрии равен нулю. Если правая ветвь кривой, начиная от вершины больше левой (<i>правосторонняя асимметрия</i>), то коэффициент асимметрии больше нуля. Если левая ветвь кривой больше правой (<i>левосторонняя асимметрия</i>), то коэффициент асимметрии меньше нуля. Асимметрия менее 0,5 считается малой.</p>
<p>эксцесса ($=3$)</p> <p>Kurtosis</p>	<p>характеризует крутизну кривой распределения - ее заостренность или пологость по сравнению с нормальной кривой.</p>	<p>Коэффициент эксцесса нормального распределения равен 3. Коэффициент островершинного распределения превышает 3, а плосковершинного – меньше 3.</p>

Таблица 2.7 - Descriptive Statistics – Stats Table (описательная статистика)

Mean	среднее
Median	медиана
Maximum	максимум
Minimum	минимум
Std.Dev.	Стандартное отклонение (корень из дисперсии)
Skewness	Асимметрия (0 у нормального распределения)
Kurtosis	Экссесс (3 у нормального распределения)
Jarque-Bera	Статистика Жака-Бера (для проверки того, что выборка из нормального распределения)
Probability	Вероятность ошибиться, отвергнув гипотезу о том, что выборка из нормального распределения

б. Определить, является ли совокупность однородной. Необходимо вычислить коэффициент вариации. Совокупность считается однородной, если коэффициент вариации не превышает 33%. В нашем примере, значение коэффициента вариации равно 1,688855%. Смотрите также анализ результатов вычисления (с.46).

2.2.10 Технология вычислений в пакете Statistica

Пример 1.1 Имеются наблюдения: темпы роста ВВП Беларуси (%), представленные в таблице 2.1. Произвести проверку соответствия выборки нормальному закону распределения при уровне значимости 5%. Произвести расчет и анализ описательной статистики. Определить, является ли выборка количественно однородной по данному признаку.

1. Необходимо создать файл, содержащий таблицу наблюдений. Смотрите приложение И. Сохранить таблицу данных.

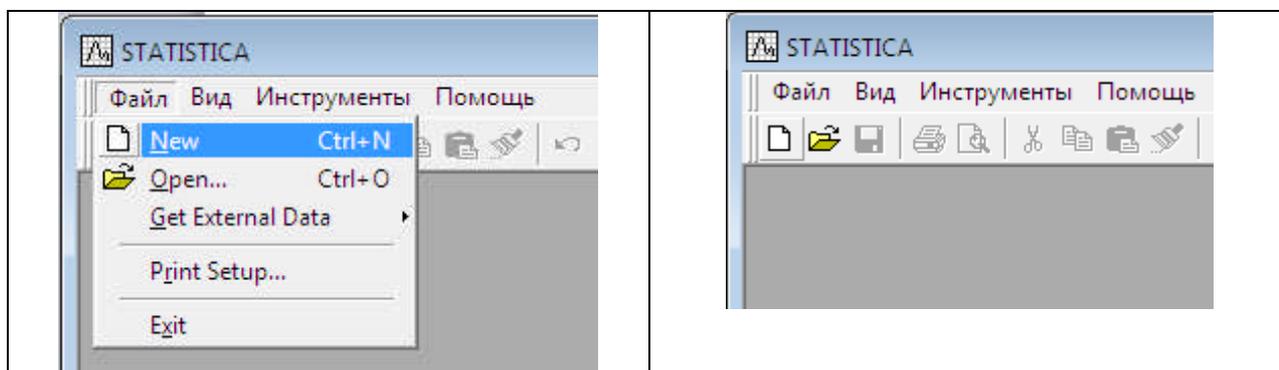


Рисунок 2.17 - Создание нового файла

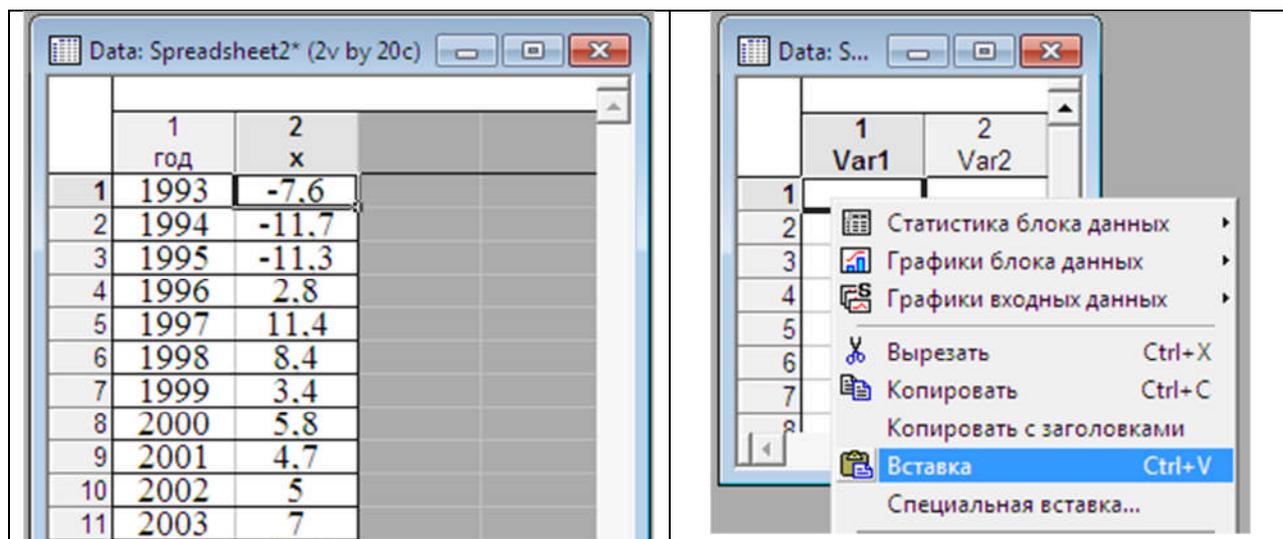


Рисунок 2.18 - Таблица данных

2. Представить данные в графическом виде (построить график линии).
- Графики\2D графики\Графики линии (переменные)\переменные (выбрать)\ОК.

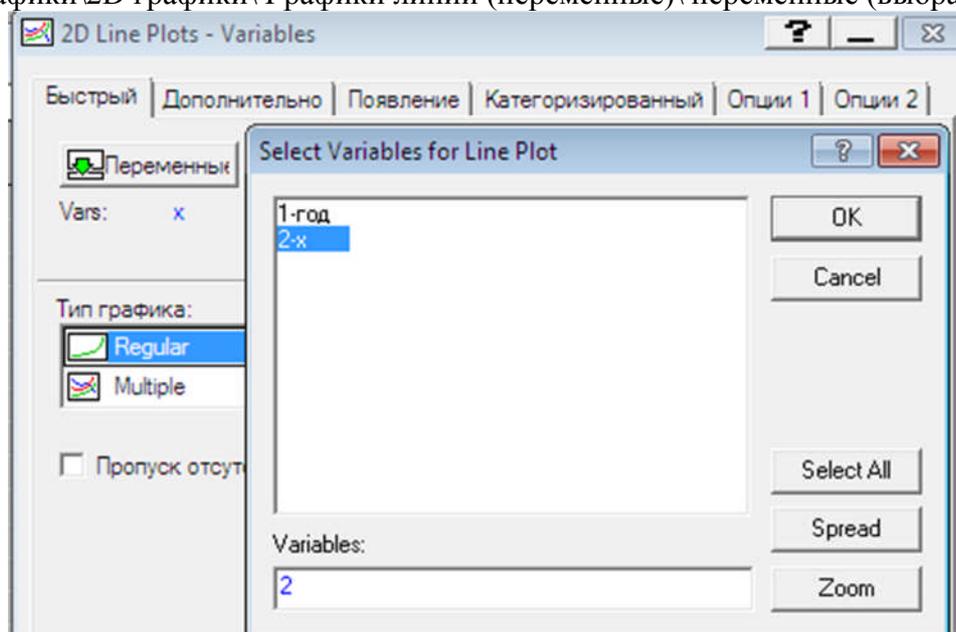


Рисунок 2.19 - Выбор переменных

- Оформить надписи дат (они не соответствуют значениям ряда Год): выделить значения дат в таблице с данными и скопировать в буфер обмена.
- Выделить надписи по оси Oх, вызвать контекстное меню **Properties\All Options...\Plot Point labels**. Извлечь из буфера обмена (вставить) диапазон дат в поле (столбец с пронумерованными строками).

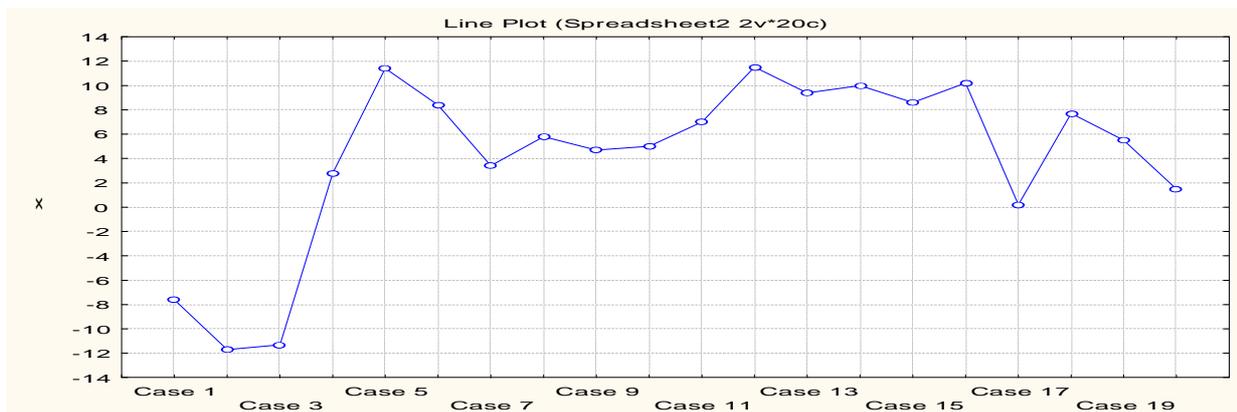


Рисунок 2.20 - График данных (Statistica)

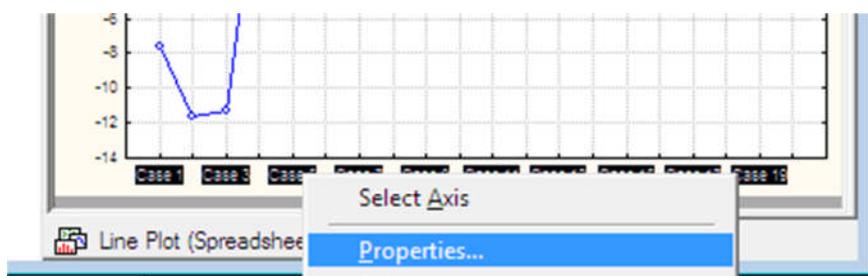


Рисунок 2.21 - Свойства

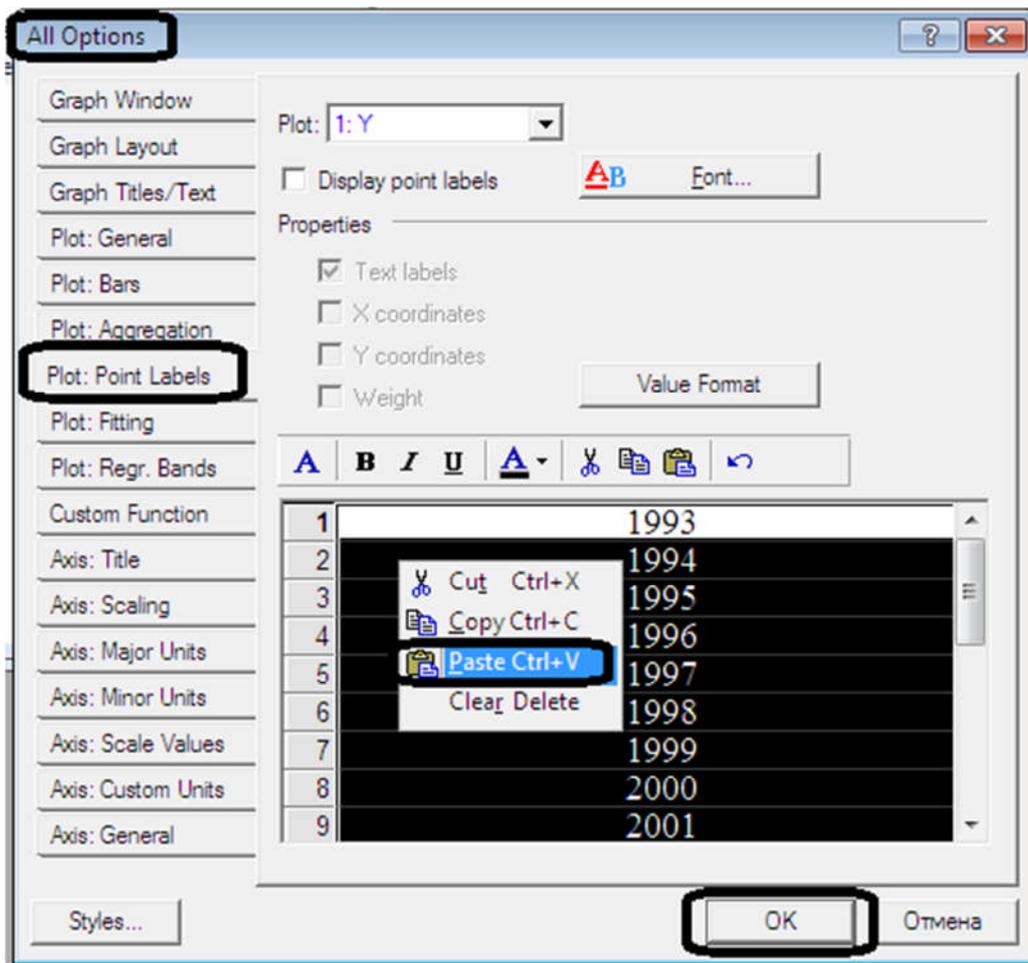


Рисунок 2.22 - Вставка

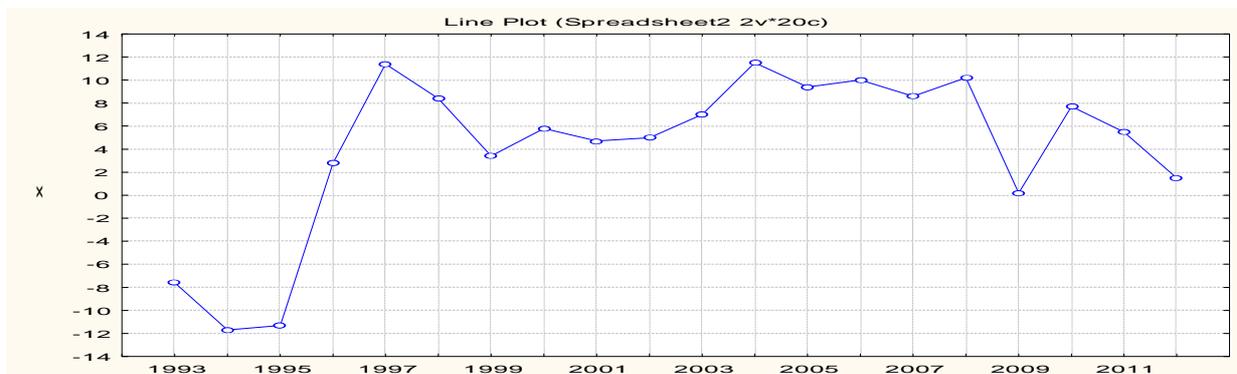


Рисунок 2.23 - Надписи (соответствующие условию примера)

3. Сделать анализ графика. Указать отрезки времени, где возрастает и убывает показатель. Сохранить график.
4. Вычислить описательную статистику.
 - Активизировать окно с данными.
 - Выполнить команду главного меню Статистика\Основная статистика/Таблицы \ в окне Quick нажать *OK*.

- В окне Descriptive statistics:... выбрать закладку Advanced. Следует поставить флажок напротив статистик, которые подлежат вычислению. Отметить: *Valid N, Mean, Median, Mode, Standard Deviation, Variance, Skewness, Std. Skewness, Kurtosis, Std. Kurtosis, Minimum & maximum, Range*.

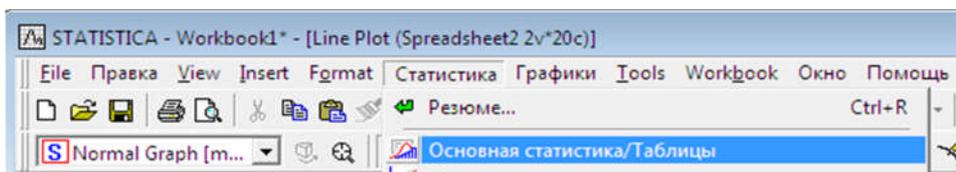


Рисунок 2.24 - Основная статистика

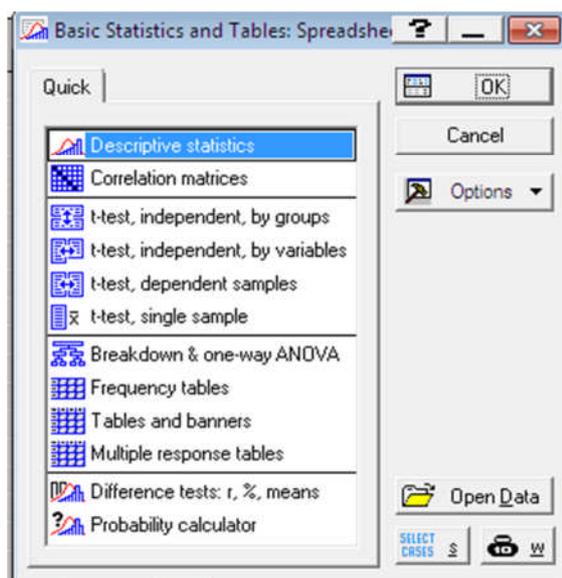


Рисунок 2.25 - Описательная статистика

- Выбрать переменную x с помощью кнопки *Variables*.
- Нажать кнопку *Summary*. Появится таблица с результатами расчетов описательных статистик.

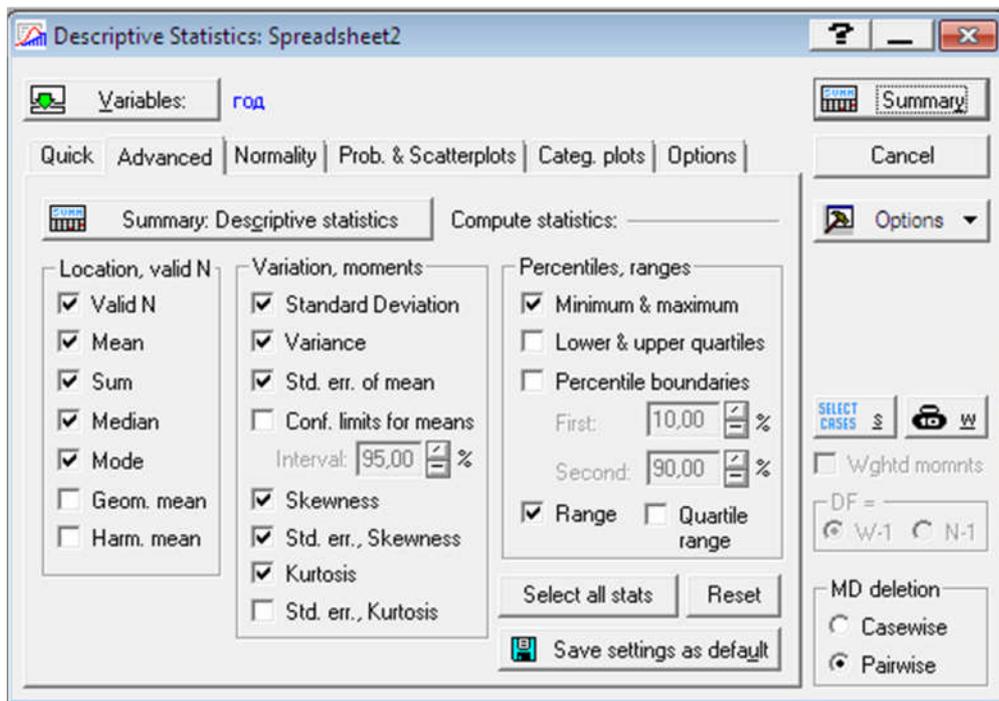
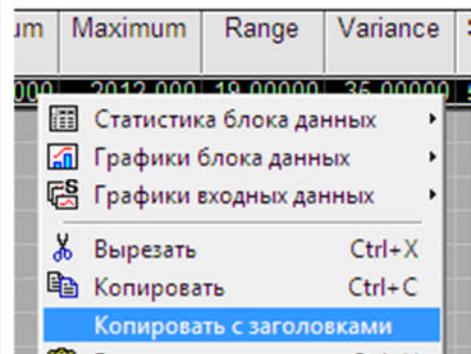


Рисунок 2.26 - Окно Описательная статистика

Таблица 2.8 – Описательная статистика (Statistica)

	x
Valid N	20
Mean	4,125000
Median	5,650000
Mode	Multiple
Sum	82,50000
Minimum	-11,7000
Maximum	11,50000
Range	23,20000
Variance	48,53250
Std.Dev.	6,966527
Standard	1,557763
Skewness	-1,33546
Std.Err.	0,512103
Kurtosis	1,017686
Std.Err.	0,992384



5. Произвести анализ описательной статистики. Смотрите анализ результатов вычисления (с. 46).
6. Определить, является ли выборка количественно однородной по данному признаку x . Необходимо вычислить коэффициент вариации. Совокупность считается однородной, если коэффициент вариации не превышает 33%. В нашем примере, значение коэффициента вариации равно 1,688855%. Смотрите также анализ результатов вычисления (с.46).
7. Построить кривую теоретического закона распределения и гистограммы эмпирического распределения.

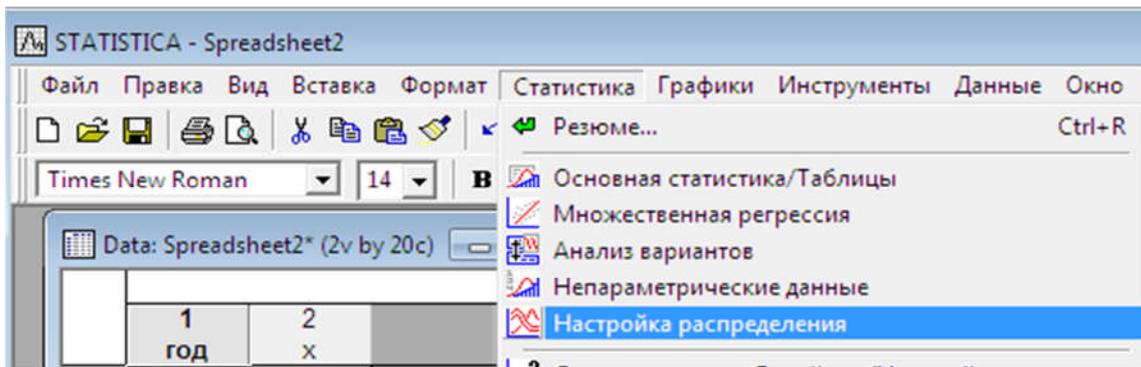


Рисунок 2.27 - Настройка распределения

- Главное меню **Статистика\Настройка распределения**. В открывшемся окне указать предполагаемую природу случайной величины (непрерывная) и предполагаемый нормальный закон распределения: флажок Continuous Distributions (непрерывное распределение, модуль), Normal и нажать ОК.

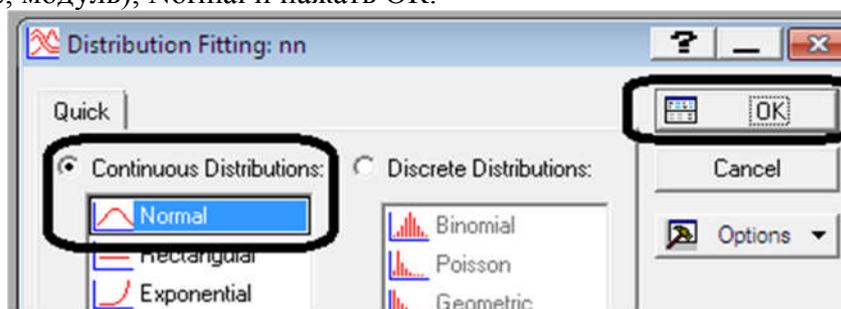


Рисунок 2.28 - Предполагаемый закон распределения (нормальный)

- Во вкладке *Options* (опции) включить расчет теста Колмогоров-Смирнов (по категориям), и снять флажок *Combine Categories* (запрет объединения соседних интервалов при расчете теста χ^2 – Пирсона).

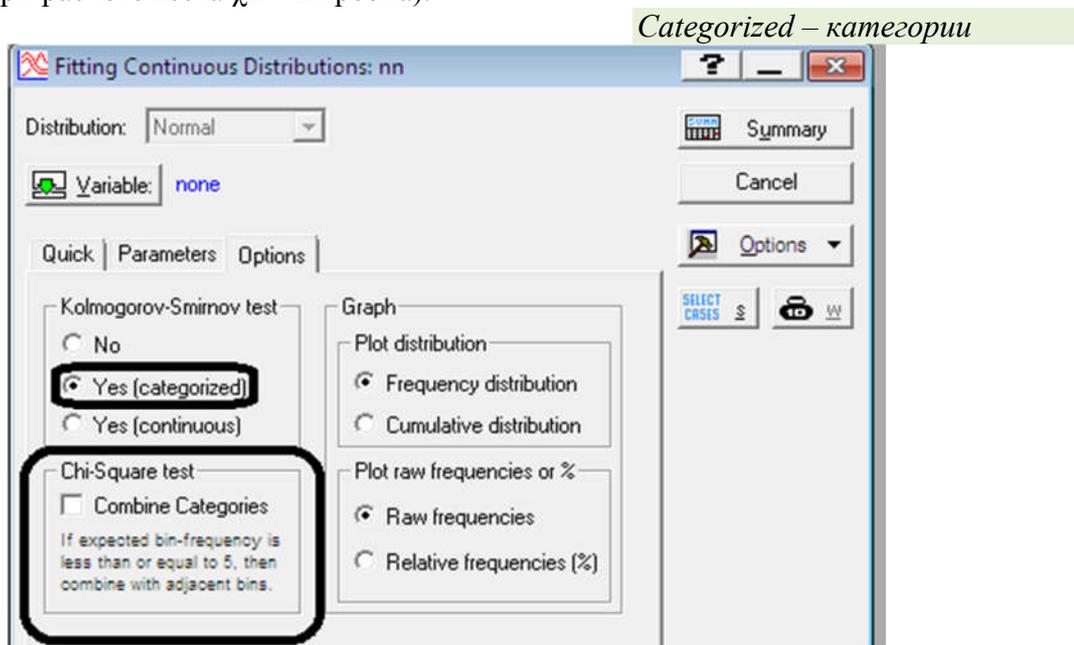


Рисунок 2.29 – Опции

- Построить кривую теоретического закона распределения и гистограммы эмпирического распределения. На вкладке *Quick* нажать кнопки *Plot of Observed and expected distributions*, выбрать вторую переменную, нажать *OK*, снова нажать кнопку *Plot of Observed and expected distributions*.

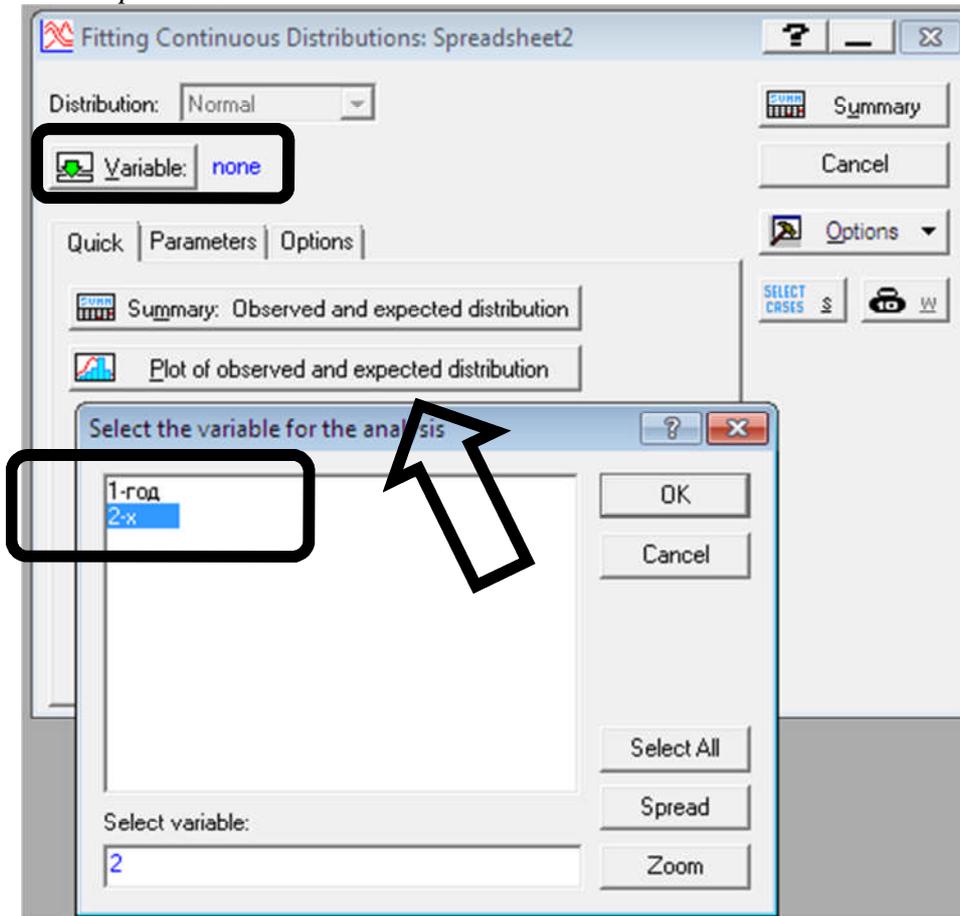


Рисунок 2.30 - Выбор переменных

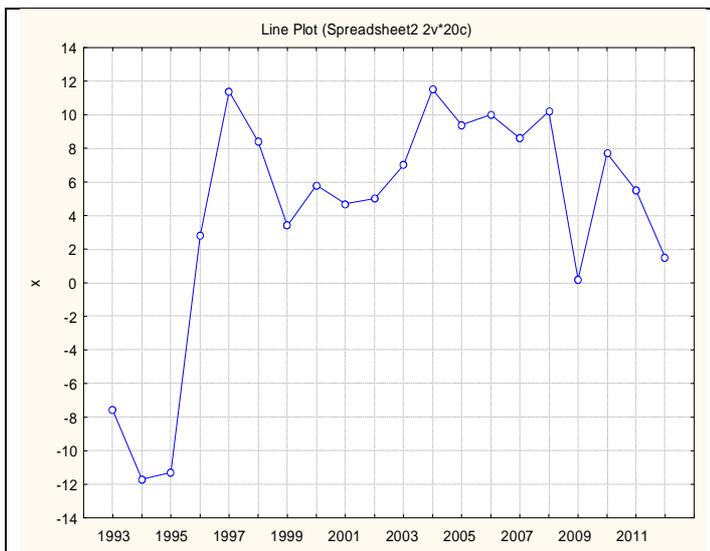
8. Сделать вывод о принятии или отклонении нулевой гипотезы.

2.2.11 Анализ результатов вычисления

Пример 1.1

1. Произвести анализ графика.

Анализ графика значений случайных величин выборки:



- Наблюдается тренд – да/нет. На каких участках наблюдается рост или убывание значений?

- Наличие колебаний – есть/нет. Какие колебания? (Различные, равномерные и т.д.).

Наблюдается резкий рост значений с 1995 по 1997.

Имеются колебания.

2. Определить, является ли выборка количественно однородной по данному признаку. Необходимо вычислить для данного признака следующий показатель: Vx (коэффициент вариации).

Примечание. Коэффициент вариации (%) есть отношение стандартного отклонения к среднему значению (среднему арифметическому), умноженное на 100%.

Среднее квадратическое (стандартное) отклонение дает обобщенную характеристику признака совокупности и показывает во сколько раз в среднем колеблется величина признака совокупности. Среднее квадратическое отклонение является мерой надежности средней величины: чем оно меньше, тем точнее средняя арифметическая.

Сопоставление линейных или среднеквадратических отклонений по признакам совокупности дает возможность определить статистическую однородность совокупности: чем меньше размер, тем совокупность более однородна.

- Произвести расчет коэффициента вариации, (%).

Сравнивая значения средней величины (*Mean*) и стандартного отклонения (*Standard deviation*), определить коэффициент вариации как их отношение ($Vx = 1,68\%$).

	x
Среднее арифметическое (\bar{x}), Mean	4,125000
Стандартное отклонение (s), Std.Dev.	6,966527
Коэффициент вариации (Vx), %	1,688855

Вывод: Для признака x показатель $V_x \leq 33\%$, следовательно, выборка количественно однородная по данному признаку.

3. Сделать выводы о близости изучаемого распределения к нормальному распределению.

- Проверить выполнение условия: $\bar{x} = M_e = M_o$.

Признак	$\bar{x} = M_e = M_o$
x	не выполняется

Mean	Median	Mode
4,125000	5,650000	Multiple

- Проверить выполнение условия: $E_k=0$.

	Эксцесс, Kurtosis (E_k)	>0, островершинной (<0, плосковершинной)	незначительно (значительно) отличается от нуля	Std.Err.
x	1,017686	островершинной	значительно	0,992384

Вывод: Так как для признака x $E_k > 0$, то кривая распределения является более островершинной по сравнению с нормальной кривой. При этом E_k значительно отличается от нуля ($E_k = 1,017686$). Следовательно, по данному признаку форма кривой эмпирического распределения значительно отличается от формы нормального распределения.

- Проверить выполнения условия: $A_s=0$.

Чем больше величина $|A_s|$, тем более асимметрично распределение. Оценочная шкала асимметрии:

- $|A_s| \leq 0,25$ - асимметрия незначительная;
- $0,25 < |A_s| \leq 0,5$ - асимметрия заметная (умеренная);
- $|A_s| > 0,5$ - асимметрия существенная.

	Асимметрия Skewness, A_s	>0, правосторонняя (<0, левосторонняя)	Значительная, заметная, существенная	Std.Err.
x	-1,33546	левосторонняя	существенная	0,512103

Вывод: Для признака x наблюдается существенная левосторонняя асимметрия.

Значения коэффициентов асимметрии (*Skewness*), эксцесса (*Kurtosis*) не превышают двукратных стандартных ошибок (графы *Standard error of skewness*, *Standard error of kurtosis*). Это указывает на отсутствие значимой скошенности и остро-(плоско) вершинности фактического распределения значений признака по сравнению с нормальным распределением.

Примечание. Критерии согласия

Нормальное распределение наиболее часто встречается, но, чтобы использовать соответствующие приемы, относящиеся к нему, необходимо убедиться, что обрабатываемые данные действительно имеют нормальное распределение.

Осуществить проверку гипотезы о соответствии реального распределения значений временных рядов нормальному закону распределения с помощью тестов: χ^2 -Пирсона, Колмогорова-Смирнова, Жака-Бера.

Решение о принятии или отклонении нулевой гипотезы (H_0) осуществляется с помощью нескольких критериев: симметричный двухсторонний (критерий 1), односторонний (правосторонний) (критерий 2), эквивалентной формы критерия (критерий 3).

Критерий 1	<i>Гипотеза H_0</i>	$\left\{ \begin{array}{l} \text{не отклоняется,} \\ \text{отклоняется,} \end{array} \right.$	$\left\{ \begin{array}{l} \text{если } D_n < D_0 \\ \text{если } D_n \geq D_0 \end{array} \right.$
Критерий 2	<i>Гипотеза H_0</i>	$\left\{ \begin{array}{l} \text{не отклоняется,} \\ \text{отклоняется,} \end{array} \right.$	$\left\{ \begin{array}{l} \text{если } D_n < D_0 \\ \text{если } D_n \geq D_0 \end{array} \right.$
Критерий 3	<i>Гипотеза H_0</i>	$\left\{ \begin{array}{l} \text{не отклоняется,} \\ \text{отклоняется,} \end{array} \right.$	$\left\{ \begin{array}{l} \text{если } P > \varepsilon \\ \text{если } P \leq \varepsilon \end{array} \right.$

Примечание. Значение уровня значимости, равное 0,05, означает, что в среднем в 5 случаях из 100 при многократном применении статистического критерия будет отвергаться нулевая гипотеза.

- Используя критерий согласия Колмогорова-Смирнова, выяснить: подчиняются ли нормальному закону распределения значения временного ряда (x) на уровне значимости 5%.

Нулевая гипотеза: Значения временного ряда подчиняются нормальному закону распределения на уровне значимости 5%.

Найти расчетное значение статистики Колмогорова $D_0 = |d|$ ($|d|$ – модуль статистического критерия, вычисляемый по известным эмпирическим данным), критерий 1 для уровня значимости $\varepsilon = 0,05$ по формуле: $D_0 = \frac{\lambda_\varepsilon}{\sqrt{n}}$, где n – количество наблюдений.

Таблица 2.9 - Функция распределения Колмогорова

ε	0,1	0,05	0,02	0,01	0,001
λ_ε	1,224	1,358	1,520	1,627	1,950

Вычислить с помощью приложения MS Excel: формула имеет вид = 1,358/корень(100).

Сравнить *расчетное* $|d|$ (смотреть рисунок, отображающий кривую теоретического закона распределения) и *пороговое (критическое) значение* ($D_n = \Delta = 0.303658$) статистики Колмогорова-Смирнова.

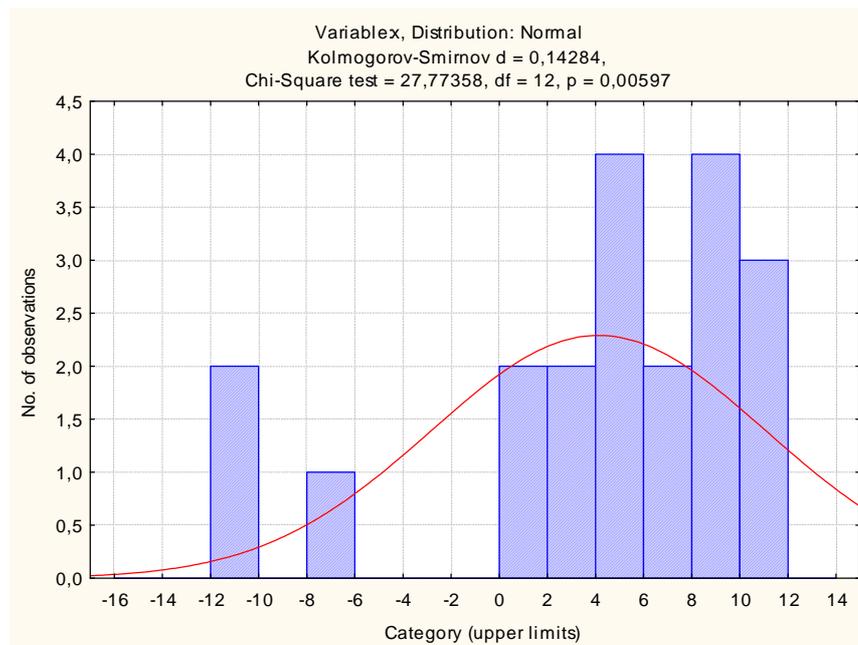


Рисунок 2.31 - Гистограмма (Statistica)

Variable x – переменная x
 Distribution Normal – нормальный закон распределения
 Colmogorov-Smirnov - Тест Колмогорова-Смирнова
 chi-square test - Тест χ^2 – Пирсона
 category upper limits - Категория верхние пределы

Решение о принятии или отклонении H_0 осуществляем по критерию 1.

Сделать выводы. Так как расчетное значение $|d|=|?|$ меньше/больше порогового $\Delta=?$, то нулевая гипотеза не отклоняется/отклоняется. Следовательно, значения временного ряда подчиняются/не подчиняются нормальному закону распределения на уровне значимости 5%.

Выполняя проверку гипотезы о нормальном законе распределения с помощью теста Колмогорова-Смирнова, одновременно осуществили аналогичную проверку по критерию χ^2 – Пирсона.

- Используя критерий согласия χ^2 –Пирсона, выяснить: подчиняются ли нормальному закону распределения значения временного ряда (x) на уровне значимости 5%.

Нулевая гипотеза: Значения временного ряда подчиняются нормальному закону распределения на уровне значимости 5%.

Решение о принятии или отклонении нулевой гипотезы по критерию χ^2 –Пирсона осуществляем по эквивалентной форме критерия 3.

Variable: x , Distribution: Normal
 Chi-Square = 27,77358, df = 12, p = 0,00597

- **Первый параметр** Chi-Square = 23,13450 – значение критерия χ^2
- **Второй параметр** df = 12 – число степеней свободы, определяемое как $df = N - l - 1$, где N – число интервалов, l – количество параметров закона распределения (для нормального закона распределения $l=2$ – это математическое ожидание и среднеквадратическое отклонение)
- **Третий параметр** p = 0,00597 – это статистический уровень значимости (p-уровень)

- Сравнить полученное значение p с пороговым значением $\varepsilon=5\%$.
- Сделать выводы. Так как полученное $p=?$ больше/меньше порогового значения $\varepsilon=5\%$, то нулевая гипотеза не отклоняется/отклоняется. Следовательно, значения временного ряда подчиняются/не подчиняются нормальному закону распределения.
- Используя критерий согласия *Жака-Бера*, выяснить: подчиняются ли нормальному закону распределения значения временного ряда (x) на уровне значимости 5% .
Нулевая гипотеза: Значения временного ряда подчиняются нормальному закону распределения на уровне значимости 5% .

Решение о принятии или отклонении нулевой гипотезы по критерию согласия *Жака-Бера* осуществляем по эквивалентной форме критерия Z .

- Сравнить полученный параметр *Probability* с пороговым значением $\varepsilon=5\%$.
- Сделать выводы. Так как полученное *Probability=?* больше/меньше порогового значения $\varepsilon=5\%$, то нулевая гипотеза не отклоняется/отклоняется. Следовательно, значения временного ряда подчиняются/не подчиняются нормальному закону распределения.

4. Исходя из полученных результатов с помощью тестов: χ^2 -Пирсона, Колмогорова-Смирнова, Жака-Бера, сделать заключение о близости изучаемого распределения к нормальному.

Вывод: выборка количественно однородна по данному признаку и для ее изучения могут использоваться метод наименьших квадратов и вероятностные методы оценки статистических гипотез.

2.3 Задания для самостоятельной работы

Задание 1

Одним и тем же вольтметром было измерено 25 раз напряжение на участке цепи. В результате опытов получены следующие значения напряжения в вольтах: 32, 32, 35, 37, 35, 38, 32, 33, 34, 37, 32, 32, 35, 34, 32, 34, 35, 39, 34, 38, 36, 30, 37, 28, 30. Найдите выборочные среднюю, дисперсию, стандартное отклонение, размах варьирования, моду, медиану. Проверить отклонение от нормального распределения, вычислив асимметрию и эксцесс.

Задание 2

Сгенерировать 500 случайных чисел, распределенных нормально. Построить гистограмму и полный список статистических характеристик с помощью *инструмента Описательная статистика*.

Задание 3

Построить последовательность случайных чисел распределённых по равномерному закону. Для этого нужно повторить все шаги алгоритма с небольшими изменениями. На 4 шаге нужно выбрать **Равномерное**, а на 5 шаге ввести в поле ввода **Между** числа 0 и 1. Выходной интервал начните с ячейки B2.

Задание 4

Имеются данные о количественном составе 60 семей.

2	4	5	6	5	2	3	4	1	4	3	3
4	3	3	4	4	4	4	5	5	3	4	1
3	4	3	5	4	3	5	3	3	2	3	4
6	5	4	4	4	2	3	4	4	6	5	1
5	2	6	2	3	3	4	5	4	4	6	4

Построить вариационный ряд и полигон распределения

Задание 5

Имеются данные о выбросах загрязняющих веществ из 50 источников:

10,4	18,6	10,3	26,0	45,0	18,2	17,3	19,2	25,8	18,7
28,2	25,2	18,4	17,5	41,8	14,6	10,0	37,8	10,5	16,0
18,1	16,8	38,5	37,7	17,9	29,0	10,1	28,0	12,0	14,0
14,2	20,8	13,5	42,4	15,5	17,9	19,	10,8	12,1	12,4
12,9	12,6	16,8	19,7	18,3	36,8	15,0	37,0	13,0	19,5

Составить равноинтервальный ряд, построить гистограмму.

3 Лабораторная работа №2

Тема: Построить диаграмму рассеяния. Составить уравнение регрессии, воспользовавшись методом наименьших квадратов, дать экономическую интерпретацию коэффициентов регрессии. Произвести расчет линейного коэффициента парной корреляции (матрицы парных коэффициентов корреляции), коэффициентов детерминации (обычного и исправленного).

Цель: Освоить основные принципы работы с указанными приложениями при построении уравнения регрессии, расчета линейного коэффициента парной корреляции (матрицы парных коэффициентов корреляции), коэффициентов детерминации (обычного и исправленного).

3.1 Задания

Задание 2.1 Скопировать исходные данные (файлы, которые созданные в Блокноте) из материалов для студентов (\ForStudent\КИС...).

Построить диаграмму рассеяния. Вычислить коэффициент корреляции.

Задание 2.2 Составить уравнение парной регрессии, воспользовавшись методом наименьших квадратов, дать экономическую интерпретацию коэффициентов регрессии. Рассчитать линейный коэффициент парной корреляции, коэффициенты детерминации (обычного и исправленного).

Рассмотрим примеры.

Пример 2.1 Исследовать зависимость объема экспорта (у, усл. ед.) от валового внутреннего продукта (ВВП) (х, усл. ед.), используя данные 20 наблюдений, приведенные в таблице 1. Построить регрессионную модель $y = f(x) + \varepsilon$. Дать экономическую интерпретацию коэффициентов регрессии.

Рассчитать линейный коэффициент парной корреляции, коэффициенты детерминации (обычного и исправленного).

Таблица 3.1 – Исходные данные для построения модели

ВВП	1000	1090	1150	1230	1300	1360	1400	1470	1500	1580
Экспорт	190	220	240	240	260	250	280	290	310	350

ВВП	1600	1630	1700	1780	1800	1850	1910	1990	2010	2100
Экспорт	340	360	380	400	420	400	400	440	450	470

Пример 2.2. Исследовать зависимость заработной платы (y , тыс. р.) от возраста (x_1 , лет) и стажа по данной специальности (x_2 , лет), используя данные наблюдений, приведенные в таблице 3.2. Построить регрессионную модель $y = f(x_1, x_2) + \varepsilon$.

Рассчитать коэффициенты детерминации (обычного и исправленного) и построить матрицу парных коэффициентов корреляции.

Таблица 3.2 – Данные наблюдений

Заработная плата	Возраст	Стаж
684,48	28	10
674,45	26	8
729,62	30	14
748,86	35	15
761,45	42	16
775,51	40	16
628,81	27	3
731,84	35	13
698,81	30	10
645,64	29	7
711,18	33	11
798,07	40	20
832,83	40	24
607,61	23	2
728,62	36	12
667,14	30	6
683,71	31	10

3.2 Методические указания

3.2.1 Парная регрессия

Суть регрессионного анализа: построение математической модели и определение ее статистической надежности.

Парная (простая) линейная регрессия представляет собой модель, где среднее значение зависимой (объясняемой) переменной рассматривается как функция одной переменной (объясняющей) переменной.

Уравнение парной (простой) линейной регрессии выглядит следующим образом:

$$y = b_1x + b_0 + \varepsilon \quad (18)$$

где ε (e) – отклонение (ошибка, возмущение) от теоретически предполагаемого значения; b_1 и b_0 - неизвестные параметры (коэффициенты регрессии).

Переменную x называют фактором (входной переменной, независимой переменной, признаком-фактором, экзогенной переменной), переменную y называют откликом (реакцией, зависимой переменной, результативным признаком, эндогенной переменной).

При помощи этого уравнения переменная y выражается через константу b_0 и угол наклона прямой (или угловой коэффициент) b_1 , умноженный на значение переменной x . Константу b_0 также называют свободным членом, а угловой коэффициент b_1 - коэффициентом регрессии. Параметры уравнения могут быть определены с помощью метода наименьших квадратов (МНК).

Расчетные формулы для нахождения коэффициентов уравнения регрессии:

$$b_1 = a = \frac{n(\sum y_i x_i) - \sum y_i \sum x_i}{n(\sum x_i^2) - (\sum x_i)^2} \quad (19)$$

$$b_0 = b = \frac{1}{n} \left(\sum y_i - a \sum x_i \right) \quad (20)$$

где n – количество пар исходных данных.

Уравнение парной регрессии можно записать иначе:

$$y = \hat{y} + \varepsilon, \quad (21)$$

где y – фактическое значение результативного признака,

\hat{y} – теоретическое значение результативного признака, найденное по формуле

$$\hat{y} = b_1 x + b_0,$$

ε – случайная величина, характеризующая отклонения реального значения результирующего признака от теоретического (рассчитанного) значения. Знак « \wedge » означает, что между переменными x и y нет строгой функциональной зависимости.

3.2.2 Множественная регрессия

Вид множественной линейной модели регрессионного анализа:

$$y = b_0 + b_1 x_{i1} + \dots + b_j x_{ij} + \dots + b_k x_{jk} + e_i \quad (22)$$

где e_i – случайные ошибки наблюдения, независимые между собой; имеют нулевую среднюю и дисперсию s .

Назначение множественной регрессии: анализ связи между несколькими независимыми переменными и зависимой переменной.

Экономический смысл параметров множественной регрессии: коэффициент множественной регрессии b_j показывает, на какую величину в среднем изменится результативный признак Y , если переменную X_j увеличить на единицу измерения; т. е. является нормативным коэффициентом.

Матричная запись множественной линейной модели регрессионного анализа:

$$Y = bX + e \quad (23)$$

где Y – случайный вектор-столбец размерности $(n \times 1)$ наблюдаемых значений результативного признака (y_1, y_2, \dots, y_n) ;

X – матрица размерности $[n \times (k+1)]$ наблюдаемых значений аргументов;

b – вектор-столбец размерности $[(k+1) \times 1]$ неизвестных, подлежащих оценке параметров (коэффициентов регрессии) модели;

e – случайный вектор-столбец размерности $(n \times 1)$ ошибок наблюдений (остатков).

На практике рекомендуется, чтобы n превышало k не менее чем в три раза.

3.2.3 Корреляционный анализ

Совокупность методов оценки корреляционных характеристик и проверка статистических гипотез о них по выборочным данным называется **корреляционным анализом**. В корреляционном анализе используются следующие основные приемы:

- 1) построение корреляционного поля (диаграммы рассеяния) для двух экономических показателей или двумерных сечений;
- 2) определение выборочных коэффициентов корреляции или составление корреляционных матриц;
- 3) проверка статистических гипотез о значимости связи между показателями.

Коэффициент корреляции является мерой линейной зависимости двух величин. Чем больше коэффициент корреляции по модулю, тем сильнее линейная зависимость. Значение коэффициента корреляции лежит в интервале $[-1; 1]$. Ориентировочно определить значение коэффициента корреляции можно, анализируя диаграмму рассеяния. Чем теснее расположены точки относительно некоторой прямой (существует линейная зависимость), тем больше по абсолютной величине коэффициент корреляции стремится к единице, и наоборот, чем более расплывчата диаграмма рассеяния, тем ближе он к нулю.

Анализ парных взаимосвязей

Связь как *синхронность* (согласованность) – корреляционный анализ.

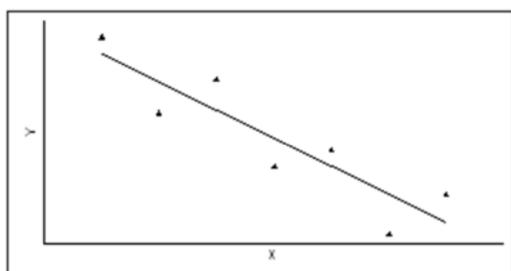
Связь как *зависимость* (влияние) – регрессионный анализ (причинно-следственные связи).

В регрессионном анализе один из признаков зависит от другого. Первый (зависимый) признак называется в регрессионном анализе *результатирующим*, второй (независимый) – *факторным*. Не всегда можно однозначно определить, какой из признаков является независимым, а какой – зависимым. Часто связь может рассматриваться как двунаправленная.

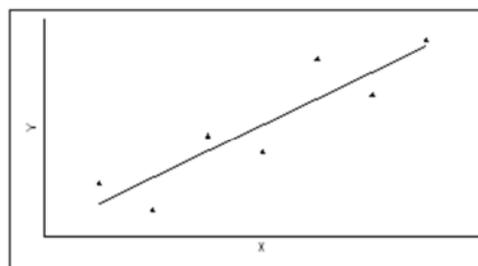
Этапы анализа

- Выявление наличия взаимосвязи между признаками;
- Определение формы связи;
- Определение силы (тесноты) и направления связи.

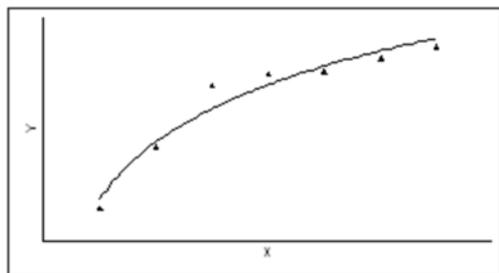
Для определения формы связи необходимо произвести визуальный анализ диаграммы рассеяния. *Существуют следующие формы связи: нелинейная и линейная.* Поскольку наиболее простой формой зависимости в математике является прямая, то в корреляционном и регрессионном анализе наиболее популярны *линейные модели*.



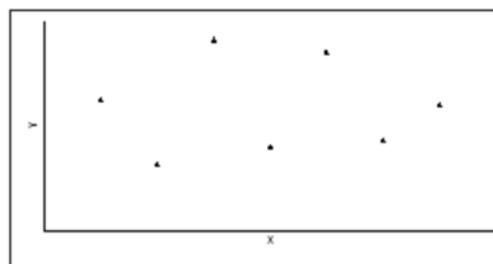
Форма связи – линейная (прямая), направление связи - обратное (наблюдается убыв.)



Форма связи – линейная (прямая), направление связи – прямое (наблюдается возр.)



Форма – нелинейная (кривая), направление – прямое



Отсутствие связи между признаками

Рисунок 3.1 – Формы и направления связи между признаками

3.2.4 Задачи регрессионного анализа

Основная задача регрессионного анализа заключается в нахождении по выборке объемом n оценки неизвестных коэффициентов регрессии b_0, b_1, \dots, b_k . Задачи регрессионного анализа состоят в том, чтобы по имеющимся статистическим данным для переменных X_i и Y :

- получить наилучшие оценки неизвестных параметров b_0, b_1, \dots, b_k ;
- проверить статистические гипотезы о параметрах модели;
- проверить, достаточно ли хорошо модель согласуется со статистическими данными (адекватность модели данным наблюдений).

Построение моделей множественной регрессии состоит из следующих этапов: выбор формы связи (уравнения регрессии); определение параметров выбранного уравнения; анализ качества уравнения и проверка адекватности уравнения эмпирическим данным, совершенствование уравнения.

3.2.5 Оценка параметров уравнения регрессии

Построение уравнения регрессии сводится к оценке ее параметров. Для оценки параметров регрессий, линейных по параметрам, используют метод наименьших квадратов (МНК). МНК позволяет получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака y от теоретических \widehat{y}_x минимальна, т.е.

$$\sum (y - \widehat{y}_x)^2 \rightarrow \min \quad (24)$$

Для линейных и нелинейных уравнений, приводимых к линейным, решается следующая система относительно a и b :

$$\begin{cases} na + b \sum x = \sum y \\ a \sum x + b \sum x^2 = \sum xy \end{cases} \quad (25)$$

Можно воспользоваться готовыми формулами, которые вытекают непосредственно из решения этой системы:

$$a = \bar{y} - b\bar{x}, \quad b = \frac{cov(x,y)}{\sigma_x^2} \quad (26)$$

где $cov(x,y) = \overline{yx} - \bar{y}\bar{x}$ - ковариация признаков x и y , $\sigma_x^2 = \overline{x^2} - \bar{x}^2$ - дисперсия признака x и $\bar{x} = \frac{1}{n} \sum x$, $\bar{y} = \frac{1}{n} \sum y$, $\overline{yx} = \frac{1}{n} \sum yx$, $\overline{x^2} = \frac{1}{n} \sum x^2$.

Ковариация – числовая характеристика совместного распределения двух случайных величин, равная математическому ожиданию произведения отклонений этих случайных величин от их математических ожиданий. Дисперсия – характеристика случайной величины, определяемая как математическое ожидание квадрата отклонения случайной величины от ее математического ожидания. Математическое ожидание – сумма произведений значений случайной величины на соответствующие вероятности.

Тесноту связи изучаемых явлений оценивает линейный коэффициент парной корреляции r_{xy} для линейной регрессии ($-1 \leq r_{xy} \leq 1$):

$$r_{xy} = b \frac{\sigma_x}{\sigma_y} = \frac{cov(x,y)}{\sigma_x \sigma_y} \quad (27)$$

Индекс корреляции ρ_{xy} - для нелинейной регрессии ($0 \leq \rho_{xy} \leq 1$):

$$\rho_{xy} = \sqrt{1 - \frac{\sigma_{ост}^2}{\sigma_y^2}} = \sqrt{1 - \frac{\sum (y - \widehat{y}_x)^2}{\sum (y - \bar{y})^2}}, \quad (28)$$

где $\sigma_y^2 = \sum(y - \bar{y})^2$ – общая дисперсия результативного признака y ;

$\sigma_{\text{ост}}^2$ – остаточная дисперсия, определяемая исходя из уравнения регрессии
 $\bar{y}_x = f(x)$.

Оценку качества построенной модели даст коэффициент (индекс) детерминации $r_{xy}^2 = R^2$ (для линейной регрессии) либо ρ_{xy}^2 (для нелинейной регрессии), а также средняя ошибка аппроксимации.

3.2.6 Технологии вычислений в MS Excel

Пример 2.1 Построение диаграммы рассеяния. Построение линейного уравнения. Расчет коэффициента корреляции, обычного и исправленного коэффициентов, средней ошибки аппроксимации.

1. Построить диаграмму рассеяния. Построить линейное уравнение парной регрессии y от x .

Для построения диаграммы рассеяния необходимо построить таблицу с данными и построить точечную диаграмму.

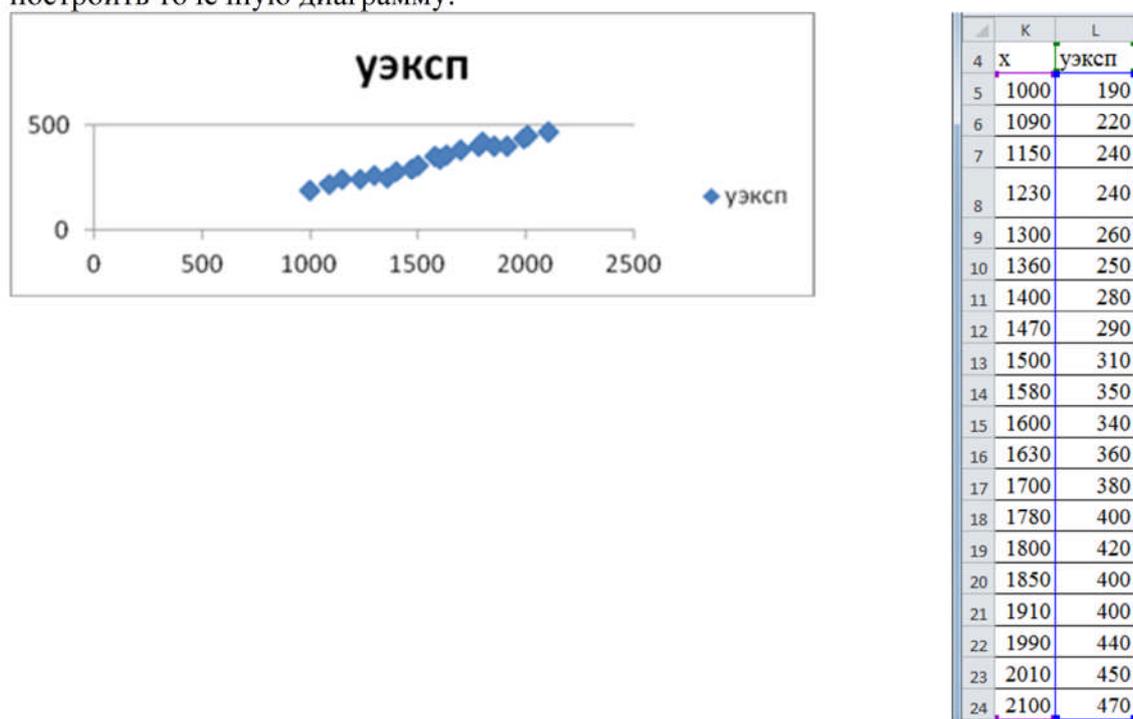
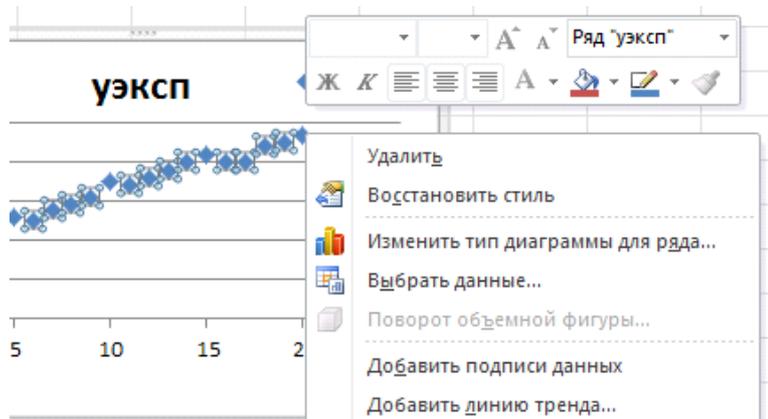


Рисунок 3.2 – Таблица данных и диаграмма рассеяния

Необходимо найти значения параметров линейной регрессии: b_1 и b_0 .

Способ 1 (с помощью добавления тренда):



- Выделить точки наблюдения, вызвать контекстное меню, в нем выбрать пункт *Добавить линию тренда...*

Рисунок 3.3 - Добавление линии тренда

Появится окно *Формат линии тренда*, в котором установить флажки: *Линейная*, *Показать уравнение на диаграмме*, *Поместить на диаграмму величину достоверности аппроксимации (R^2)*, нажать *Заккрыть*.

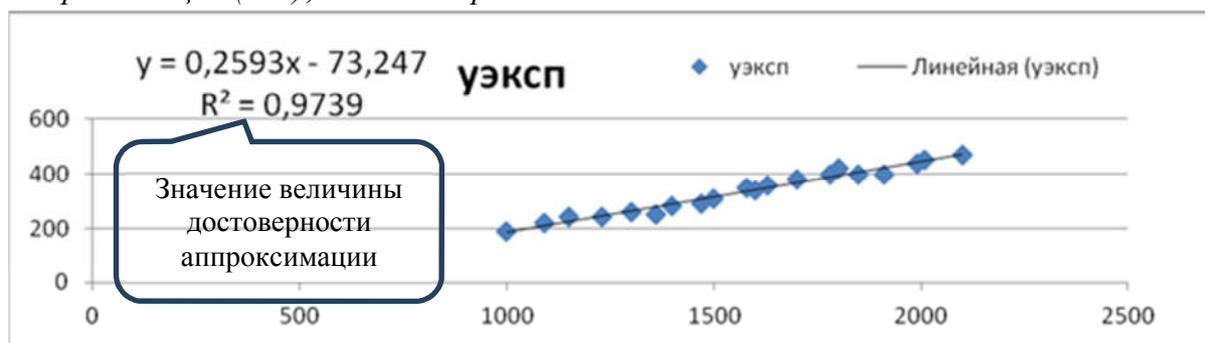


Рисунок 3.4 – Построенная линия тренда

На диаграмме рассеяния показаны не только точки - объекты, но и теоретическая линия регрессии (прямая).

Примечание. Результат-график у каждого варианта свой, отличный от других.

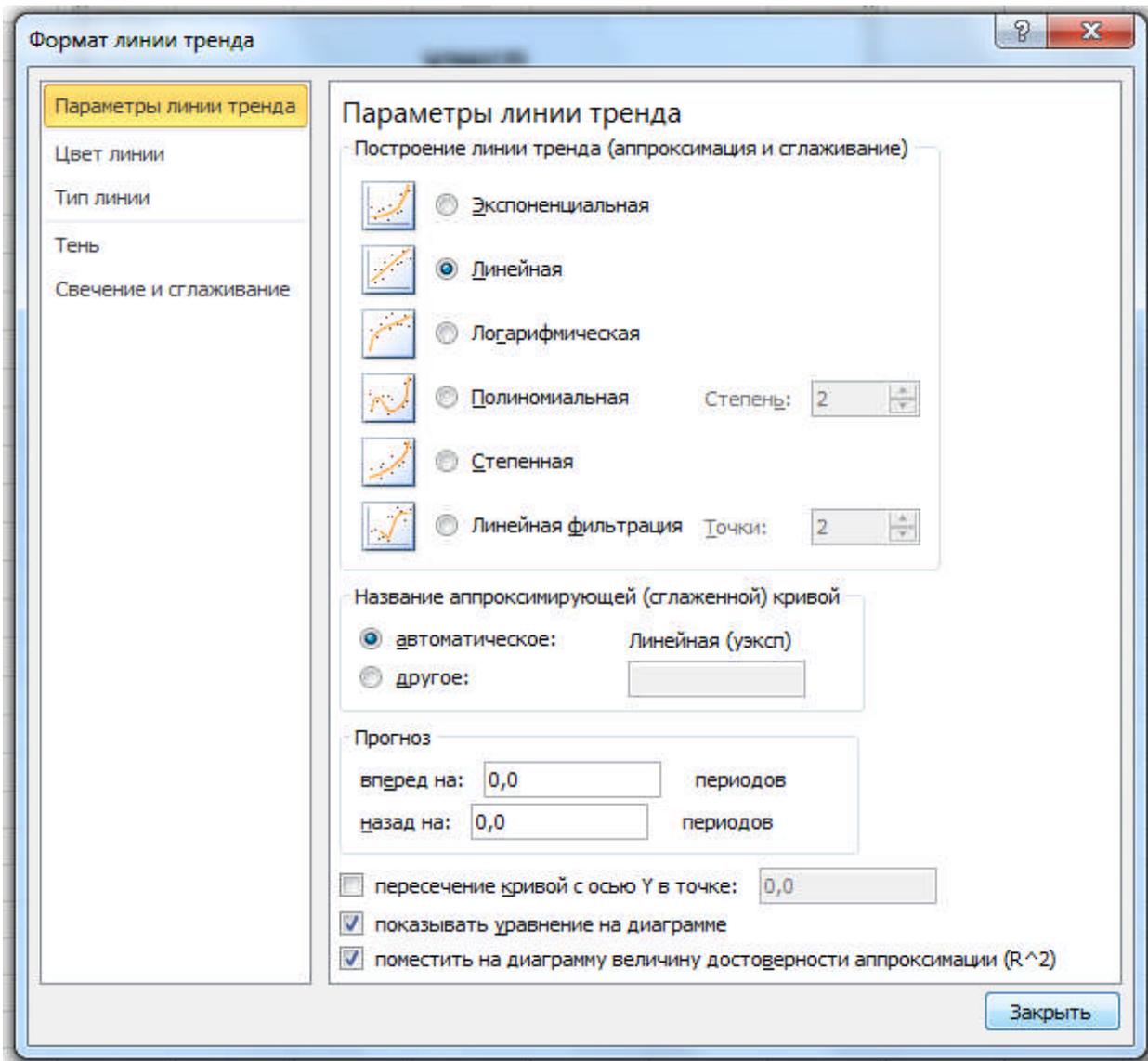


Рисунок 3.5 – Формат линии тренда

Способ 2 (с помощью функций):

- с помощью функции ЛИНЕЙН(массив Y; массив X; 1; 1), ИЛИ ЛИНЕЙН(массив Y; массив X; 1; 0)
 - с помощью функций НАКЛОН и ОТРЕЗОК.
- Функция ЛИНЕЙН(массив Y; массив X; 1; 1) или ЛИНЕЙН(массив Y; массив X; 1; 0)
- Выделить диапазон ячеек (5 строчек и 2 столбца, например N5: O9 или 1 строчек и 2 столбца).
 - С помощью мастера функций f_x вставить функцию ЛИНЕЙН() из категории Статистические.
 - В окне Аргументы функции заполнить поля: Известные_значения_y, Известные_значения_x, Константа, Статистика.
 - Нажать одновременно <Ctrl>+<Shift>+<Enter>.
 - Появится таблица в диапазоне ячеек N5:O9.

Таблица 3.3 – Результат вычисления ЛИНЕЙН()

Значение коэффициента b_1	Значение коэффициента b_0
Стандартная ошибка b_1	Стандартная ошибка b_0
Коэффициент детерминации R^2	Стандартная ошибка y
F-статистика	Число степеней свободы
Регрессионная сумма квадратов	Остаточная сумма квадратов

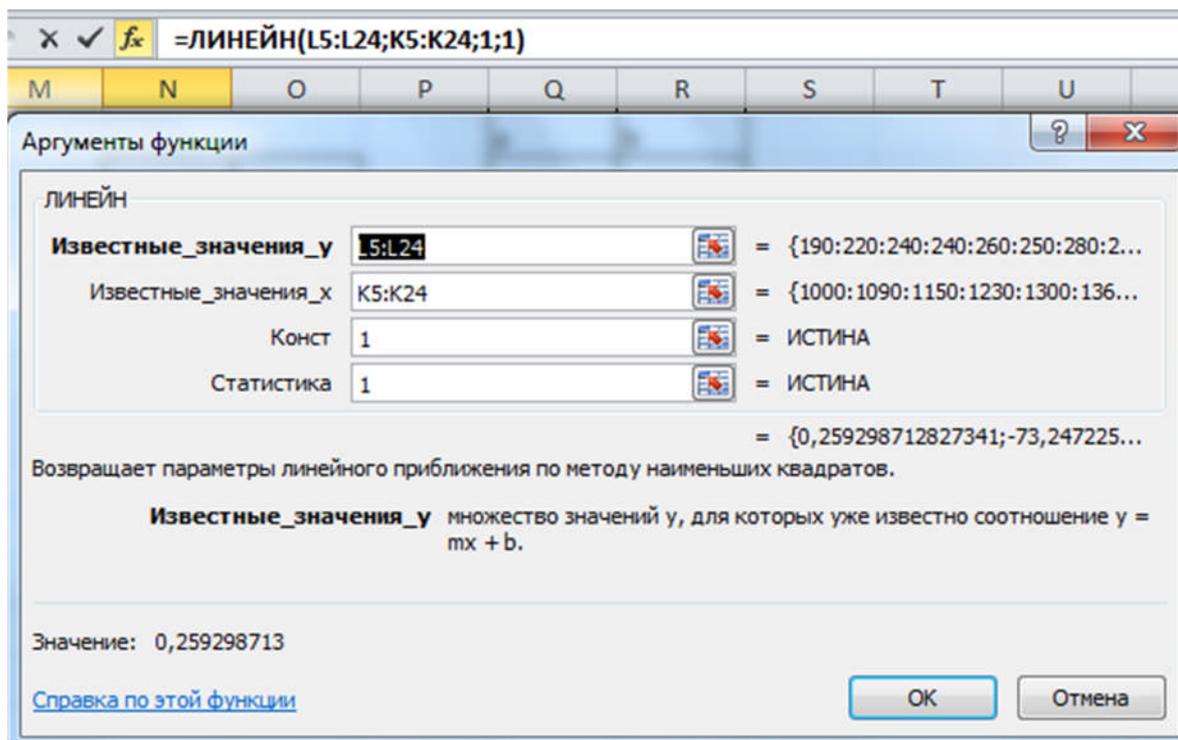


Рисунок 3.6 – Аргументы функции ЛИНЕЙН()

	N	O	P	Q	R	S
5	0,259299	-73,2472		0,259298713	-73,24722592	
6	0,009999	16,03423				
7	0,97393	14,03959				
8	672,4519	18				
9	132547	3547,98				

Рисунок 3.7 – Результат вычисления ЛИНЕЙН()

Получили уравнение вида: $y = -0.45034x + 5.919105$.

функции НАКЛОН (для нахождения b_1) и *ОТРЕЗОК* (для b_0)

	Q	R		Q	R
4	b1	b0	4	b1	b0
5	0,25929871	-73,24722592	5	=ЛИНЕЙН(L5:L24;K5:K24;1;0)	=ЛИНЕЙН(L5:L24;K5:K24;1;0)
6	0,25929871	-73,24722592	6	=НАКЛОН(L5:L24;K5:K24)	=ОТРЕЗОК(L5:L24;K5:K24)

Получено уравнение регрессии: $y = 0,259299 - 73,2472$.

Способ 3 (анализ данных Регрессия):

Использовать инструмент анализа данных *Регрессия* для получения: результатов регрессионной статистики, дисперсионного анализа, доверительных интервалов, графиков подбора линии регрессии, остатков и нормальной вероятности. Порядок действий следующий: вкладка *Файл\Параметры\Надстройки\нажать Перейти\ в окне Надстройки установить флажок *Пакет анализа\ нажать ОК и вкладка Данные\ группа Анализ \ Анализ данных \ в окне Анализ данных выбрать Регрессия и нажать ОК.**

Появится диалоговое окно *Регрессия* входных данных и параметров вывода. Необходимо заполнить в нем пустые поля.

Входной интервал y – диапазон, содержащий данные результативного признака;

Входной интервал x – диапазон, содержащий данные факторного признака;

Метки – флажок, который указывает, содержит ли первая строка названия столбцов или нет;

Константа - ноль – логическое значение, которое указывает на наличие или отсутствие свободного члена в уравнении регрессии;

Выходной интервал – достаточно указать левую верхнюю ячейку будущего диапазона;

Новый рабочий лист – можно создать произвольное имя нового листа.

Примечание. Если необходимо получить информацию об остатках и графики остатков, установите соответствующие флажки в диалоговом окне.

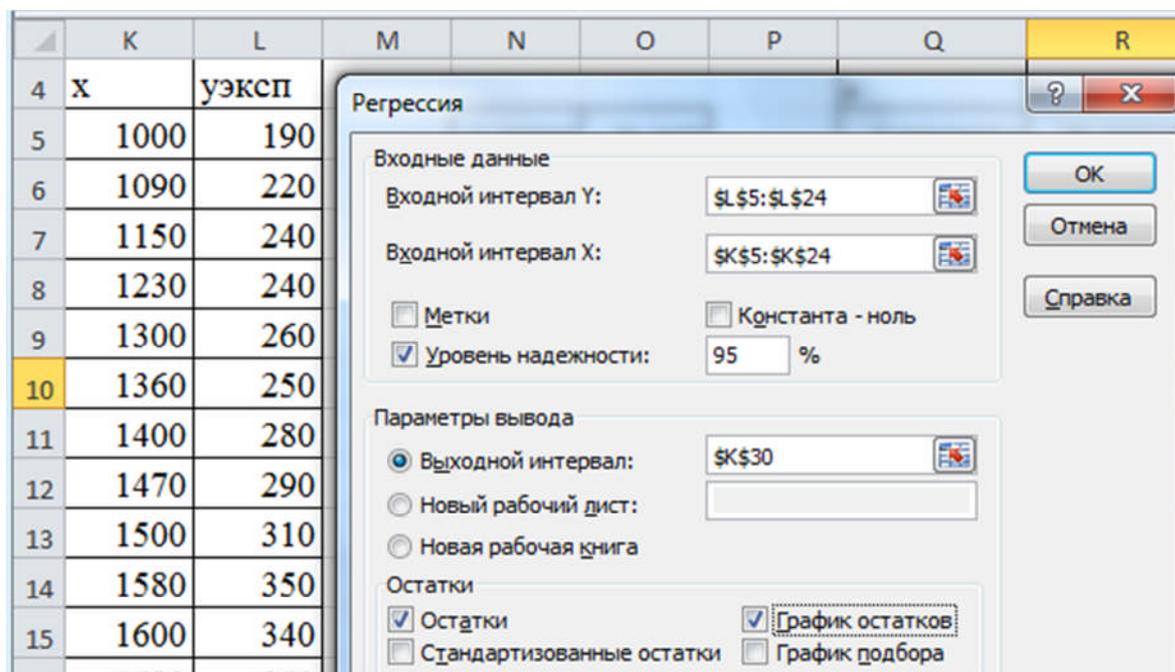


Рисунок 3.8 - Диалоговое окно ввода параметров инструмента РЕГРЕССИЯ

Результаты регрессионного анализа:

	К	L	M	N	O	Р
30	ВЫВОД ИТОГОВ					
31						
32	<i>Регрессионная статистика</i>					
33	Множественный R	0,986878979				
34	R-квадрат	0,973930119				
35	Нормированный R-квадрат	0,972481792				
36	Стандартная ошибка	14,03958782				
37	Наблюдения	20				
38						
39	<i>Дисперсионный анализ</i>					
40		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>
41	Регрессия	1	132547,0195	132547,0195	672,4519403	1,04396E-15
42	Остаток	18	3547,98047	197,1100261		
43	Итого	19	136095			
44						
45		<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>
46	Y-пересечение	-73,24722592	16,03423416	-4,568177389	0,000238365	-106,9339019
47	Переменная X 1	0,259298713	0,009999303	25,93167832	1,04396E-15	0,238290956

Рисунок 3.9 - Результат применения инструмента Регрессия

Способ 4 (с помощью вычислений):

Аналитический способ нахождения значений b_1 , b_0 .

- Построить таблицу вида:

x_i	$y_{\text{эксп}} = y_i$	$y_{\text{теор}} = y(x_i)$	$y_i * x_i$	x_i^2	y_i^2	$(y_i - y_{\text{ср}})^2$	$(y_i - y(x_i))^2$	$(x_i - x_{\text{ср}})^2$	$ y_i - y(x_i) : y_i$	№
										1
										2
										...
										20
										сумма
$x_{\text{ср}}$	$y_{\text{ср}}$									Ср. знач.

- Вычислить значения b_1 , b_0 с помощью формул (19) и (20).

2. Рассчитать линейный коэффициент парной корреляции, коэффициенты детерминации (обычный и исправленный) и среднюю ошибку аппроксимации.

Выборочные средние:

$$x_{\text{ср}} = \sum x_i / n; y_{\text{ср}} = \sum y_i / n; x y_{\text{ср}} = \sum x_i y_i / n$$

Выборочные дисперсии:

$$s^2(x) = \sum x_i^2 / n - x_{\text{ср}}^2; s^2(y) = \sum y_i^2 / n - y_{\text{ср}}^2$$

Таблица 3.4 – Коэффициенты: регрессии, парной корреляции, детерминации, средняя ошибка аппроксимации

Обозначение	Значение	Вычисление
Линейный коэффициент парной корреляции (r_{xy})	0,98687879	КОРРЕЛ (диапазон X; диапазон Y) $r_{xy} = \frac{\sum_{i=1}^n (x_i - x_{cp})(y_i - y_{cp})}{\sqrt{\sum_{i=1}^n (x_i - x_{cp})^2 \sum_{i=1}^n (y_i - y_{cp})^2}}$
Коэффициент детерминации обычный (R^2)	0,973930119	Рассчитан (см. рис.2.6 и 2.11) $R^2 = (r_{xy})^2$ или $R^2 = 1 - \frac{\sum (y_i - y(x_i))^2}{\sum (y_i - y_{cp})^2}$
Коэффициент детерминации исправленный (нормированный) ($R^2_{\text{скор}}, \bar{R}^2$)	0,972482	Рассчитан (см. рис.2.11) $\bar{R}^2 = 1 - (1 - R^2) \frac{n - 1}{n - m - 1}$ где m - количество объясняющих переменных модели (число факторов), n - число наблюдений
Средняя ошибка аппроксимации (\bar{A})	3,43%	$\bar{A} = \frac{\sum y_i - y(x_i) : y_i}{n} \cdot 100(\%)$
Коэффициент регрессии (b_1)	0,26	Рассчитан

	Q	R	S
8	коэфф корреляции	R^2	R^2 норм
9	0,986878979	0,973930119	0,972482

	Q	R	S
8	коэфф корреляции	R^2	R^2 норм
9	=КОРРЕЛ(K5:K24;L5:L24)	=Q9^2	=1-(1-R9)*(20-1)/(20-1-1)

Рисунок 3.13 - Результат расчета коэффициентов

Пример 2.2 Матрица парных коэффициентов корреляции переменных

3. Рассчитать матрицу парных коэффициентов корреляции.

Примечание. Матрицу парных коэффициентов корреляции переменных можно рассчитать, используя инструмент анализа данных **Корреляция**. Для этого:

- 1) Выполнить команду: вкладка *Данные* \ группа *Анализ данных* \ *Корреляция*.
- 2) Указать диапазон данных;

После вычисления парных коэффициентов корреляции построится матрица, которая должна иметь следующее соответствие:

$\sum n$	$\sum y$	$\sum x_1$	$\sum x_2$	$\sum x_3$
$\sum y$	$\sum y^2$	$\sum x_1 y$	$\sum x_2 y$	$\sum x_3 y$
$\sum x_1$	$\sum yx_1$	$\sum x_1^2$	$\sum x_2 x_1$	$\sum x_3 x_1$
$\sum x_2$	$\sum yx_2$	$\sum x_1 x_2$	$\sum x_2^2$	$\sum x_3 x_2$
$\sum x_3$	$\sum yx_3$	$\sum x_1 x_3$	$\sum x_2 x_3$	$\sum x_3^2$

Способ 1 (использование функций)

G10		fx		=КОРРЕЛ(B3:B19;C3:C19)	
	E	F	G	H	I
7		y	x1	x2	
8	y	1			
9	x1	0,906518213	1		
10	x2	0,986942369	0,868169123	1	

	E	F	G	H
7		y	x1	x2
8	y	=КОРРЕЛ(A3:A19;A3:A19)		
9	x1	=КОРРЕЛ(A3:A19;B3:B19)	=КОРРЕЛ(B3:B19;B3:B19)	
10	x2	=КОРРЕЛ(A3:A19;C3:C19)	=КОРРЕЛ(B3:B19;C3:C19)	=КОРРЕЛ(C3:C19;C3:C19)

Рисунок 3.14 - КОРРЕЛ()

Способ 2 (инструмент Корреляция)

вкладка Данные / группа Анализ / Анализ данных / Корреляция.

	A	B	C	D	E	F	G	H	I	J
1	Заработная плата	Возраст	Стаж							
2	y	x1	x2							
3	684,48	28	10							
4	674,45	26	8							
5	729,62	30	14							
6	748,86	35	15							
7	761,45	42	16							
8	775,51	40	16							
9	628,81	27	3							
10	731,84	35	13							
11	698,81	30	10							
12	645,64	29	7							
13	711,18	33	11							
14	798,07	40	20							
15	832,83	40	24							
16	607,61	23	2							
17	728,62	36	12							
18	667,14	30	6							
19	683,71	31	10							

Корреляция

Входные данные
 Входной интервал:

Группирование:
 по столбцам
 по строкам

Метки в первой строке

Параметры вывода
 Выходной интервал:
 Новый рабочий лист:
 Новая рабочая книга

Рисунок 3.15 - Инструмент Корреляция

G11		fx		
	E	F	G	H
2		Столбец 1	Столбец 2	Столбец 3
3	Столбец 1	1		
4	Столбец 2	0,906518213	1	
5	Столбец 3	0,986942369	0,868169123	1

Рисунок 3.16 - Результат вычислений инструмента Корреляция

Пример 2.2 Линейное уравнение множественной регрессии

3. Построить линейное уравнение множественной регрессии.

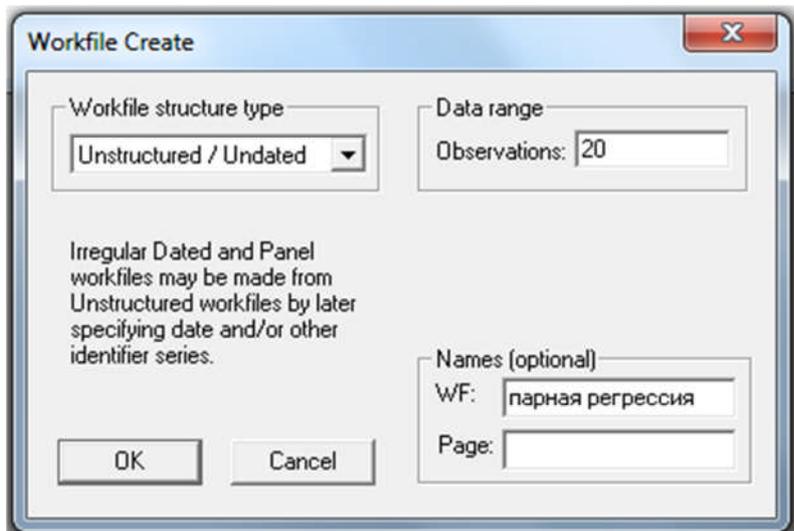
Таблица 3.5 – Значения коэффициентов

	Коэффициенты
У-пересечение	541,8833652
Переменная X 1	2,176832561
Переменная X 2	8,57091736

Пример 2.1 Построение диаграммы рассеяния. Построение линейного уравнения. Расчет коэффициента корреляции, обычного и исправленного коэффициентов, средней ошибки аппроксимации.

3.2.7 Технологии вычислений в EViews

1. Необходимо создать файл (Workfile) со своим именем и импортировать исходные данные из файла. Смотрите приложения Ж, З, И .

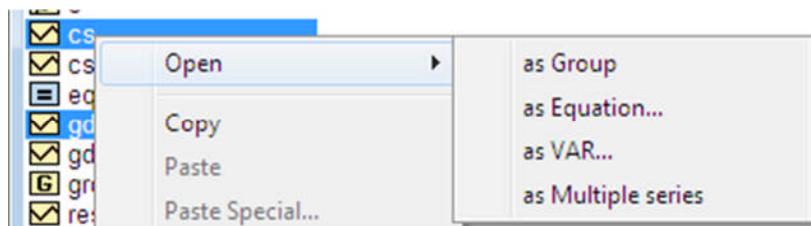


2. Необходимо построить графики временных рядов (представить данные в графическом виде - *построить диаграмму рассеивания*).

Для построения диаграммы рассеяния необходима таблица (группа), отображающая значения независимой (фактор) и зависимой переменной (результат).

- Сохранить набор переменных в объекте Group.

Способ 1 (предпочтительный):



Удерживая клавишу CTRL, выделить сначала независимую (фактор), а затем зависимую переменную (результат). Вызвать контекстное меню. В нем выбрать команду **Open- as / Group**.

Рисунок 3.17 - Создание объекта Группа

Способ 2

Необходимо в объекте Equation выбрать пункт *Procs, Make Regressor Group*.

Появится объект Group с зависимой и независимой переменными. Для сохранения выбрать закладку Name и ввести допустимое имя.

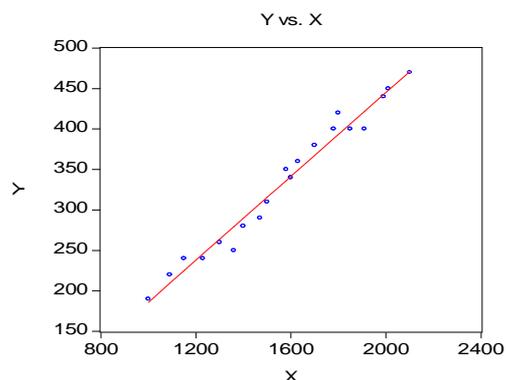


Рисунок 3.18 - Диаграмма рассеяния (EViews)

- Построить диаграмму рассеяния: в окне объекта *Group* выбрать закладку *View*, а затем *Graph\Scatter\Scatter with Regression*.

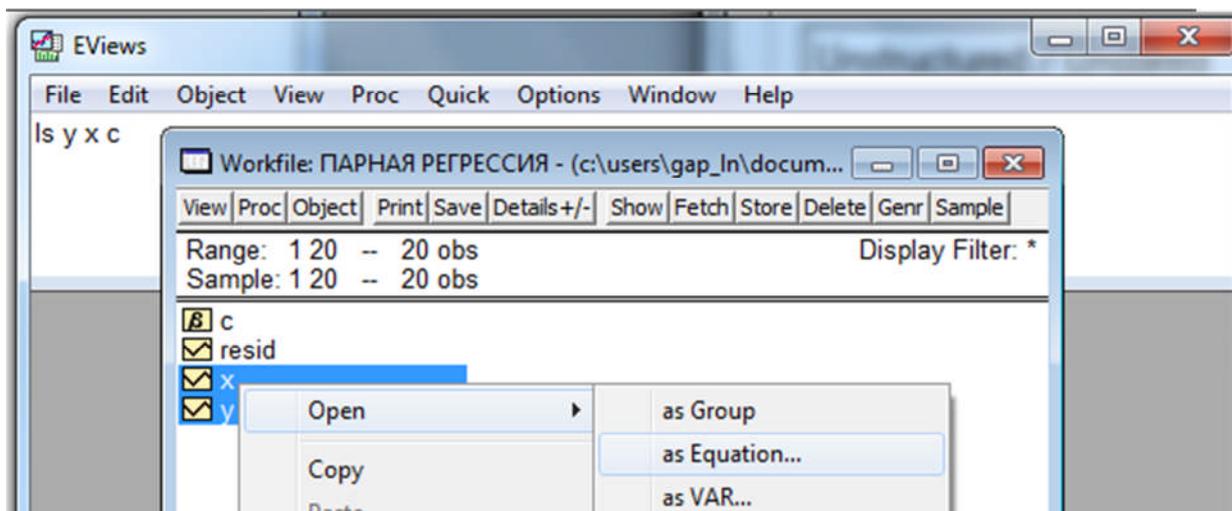


Рисунок 3.19 – Команда меню (создание диаграммы рассеяния)

3. Создать уравнение парной регрессии и его оценки.
Можно несколькими способами:
 - Необходимо *последовательно выделить* левой кнопкой мыши сначала *зависимую переменную*, а затем *последовательно независимые переменные*, удерживая клавишу *Ctrl*. Отпустите клавишу *Ctrl*. Нажать правую кнопку мыши и выбрать в появившемся контекстном меню *Open as Equation*. Появится диалоговое окно *Equation Specification*.
 В окне будет указана спецификация уравнения для оценивания в *EViews*.
 - Необходимо в окне ввода команд ввести команду *ls*, указать *имя зависимой переменной*, затем через пробел *имена независимых переменных*, и в конце указать служебную переменную *C*.
 - Необходимо в окне ввода команд ввести команду и нажать *Enter*.

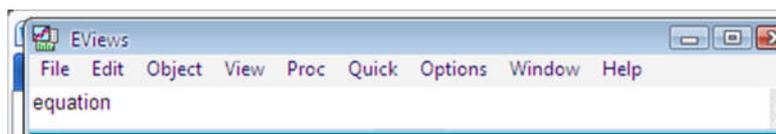


Рисунок 3.20 – Создание объекта Equation

- Необходимо в основном меню *EViews* выбрать пункт *Quick, Estimate Equation*.
 Появится диалоговое окно *Equation Specification*. Затем в появившемся диалоговом окне необходимо указать спецификацию уравнения, указав сначала *имя зависимой переменной*, затем через пробел *имена независимых переменных*, и в конце указать служебную переменную *C*, в которой будут сохранены оценки коэффициентов регрессии.

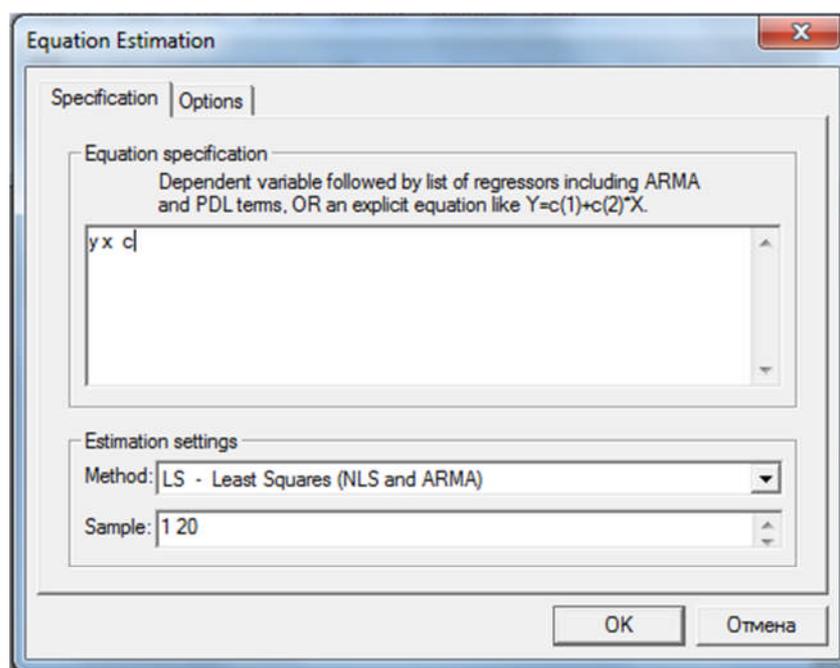


Рисунок 3.21 - Окно Equation Specification

В поле Method необходимо выбрать метод оценки коэффициентов уравнения. По умолчанию выбран метод наименьших квадратов LS – Least Squares (NLS and ARMA). В поле Sample можно указать произвольный поддиапазон данных, по которому будет производиться оценивание. После нажатия кнопки ОК будет создан объект Equation, в котором показаны результаты оценивания.

Примечание. Оценки метода наименьших квадратов существуют не всегда. В данном случае объекта Equation создано не будет, и появится предупреждение об ошибке Near singular matrix (матрица вырождена), что говорит о несуществовании МНК-оценок.

Таблица 3.6 – Характеристики регрессионного уравнения

Dependent Variable: Y
 Method: Least Squares
 Date: 10/24/14 Time: 12:42
 Sample: 1 20
 Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X	0.259299	0.009999	25.93168	0.0000
C	-73.24723	16.03423	-4.568177	0.0002
R-squared	0.973930	Mean dependent var		334.5000
Adjusted R-squared	0.972482	S.D. dependent var		84.63389
S.E. of regression	14.03959	Akaike info criterion		8.216279
Sum squared resid	3547.980	Schwarz criterion		8.315852
Log likelihood	-80.16279	F-statistic		672.4519
Durbin-Watson stat	1.257096	Prob(F-statistic)		0.000000

4. *Сохранить новый объект Equation.*

Для сохранения нового объекта необходимо выбрать в окне *объекта Equation* закладку *Name* и ввести допустимое имя без символов кириллицы. После нажатия кнопки «ОК» объект *Equation* будет сохранен с введенным вами именем и отображен в основном рабочем окне в виде пиктограммы со знаком равенства.

5. *Получить в текстовом виде спецификацию регрессионного уравнения, оцениваемое уравнение регрессии.*

Выбрав в окне *объекта Equation* закладку *View* и в раскрывшемся меню пункт *Representations*, можно получить в текстовом виде спецификацию регрессионного уравнения, оцениваемое уравнение регрессии и, что особенно важно, уравнение регрессионной прямой (регрессионное уравнение с подставленными коэффициентами). Выделив текст, его можно копировать в другое приложение, например, текстовый редактор.

Estimation Command:

=====
LS Y X C

Estimation Equation:

=====
 $Y = C(1)*X + C(2)$

Substituted Coefficients:

=====
 $Y = 0.2592987128*X - 73.24722592$

6. *Создать новый объект Table*

Для сохранения результатов (расчетов описательной статистики - закладке *Stats*) можно выбрать закладку *Freeze объекта Equation*. В результате будет создан новый объект Table, куда будет скопирован текущий текст результатов расчетов.

7. *Сохранить объект Table.*

Для сохранения необходимо выбрать закладку *Name объекта Table* и в появившемся диалоговом окне ввести допустимое имя без символов кириллицы. После нажатия кнопки *Ok* объект *Table* будет сохранен с введенным вами именем и отображен в рабочем окне *Eviews* в виде пиктограммы с изображением таблицы и надписью «Table».

8. *Экономическая интерпретация оценок коэффициентов регрессии.*

Обозначили за *y* ряд данных для реального объема экспорта, за *x* – реальный ВВП. Вывести результаты, отображаемые в объекте *Equation* после проведения оценивания коэффициентов регрессионного уравнения $y = C(1) \times x + C(2)$ методом наименьших квадратов.

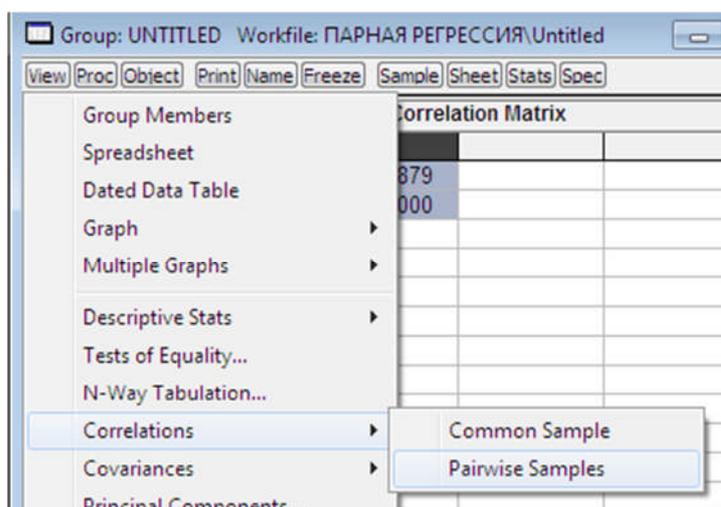
Коэффициент C(1) есть наклон регрессионной прямой, *C(2)* – константа данной парной регрессии. *Коэффициент при переменной x* означает, что, по имеющимся наблюдениям, при росте (снижении) реального ВВП на единицу реальный объем экспорта увеличивается (уменьшается) в среднем на *C(1)*.

9. *Рассчитать коэффициент корреляции.*

- Построить группу вида:

у	х
значения Y	значения X

- Выполнить *View\Correlations\...*



Common	общий
Pairwise	парно

	Y	X
Y	1	0.986878978843904
X	0.986878978843904	1

Пример 2.2 Матрица парных коэффициентов корреляции. Линейное уравнение множественной регрессии.

	Y	X1	X2
Y	1	0.906518212672502	0.986942368829694
X1	0.906518212672502	1	0.868169122803393
X2	0.986942368829694	0.868169122803393	1

Dependent Variable: Y
 Method: Least Squares
 Date: 11/11/14 Time: 12:55
 Sample: 1 17
 Included observations: 17

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X1	2.176833	0.733225	2.968847	0.0102
X2	8.570917	0.717442	11.94649	0.0000
C	541.8834	17.33400	31.26131	0.0000

R-squared	0.984079	Mean dependent var	712.2724
Adjusted R-squared	0.981804	S.D. dependent var	60.31206
S.E. of regression	8.135566	Akaike info criterion	7.189153
Sum squared resid	926.6240	Schwarz criterion	7.336190
Log likelihood	-58.10780	F-statistic	432.6659
Durbin-Watson stat	2.497452	Prob(F-statistic)	0.000000

$$Y = 2.176833 * X1 + 8.570917 * X2 + 541.8834 \text{ (специфика уравнения)}$$

3.2.8 Технологии вычислений в Statistica

Пример 2.1 Построение диаграммы рассеяния. Построение линейного уравнения. Расчет коэффициента корреляции, обычного и исправленного коэффициентов.

1. Запустите программу Statistica. Необходимо создать файл, содержащий таблицу наблюдений.

2. Построить диаграмму рассеяния (матричный график):

Вернуться в окно *Product-Moment Correlation*. Нажмите на кнопку *Two lists*. Откроется окно выбора переменных. Выберите в первом списке независимую переменную, во втором – зависимую переменную и нажмите ОК.

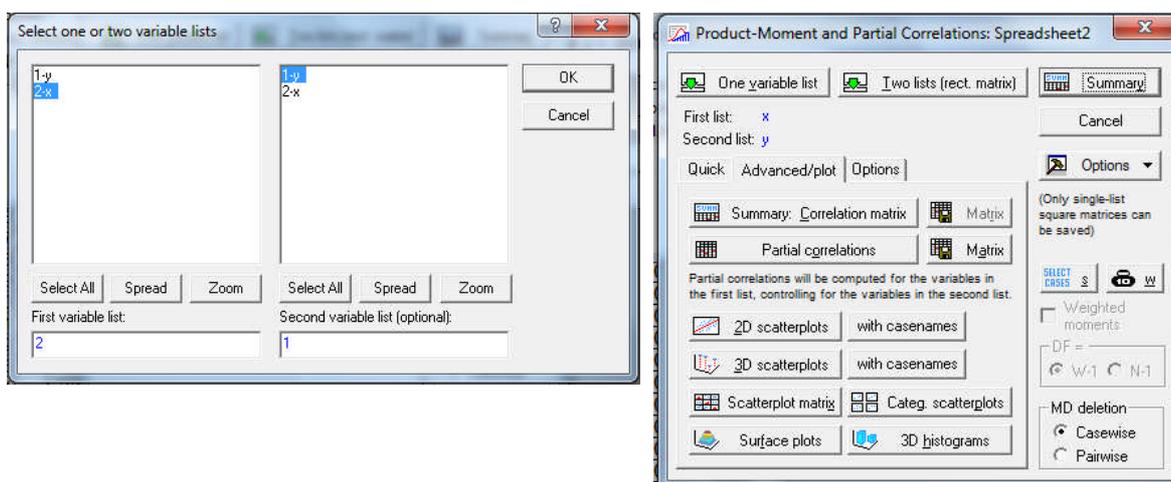


Рисунок 3.22 - Окно Product-Moment Correlation

Перейдите в окне *Product-Moment Correlation* к вкладке *Advanced/plot*. Нажмите кнопку *2D scatterplot* (2D диаграмма рассеяния). Появится окно диаграммы рассеяния для выбранных переменных.

Из графика отчетливо видно, что зависимость линейная, за исключением некоторых точек, отдаленных от линии. На графике программой предложена лучшая прямая. При смене коэффициента наклона подгонка будет хуже.

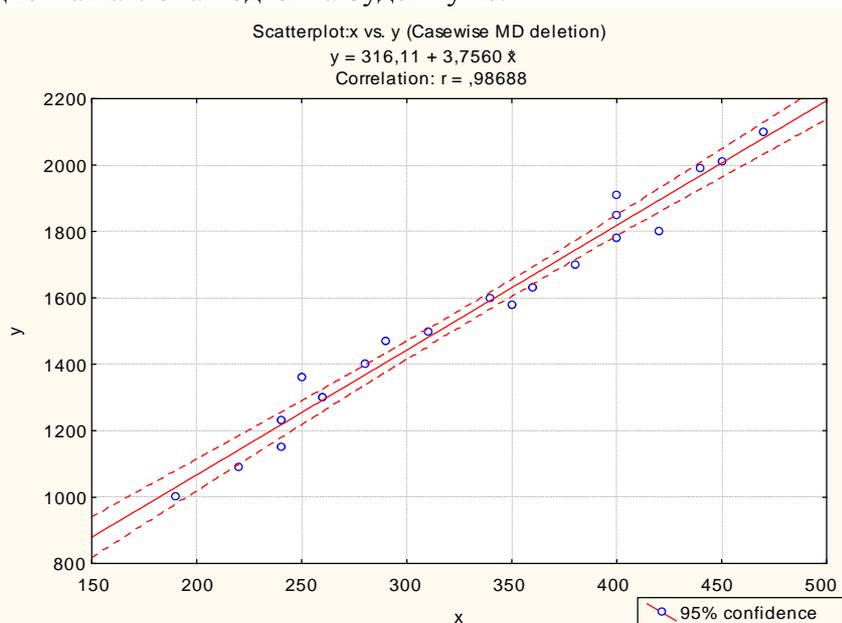


Рисунок 3.23 – Диаграмма рассеяния (Statistica)

Пример 2.2 Матричный график

3. Построить матричный график.

Вернитесь в окно *Product-Moment Correlation*. Щелкните по кнопке *One variable list* (Один список переменных). Выберите все переменные. Нажмите кнопку ОК. Далее в окне *Product-Moment Correlation* нажмите кнопку *Scatterplot matrix*. Нажмите ОК. На экране появится корреляционная матрица в графическом виде, позволяющая оценить линейные связи визуально.

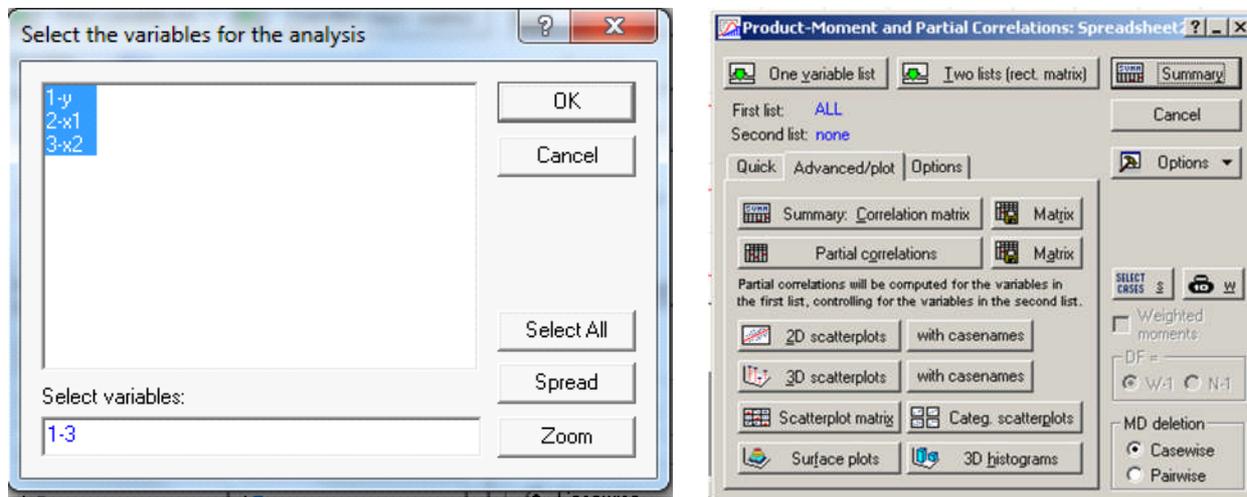


Рисунок 3.24 – Выбор переменных (Statistica)

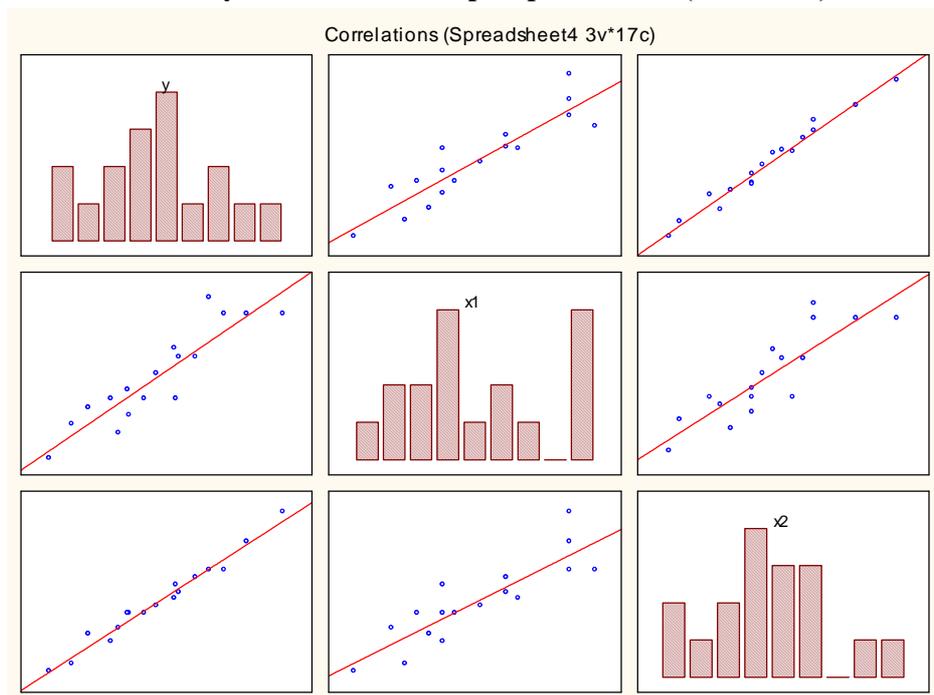


Рисунок 3.25 – Матричный график (Statistica)

4. Вычислить коэффициент корреляции (построить матрицу парных коэффициентов корреляции): главное меню *Статистика* \ модуль *Основная статистика/Таблицы* \ пункт *Correlation matrices* и ОК.

В модуле Основные статистики легко можно вычислить и проанализировать корреляционную матрицу выбранных вами переменных.

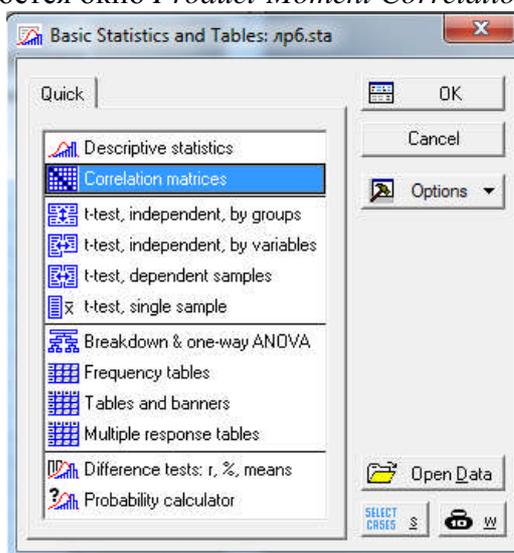
Примечание. Корреляция — статистическая взаимосвязь двух или нескольких случайных величин. При этом изменения одной или нескольких из этих величин приводят к систематическому изменению другой или других величин. Математической мерой корреляции (взаимосвязи) двух случайных величин служит коэффициент корреляции.

Метод вычисления коэффициента корреляции зависит от вида шкалы, к которой относятся переменные. Так, для измерения переменных с интервальной и количественной шкалами необходимо использовать коэффициент корреляции Пирсона.

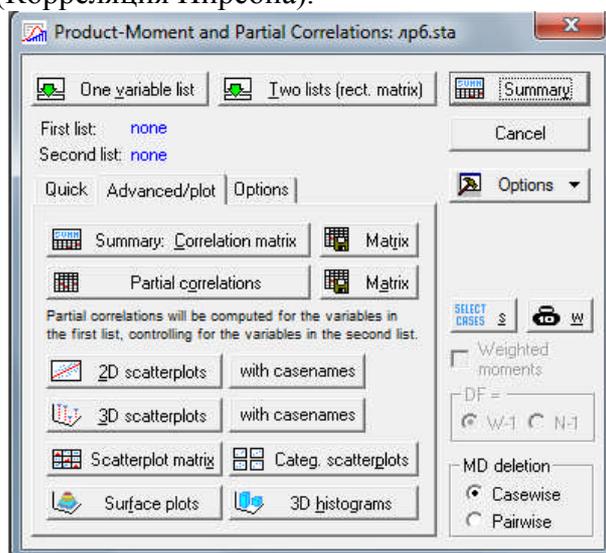
Коэффициент корреляции – это мера того, насколько хорошо два набора данных ложатся на прямую. Любой коэффициент корреляции изменяется в пределах от -1 до +1. Отрицательные значения говорят про обратнопропорциональную взаимосвязь, положительные - о прямопропорциональной.

Пример 2.1 Коэффициент корреляции

В стартовой панели модуля *Основные статистики* выберите *Correlation matrices*. Откроется окно *Product-Moment Correlation* (Корреляция Пирсона).

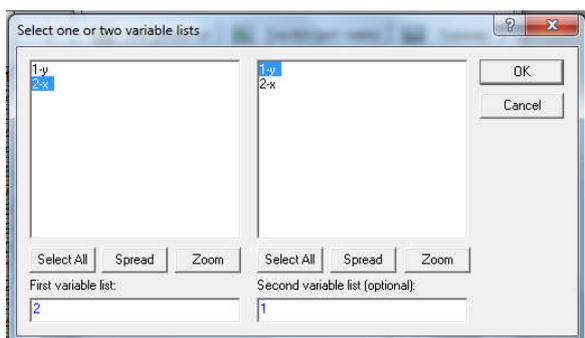


1

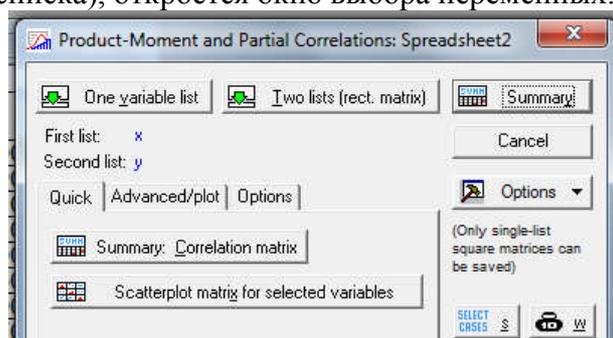


2

Нажмите на кнопку *Two lists* (Два списка), откроется окно выбора переменных.



3



4

Выберите независимые переменные в *First variables list* (Первый список) и зависимую переменную переменную в *Second variables list* (Второй список переменных). Нажмите в окне кнопку ОК, вернетесь в предыдущее окно.

В окне *Product-Moment Correlation* нажмите кнопку *Summary*. На экране вы увидите корреляционную матрицу.

В этой матрице имеется только один столбец, так как во втором списке выбрана одна зависимая переменная. Красным цветом автоматически выделены коэффициенты, значимые на уровне $p < 0,05$. Именно на эти коэффициенты следует обратить наибольшее внимание. Зависимость между переменными с выделенными красным цветом коэффициентами корреляции наиболее значимая.

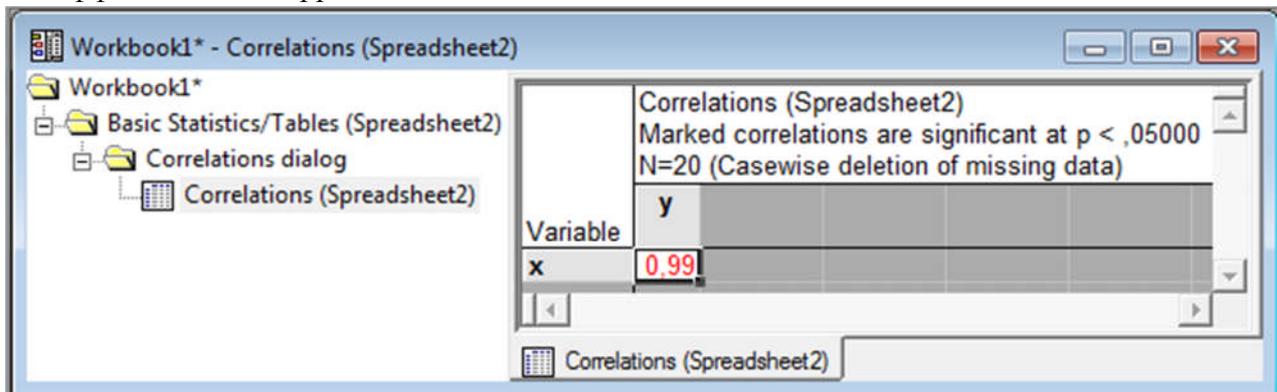


Рисунок 3.26 – Коэффициент корреляции

Пример 2.2 Матрица коэффициентов корреляции

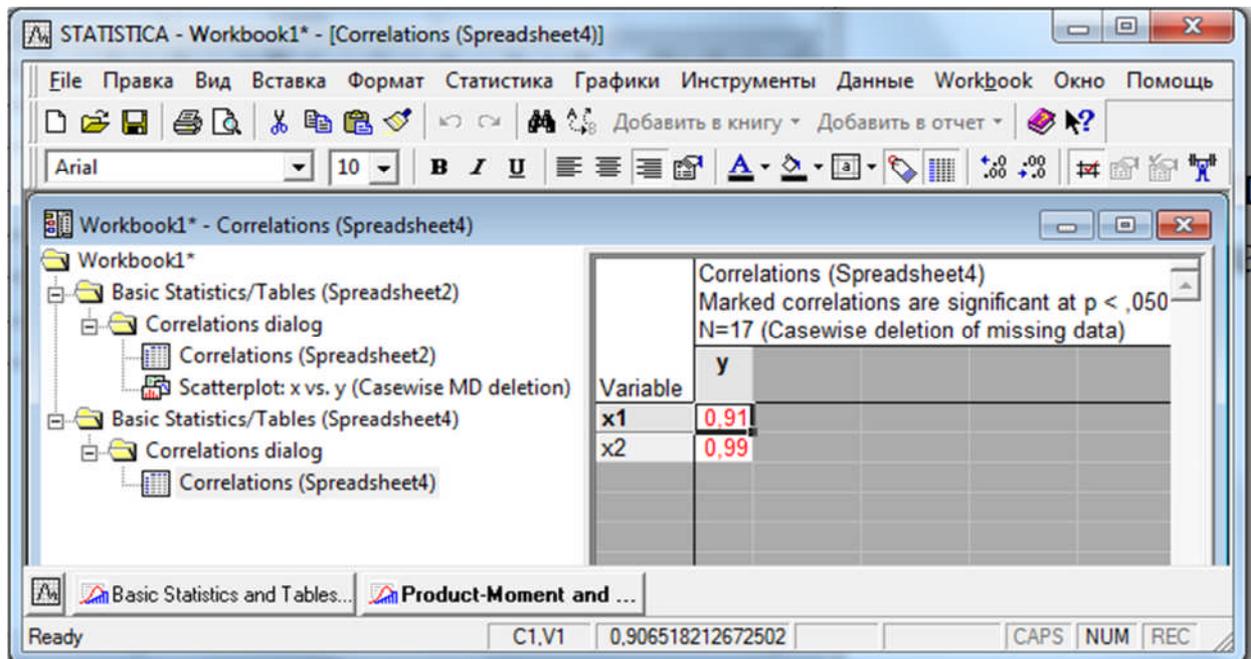


Рисунок 3.27 – Связь независимых переменных с зависимой

Построить матрицу парных коэффициентов корреляции: Вернитесь в окно *Product-Moment Correlation*. Нажмите на кнопку *Two lists*. Откроется окно выбора переменных. Выберите все переменные в первом и во втором списке, нажмите кнопку ОК. Вернетесь в предыдущее окно. В окне *Product-Moment Correlation* нажмите кнопку *Summary*. На экране появится корреляционная матрица.

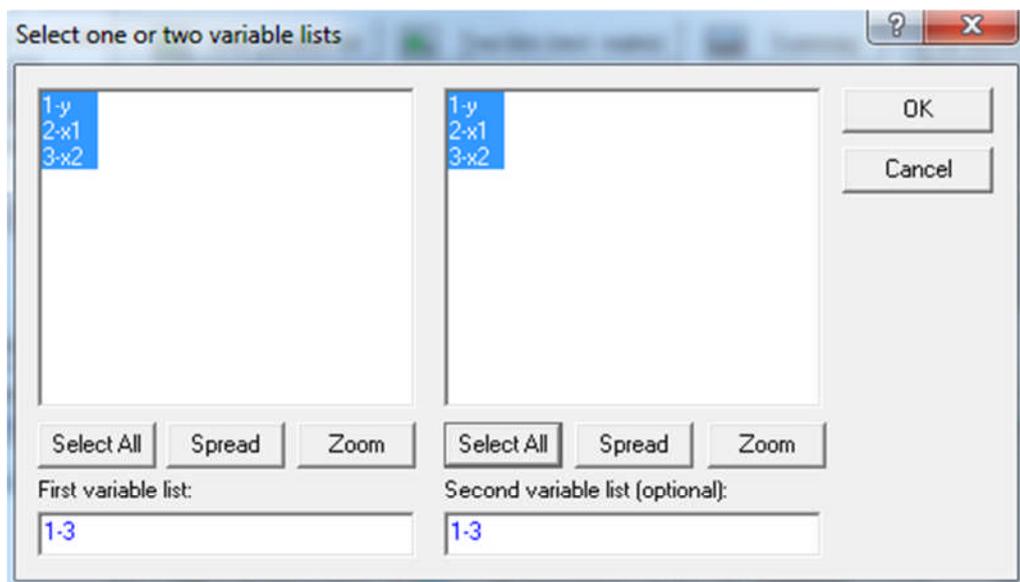


Рисунок 3.28 – Выбор переменных

Примечание. Рассчитаем для этих переменных матрицу парных коэффициентов корреляции с указанием для каждого коэффициента корреляции **p**-величины.

Variable	y	x1	x2
y	1.0000	.9065	.9869
	p=---	p=.000	p=.000
x1	.9065	1.0000	.8682
	p=.000	p=---	0
x2	.9869	.8682	1.0000
	0	0	p=---

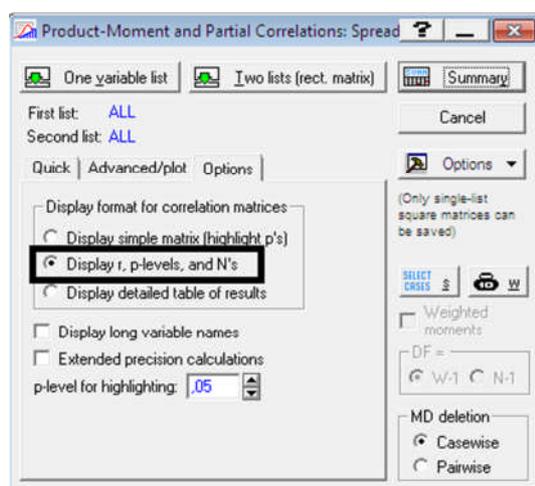


Рисунок 3.29 – Матрица парных коэффициентов корреляции с p-величинами

Свободный член и коэффициенты регрессии представлены в *графе В* $b_0=541,8834$; $b_1=2,1768$; $b_2=8,5709$. При этом уравнение множественной регрессии примет вид: $y=541,8834+2,1768*x_1+8,5709*x_2$.

Regression Summary for Dependent Variable: y (Spreadsheet2)						
R= ,99200747 R?= ,98407882 Adjusted R?= ,98180437						
F(2, 14)=432,67 p<,00000 Std.Error of estimate: 8,1356						
N=17	Beta	Std.Err. of Beta	B	Std.Err. of B	t(14)	p-level
Intercept			541,8834	17,33400	31,26131	0,000000
x1	0,201741	0,067953	2,1768	0,73322	2,96885	0,010160
x2	0,811797	0,067953	8,5709	0,71744	11,94649	0,000000

Рисунок 3.32 - Результаты построения множественной линейной регрессионной модели

Имеется возможность выполнить примеры с помощью онлайн-решения: <http://math.semestr.ru/corel/corel.php> (все калькуляторы: <http://math.semestr.ru/>)

3.2.8 Анализ результатов вычисления

Пример 2.1 Диаграммы рассеяния. Параметры линейного уравнения. Коэффициент корреляции (обычный и исправленный коэффициенты), средняя ошибка аппроксимации.

1. Произвести визуальный анализ диаграммы рассеяния.

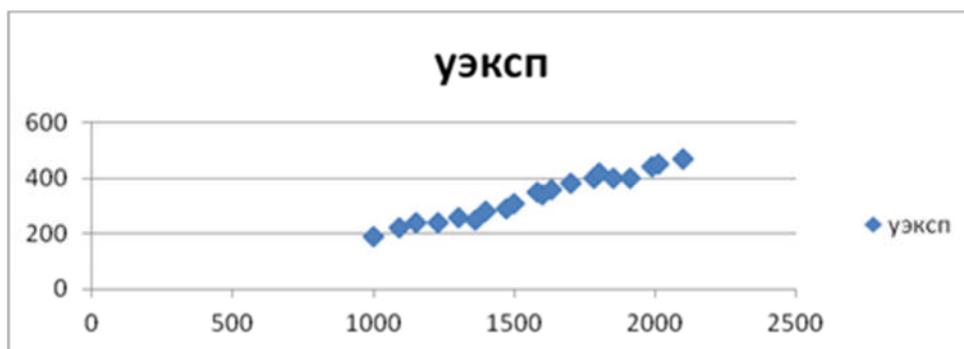


Рисунок 3.33 – Диаграмма рассеяния

Вывод. Форма – линейная, направление – прямое.

2. Определить влияние фактора на исследуемый отклик.

Примечание. Величина влияния фактора на исследуемый отклик может быть оценена при помощи коэффициента линейной парной корреляции, характеризующего тесноту (силу) линейной связи между двумя переменными.

Коэффициент обладает следующими свойствами:

- 1) не имеет размерности, следовательно, сопоставим для величин различных порядков;

- 2) изменяется в диапазоне от -1 до $+1$. Положительное значение свидетельствует о прямой линейной связи, отрицательное – об обратной. Чем ближе абсолютное значение коэффициента к единице, тем теснее связь. Считается, что связь достаточно сильная, если коэффициент по абсолютной величине превышает $0,7$, и слабая, если он менее $0,3$.

Качественная оценка тесноты связи величин X и Y может быть оценена на основе шкалы Чеддока:

Характеристика тесноты связи	Показатель тесноты связи: абсолютное значение коэффициента корреляции
Слабая	0,1-0,3
Умеренная	0,3-0,5
Заметная	0,5-0,7
Высокая (тесная, сильная)	0,7-0,9
Весьма высокая (очень тесная, весьма сильная)	0,9-0,99
Функциональная	1,0

Источник: Балинова В.С. Статистика в вопросах и ответах: Учеб. пособие. — М.: ТК Велби, Изд-во Проспект, 2004. — 344 с.

Линейный коэффициент корреляции ($r_{xy}=r_{yx}=r=0,98687898$)	Теснота связи между признаком y и фактором x весьма высокая, направление связи – прямая линейная связь.
--	---

3. Записать линейное уравнение парной регрессии. Дать интерпретацию коэффициента регрессии.

Прямая линия описывается уравнением вида ($y = b_1x + b_0$). В нашем примере параметры уравнения регрессии $y = 0,2593x - 73,247,92$ где y – результирующий признак, x – факторный признак, b_1 и b_0 – числовые параметры уравнения (b_1 есть наклон, b_0 – константа данной парной регрессии). Коэффициент b_1 в уравнении регрессии называется *коэффициентом регрессии*.

В общем случае коэффициент регрессии b_1 показывает, как в среднем изменится *результативный признак* (y) (понижится - «-» или повысится – «+»), если *факторный признак* (x) увеличится на единицу.

Коэффициент регрессии ($a=0,26$)	<i>Интерпретация коэффициента регрессии</i> . По имеющимся наблюдениям, при повышении величины фактора x на 1 единицу (ед. изм.) y повысится в среднем на 0,26 (ед. изм.).
------------------------------------	--

Свойства коэффициента регрессии

- Коэффициент регрессии принимает любые значения
- Коэффициент регрессии *не симметричен*, т.е. изменяется, если X и Y поменять местами
- *Единицей измерения* коэффициента регрессии является отношение единицы измерения Y к единице измерения X ($[Y] / [X]$)
- Коэффициент регрессии *изменяется при изменении единиц измерения* X и Y .

Примечание. Сравнение коэффициентов корреляции и регрессии представлено ниже.

Коэффициент корреляции	Коэффициент регрессии
<ul style="list-style-type: none"> • Принимает значения в диапазоне от -1 до +1 • Безразмерная величина • Показывает силу связи между признаками • Знак коэффициента говорит о направлении связи (положительная (прямая), отрицательная (обратная)) 	<ul style="list-style-type: none"> • Может принимать любые значения • Привязан к единицам измерения обоих признаков • Показывает структуру связи между признаками • Знак коэффициента говорит о направлении связи

4. Оценить коэффициенты детерминации и среднюю ошибку аппроксимации.

Величина коэффициента детерминации R^2 имеет смысл только в том случае, когда константа включена в состав регрессоров. R^2 показывает качество подгонки. Чем ближе значение к 1, тем ближе модель к собранным наблюдениям.

<p>Средняя ошибка аппроксимации - среднее отклонение расчетных значений от фактических. $\bar{A} = 3,43$</p>	<p>С помощью ошибки абсолютной аппроксимации производят оценку качества уравнения регрессии. Средняя ошибка аппроксимации. Ошибка аппроксимации в пределах 5%-7% свидетельствует о хорошем подборе уравнения регрессии к исходным данным. Поскольку ошибка меньше 7%, то данное уравнение можно использовать в качестве регрессии</p>
<p>Множественный коэффициент $R = 0,9869$</p>	<p>Тесноту совместного влияния факторов на результат оценивает индекс множественной корреляции</p>
<p>Коэффициент детерминации обычный $R^2 = 0,9739$</p>	<p>97,39% вариации зависимой переменной (у) объясняется вариацией независимой переменной (х). Остальные 2,61% вариации зависимой переменной объясняются другими факторами, неучтенными в модели. Другими словами - точность подбора уравнения регрессии – высокая</p>
<p>Скорректированный (нормированный, исправленный) коэффициент детерминации $R^2_{\text{скор}} = \bar{R}^2 = 0,9725$</p>	<p>Если значения обычного и исправленного коэффициентов детерминации различаются незначительно, а также $R^2 \geq R^2_{\text{скор}}$ и их значения велики, то считается, что регрессия аппроксимирует эмпирические данные достаточно точно и тем теснее наблюдения примыкают к линии регрессии</p>

В регрессионном анализе коэффициент детерминации R^2 свидетельствует о качестве регрессионной модели и отражает долю общей вариации результирующего признака у, объясненную изменением функции регрессии $f(x)$. Справедливо соотношение $0 \leq R^2 \leq 1$. Чем ближе этот коэффициент к единице, тем больше уравнение множественной регрессии объясняет поведение у.

Примечание. Если $R^2 > 0,8$, то качество подгонки регрессионной модели к наблюдаемым значениям y_i считается хорошим. Чем ближе коэффициент детерминации к единице, тем лучше точки на регрессионном поле укладываются на линию регрессии, т.е. тем выше уровень «подгонки» модели (Чем ближе этот коэффициент к единице, тем больше уравнение множественной регрессии объясняет поведение у).

Если $R^2 < 0,5$, то модель надо улучшить либо выбрав другие факторы, либо увеличив количество наблюдений (при функциональной связи между переменными R^2 равняется 1 (эмпирические точки (x_i, y_i) расположены на линии регрессии), а при отсутствии связи — 0).

Для того чтобы исследователи не увеличивали R^2 с помощью добавления дополнительных факторов, R^2 заменяется на скорректированный Adjusted R-squared (обозначение: \bar{R}^2), который даёт штраф за дополнительно включённые факторы.

Для небольших значений n (<30) необходимо использовать скорректированный коэффициент детерминации.

$\bar{R}^2 < R^2$ для $m > 1$, где m – количество факторов. Скорректированный коэффициент детерминации является более объективной оценкой. С ростом значения m

скорректированный коэффициент детерминации растет медленнее, чем обычный. Добавление в модель новых объясняющих переменных осуществляется до тех пор, пока растет скорректированный коэффициент детерминации. Очевидно, что $\bar{R}^2 = R^2$ только при $R^2 = 1$. \bar{R}^2 может принимать отрицательные значения.

Из двух вариантов уравнений, которые отличаются величиной скорректированного коэффициента детерминации, но имеют одинаково хорошие другие критерии качества, предпочитают вариант с большим значением скорректированного коэффициента детерминации.

Если величина скорректированного коэффициента детерминации достаточно велика, следовательно, при построении модели были учтены наиболее существенные факторы.

5. Произвести анализ матрицы парных коэффициентов и анализ матричного графика.

Пример 2.2 Матрица парных коэффициентов

	y	x1	x2
y	1		
x1	0,906518213	1	
x2	0,986942369	0,868169123	1

1) Анализ первого столбца позволяет произвести отбор факторных признаков. Факторные признаки, для которых $g_{yx_i} < 0.5$, исключают из модели. В нашем примере нет таких факторных признаков.

2) Мультиколлинеарность присутствует, так как наибольший коэффициент корреляции между независимыми переменными $g_{x_1x_2} = 0,87$, что больше 0,8.

		Влияющие (переменные)			
		Y	X1	X2	X3
Зависимые (функции)	Y	1			
	X1		1		
	X2			1	
	X3				1

Анализ произвести по алгоритму:

1. Описать расположение точек вдоль условной прямой (вдоль или хаотично).
2. Предположить форму корреляционной зависимости (линейная положительная, линейная отрицательная, отсутствует, нелинейная).
3. Оценить тесноту корреляционной связи.
4. Сделать выводы о наличии (отсутствии) линейной связи.

Следует отметить, что коэффициенты корреляции выражают *непричинную* (обусловленность одного признака другим), а *функциональную* (взаимная согласованность изменения признаков) зависимость между признаками. Различают парную (между двумя признаками) и множественную (между несколькими признаками) корреляции.

В практической деятельности, когда число коррелированных X и Y невелико ($n \leq 30$), при оценке зависимости между показателями используется следующая градация:

<i>Характеристика тесноты связи</i>	<i>Показатель тесноты связи по абсолютному значению</i>
Полная	$r=1$ или $r=-1$ - линейная связь
Высокая степень взаимосвязи	От 0.7 до 0.99 (точки расположены примерно вдоль условной прямой) или $ r \geq 0.75$
Средняя степень взаимосвязи	От 0.5 до 0.69 (имеется некоторая корреляция: точки еще расположены вдоль прямой, но хаотично) или $0.25 < r < 0.75$
Слабая степень взаимосвязи или отсутствие	От 0.2 до 0.49 (отсутствие линейной связи, возможен случай нелинейной связи) $r=0$ - отсутствие связи (независимость) или нелинейная $ r \leq 0.25$

Задания для самостоятельной работы

Задание 6

- Построить:*
 - диаграммы рассеяния,
 - регрессионную модель,
 - совмещенный график остатков, фактических и прогнозных значений зависимой переменной.
- Рассчитать:*
 - коэффициента линейной парной корреляции,
 - коэффициенты детерминации: обычный и исправленный.
- Дать:*
 - содержательную интерпретацию коэффициента регрессии, построенной модели,
 - оценку коэффициента линейной парной корреляции,
 - оценку коэффициентов детерминации: обычный и исправленный.

Вариант 1

По территориям региона приводятся данные за 199X.

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	78	133
2	82	148
3	87	134
4	79	154
5	89	162
6	106	195
7	67	139
8	88	158
9	73	152
10	87	162
11	76	152
12	115	173

Вариант 2

По 12 транспортным компаниям исследуется зависимость годового дохода (переменная Y , млн. руб.) от среднегодового количества грузовых автомобилей (переменная X). Имеются данные, для удобства, упорядоченные по фактору X :

x	y
15	235
18	250
22	247
27	287
25	260
31	262
34	307
37	280
40	357
45	410
48	389
48	311

Вариант 3

Исследуется зависимость цены акции предприятия (переменная Y , руб.) от индекса фондового рынка (переменная X , пунктов) по данным за 12 месяцев предыдущего года на первое число. Имеются временные ряды средневзвешенных за месяц значений переменных:

x	y
244	152
222	154
201	149
186	136
215	139
248	148
256	152
255	156
217	152
224	156
263	169
292	176

Вариант 4

Исследуется зависимость стоимость квартиры (переменная Y , долл.) от ее жилой площади (переменная X , кв. м.) по данным за 16 месяцев (например, первое значение за 01/01/2012). Имеются временные ряды значений переменных:

x	y
5000	30.2
5200	32
5350	32
5880	37
5430	30
5430	30
5430	30

5350 29
 5740 33
 5570 31
 5530 30
 6420 31
 7010 38
 6420 31
 7150 39
 7190 39.5

Вариант 5

Имеются статистические данные о численности населения (оценка на конец 2008 года; сотни тысяч человек) (фактор X) и об обороте розничной торговли в этих же регионах (в фактически действовавших ценах, млрд. руб.) (результативный признак Y) в некоторых регионах Российской Федерации.

№	Регион	X	Y	№	Регион	X	Y
1	Республика Адыгея	4,4	25,6	9	Ставропольский край	27,1	203,6
2	Республика Дагестан	27,1	217,3	10	Астраханская область	10,1	77,7
3	Кабардино-Балкарская Республика	8,9	48	11	Волгоградская область	26	186,1
4	Республика Калмыкия	2,8	7,4	12	Ростовская область	42,4	423,4
5	Карачаево-Черкесская Республика	4,3	22,6	13	Ивановская область	10,7	57,9
6	Республика Северная Осетия - Алания	7	37,4	14	Калужская область	10	80,7
7	Чеченская Республика	12,4	25,2	15	Костромская область	6,9	38,1
8	Краснодарский край	51,4	500,7				

Вариант 6

Исследуется зависимость расходов питания (переменная Y) от величины душевого дохода семей (переменная X) по данным, представленных по номеру группы. Имеются временные ряды значений переменных:

x y
 433 628
 616 1577
 900 2659
 1113 3701
 1305 4796
 1488 5926
 1646 7281
 1914 9350
 2411 18807

Варианты 7, 8, 9

Исследуется зависимость стоимости подержанного автомобиля ВАЗ 21XX (переменная Y) от срока эксплуатации (переменная X) по годовым данным (с 1992 года). Имеются временные ряды значений переменных.

Стоимость подержанных автомобилей, руб.			Срок эксплуатации, лет
ВАЗ 2105	ВАЗ 2107	ВАЗ 2109	
83000	99000	112000	1
86000	95000	101000	2
84000	88000	91000	3
79000	79000	82000	4
69000	82000	73000	5
66000	70000	66000	6
53000	72000	59000	7
47000	67000	53000	8
46000	59000	48000	9
44000	55000	43000	10
41000	44000	39000	11
24000	40000	35000	12
20000	32000	32000	13
19000	27000	30000	14

Задание 7

Произвести анализ матрицы парных коэффициентов корреляции.

Вариант 1

-	y	x ₁	x ₂	x ₃
y	1			
x ₁	0.802	1		
x ₂	0.823	0.644	1	
x ₃	0.487	0.246	0.872	1

Вариант 2

-	y	x ₁	x ₂	x ₃
y	1			
x ₁	0.8	1		
x ₂	0.7	0.8	1	
x ₃	0.6	0.8	0.2	1

Вариант 3

-	y	x ₁	x ₂
y	1		
x ₁	0.913	1	
x ₂	0.942	0.884	1

Вариант 4

-	y	x ₁	x ₂	x ₃
y	1			
x ₁	0.96	1		
x ₂	0.95	0.94	1	
x ₃	0.15	0.28	0.29	1

Задание 8

Вариант 1

Вычислить значение коэффициента корреляции:

x ₁	51	73	55	40	58	48	58	69	61	33
x ₂	66	69	67	58	87	54	91	95	88	55

Вариант 2

Приведен пример исходных данных измерения двух показателей интеллекта (вербального (X_i) и невербального (Y_i)) у 20 учащихся 8 класса. Рассчитать коэффициент корреляции.

Вербальный (X_i): 13, 9, 8, 9, 7, 9, 8, 13, 11, 12, 8, 9, 10, 10, 12, 10, 8, 9, 10, 11.

Невербальный (Y_i): 12, 11, 8, 12, 9, 11, 9, 13, 9, 10, 9, 8, 10, 12, 10, 10, 11, 10, 11, 13.

Вариант 3

Для каждого из 12 учащихся одного класса известно время решения текстовой арифметической задачи в секундах (X) и средний балл отметок по математике за последнюю четверть (Y). Рассчитать коэффициент корреляции и сделать вывод.

X : 122, 105, 100, 145, 130, 90, 162, 172, 120, 150, 170, 112.

Y : 4.7, 4.5, 4.4, 3.8, 3.7, 4.6, 4.0, 4.2, 4.1, 3.6, 3.5, 4.8.

Вариант 4

В эксперименте проверялась взаимосвязь между ригидностью и тревожностью человека. Предполагалось, что ригидность является одним из механизмов защиты от тревоги. В целях проверки данного предположения 15 респондентов были протестированы по опроснику ригидности и тесту на тревожность Тейлора. Результаты тестирования (сырые тестовые баллы) приведены в таблице. Проверить сформулированное предположение.

№	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	19	25	26	6	35	30	28	31	11	13	7	13	23	32	18
	28	43	34	24	42	22	45	38	21	19	18	20	14	16	18

Вариант 5

В группе специалистов проведено изучение уровней отсутствия психологического барьера при работе с компьютером (выборка А) и субъективной удовлетворенности деятельностью с применением компьютера (выборка В).

Имеется ли статистически значимая корреляционная связь между показателями, и если – да, то насколько она сильна?

№	1	2	3	4	5	6	7	8	9	10
А	10	8	6	6	13	5	12	10	3	2
В	7	9	11	3	11	7	14	11	6	1

Вариант 6

10 школьникам были даны тесты на наглядно-образное и вербальное мышление. Измерялось среднее время решения заданий теста в секундах. Исследователя интересует вопрос: существует ли взаимосвязь между временем решения этих задач? Переменная X — обозначает среднее время решения наглядно-образных, а переменная Y — среднее время решения вербальных заданий тестов.

№ испытуемых	X	Y
1	19	17
2	32	7
3	33	17
4	44	28
5	28	27
6	35	31
7	39	20
8	39	17
9	44	35
10	44	43

Вариант 7

Определите, имеется ли взаимосвязь между рождаемостью и смертностью (количество на 1000 человек) в Санкт-Петербурге:

Годы	Рождаемость	Смертность
1991	9,3	12,5
1992	7,4	13,5
1993	6,6	17,4
1994	7,1	17,2
1995	7,0	15,9
1996	6,6	14,2
1997	7,1	16
1998	8,2	13,4

Вариант 8

Имеются ежемесячные данные наблюдений за состоянием погоды и посещаемостью музеев и парков. Необходимо определить, существует ли взаимосвязь между состоянием погоды и посещаемостью музеев и парков.

Число ясных дней	Количество посетителей музея	Количество посетителей парка
8	495	132
14	503	348
20	380	643
25	305	865
20	348	743
15	465	541

Вариант 9

10 менеджеров оценивались по методике экспертных оценок психологических характеристик личности руководителя (см. Психологические тесты. Т.2. Под ред. А.А. Карелина. - М., ВЛАДОС, 1999, стр. 99). 15 экспертов производили оценку каждой

психологической характеристики по пятибалльной системе. Психолога интересует вопрос, в какой взаимосвязи находятся эти характеристики руководителя между собой.

Испытуемые п/п	тактичность	требовательность	критичность
1	70	18	36
2	60	17	29
3	70	22	40
4	46	10	12
5	58	16	31
6	69	18	32
7	32	9	13
8	62	18	35
9	46	15	30
10	62	22	36

Задание 9

Множественная регрессия

Вариант 1

Имеются временные ряды значений переменных.

y	x1	x2
7	3.9	10
7	3.9	14
7	3.7	15
7	4	16
7	4.8	17
8	5.4	19
8	4.4	20
8	5.3	20
10	6.8	20
9	6	21
11	6.4	22
12	8	28
12	5.2	29
12	8.1	30
12	8.5	31
14	9.6	32
149	9.7	36

Вариант 2

По 20 предприятиям региона изучается зависимость выработки продукции на одного работника y (тыс. руб.) от ввода в действие новых основных фондов x_1 (% от стоимости фондов на конец года) и от удельного веса рабочих высокой квалификации в общей численности x_2 (%) (p_1 – число букв в полном имени студента, p_2 – число букв в фамилии студента).

Номер предприятия	Y	x_1	x_2
1	7	$3.6 + 0.1p_1$	11
2	7	3.7	13
3	7	3.9	15
4	7	4	17
5	7	$3.8 + 0.1p_1$	18
6	7	4.8	19
7	8	5.3	19
8	8	$5.6 - 0.1p_1$	20
9	8	5.4	20
10	10	6.8	21
11	9	$6 + 0.1p_2$	21
12	11	6.4	22
13	9	6.9	22
14	11	7.2	25
15	12	$8 - 0.1p_2$	28
16	12	8.2	29
17	12	8.5	31
18	12	8.6	32
19	14	9.6	32
20	14	$9 + 0.1p_2$	36

Вариант 3

Требуется построить статистическую зависимость квартиры от трех факторов в виде множественной регрессии.

№ изм.	Общая площадь, m^2 , x_1	Жилая площадь квартиры, m^2 , x_2	Расстояние до метро, минут пешком, x_3	Стоимость квартиры, тыс. дол., y
1	80	84	3	16
2	62	37	7	22
3	69.7	42	18	24
4	97	81.2	29	19.3
5	97.3	88	8	34
6	91	64	8	25.2
7	102	66	7	27.4
8	87	83.8	10	41
9	115.3	74.7	8	38.6
10	115.6	74	8	38.6
11	90	62	8	46
12	116	81	11	38

13	107	72	11	42.3
14	93	66	18	27
15	176	129	10	78
16	96	70	8	38
17	93	73	10	23.5
18	176	110	20	68
19	74	49	18	23
20	107	73	10	48.6
21	88	61	3	34
22	74	48.3	10	24
23	74.3	81.2	10	23
24	118	77	9	36

4. Лабораторная работа №3

Тема: Оценка статистической значимости параметров регрессионного уравнения с помощью t-статистики Стьюдента, проверка гипотезы о значимости уравнения в целом с помощью статистики Фишера F, построение доверительных интервалов для коэффициентов регрессионного уравнения. Анализ адекватности регрессионных моделей.

Цель: Освоить основные принципы работы с указанными приложениями при вычислении критического значения t-статистики Стьюдента, критического значения статистики Фишера F; при проведении тестов для анализа адекватности регрессионной модели.

4.1 Задания

Задание 3.1 Скопировать исходные данные (файлы, которые созданные в Блокноте) из материалов для студентов ([\\ForStudent\КИС...](#)).

Задание 3.2. Произвести оценку статистической значимости параметров регрессионного уравнения с помощью t-статистики Стьюдента, проверку гипотезы о значимости уравнения в целом с помощью статистики Фишера F, построение доверительных интервалов для коэффициентов регрессионного уравнения. Сделать анализ адекватности регрессионных моделей.

Рассмотрим пример 2.1

4.2 Методические указания

4.2.1 t-статистики критерия Стьюдента

Осуществить проверку значимости коэффициентов уравнения регрессии.

Эмпирическое уравнение регрессии определяется на основе конечного числа статистических данных. Очевидно, что коэффициенты эмпирического уравнения регрессии являются случайными величинами, изменяющимися от выборки к выборке. При проведении статистического анализа возникает необходимость сравнения эмпирических коэффициентов регрессии β_i с некоторыми теоретически ожидаемыми значениями b_i этих коэффициентов.

Также наиболее важной проблемой на начальном этапе статистического анализа построенной модели является задача установления наличия линейной зависимости между Y и X. Эта проблема может быть решена проверкой гипотезы.

Нулевая гипотеза $H_0: b_i=0$, равенство нулю некоторого коэффициента регрессионного уравнения.

Альтернативная гипотеза $H_1: b_i \neq 0$.

Гипотеза в такой постановке обычно называется *гипотезой о статистической значимости коэффициента регрессии*. При этом если принимается нулевая гипотеза, то есть основания считать, что величина Y не зависит от X – коэффициент b_i *статистически незначим* (он слишком близок к нулю). При отклонении H_0 коэффициент считается *статистически значимым*, что указывает на наличие определённой линейной зависимости между Y и X .

Способ 1

Для проверки гипотезы используется в этом случае t – статистика, которая имеет вид: $t = \frac{b_i}{S_{b_i}}$ и при нулевой гипотезе имеет распределение Стьюдента с $(n-m-1=n-k)$ степенями свободы, где m – количество объясняющих факторов, k – количество регрессоров (параметров) включая константу, $S_{b_i} = \frac{S}{\sqrt{\sum(x_i-\bar{x})^2}}$ – стандартная ошибка коэффициента регрессии b_i , $S^2 = \frac{\sum e_i^2}{n-2}$.

Если вычисленное значение t – статистики - $|t_{факт}|$ при заданном уровне значимости p больше критического (табличного) $t_{табл}$, т.е.

$$|t_{факт}| \geq t_{табл} = t(p; n-m-1), \text{ (симметричный двухсторонний критерий)}$$

то гипотеза $H_0: b_i = 0$, отвергается в пользу альтернативной при выбранном уровне значимости. Это подтверждает статистическую значимость коэффициента регрессии b_i .

Если $|t_{факт}| < t_{табл} = t(p; n-m-1)$, то гипотеза H_0 не отвергается. Критическое значение $t_{табл} = t(p; n-m-1)$, при заданном уровне значимости p и числе степеней свободы $(n-m-1)$ находится по таблицам.

Способ 2

Необходимо *сравнить p -значение (фактическую вероятность) принятия нулевой гипотезы данного коэффициента регрессии) с выбранным уровнем значимости ϵ* . Если выполняется условие $p < \epsilon$, то нулевая гипотеза отвергается на уровне значимости ϵ , иначе нулевая гипотеза не может быть отвергнута для данного уровня значимости ϵ .

- Построить доверительные интервалы для коэффициентов уравнения регрессии.

Доверительный интервал для коэффициентов уравнения парной регрессии

$$(b_1 - t_{крит} S_{b_1}; b_1 + t_{крит} S_{b_1}), (b_0 - t_{крит} S_{b_0}; b_0 + t_{крит} S_{b_0}),$$

где b_1 , – значения коэффициентов,

S_{b_1}, S_{b_0} – их стандартные ошибки,

$t_{крит} = t_{табл}$ – критическое значение статистики Стьюдента.

4.2.2 F-статистика. Критерий Фишера

- Осуществить проверку значимости уравнения в целом.

Соответствует ли математическая модель, выражающая зависимость между переменными, имеющимся данным и достаточно ли включённых в уравнение объясняющих переменных для описания зависимой переменной.

Оценка значимости уравнения в целом дается с помощью F – критерия Фишера.

$H_0: \{b_1=b_2=\dots=0\}$, то есть тестируется одновременное равенство нулю всех коэффициентов регрессионного уравнения кроме константы регрессии.

Следовательно, фактор не оказывает влияния на результат.

Непосредственному расчету F -критерия предшествует анализ дисперсии результативного признака Y . Центральное место в нем занимает разложение общей суммы квадратов отклонений переменной y от среднего значения \bar{y} на две части – «объясненную» и «остаточную» («необъясненную»):

$\sum_i (y_i - \bar{y})^2$	=	$\sum_i (\hat{y}_i - \bar{y})^2$	+	$\sum_i (y_i - \hat{y}_i)^2$
Общая сумма квадратов отклонений		Сумма квадратов отклонений, объяснённая регрессией		Остаточная (необъяснённая) сумма квадратов отклонений

Способ 1

Если вычисленное значение F -отношения - $F_{\text{факт}}$ при заданном уровне значимости p больше критического (табличного) $F_{\text{табл}}$, т.е.

$$F_{\text{факт}} \geq F_{\text{табл}} = F(p, m, n-m-1), \text{ (односторонний правосторонний критерий)}$$

то гипотеза $H_0: \beta_1=0$ отвергается, признаётся статистическая значимость уравнения регрессии, т.е. связь между рассматриваемыми признаками есть и результаты наблюдений не противоречат предположению о её линейности.

Если $F_{\text{факт}} < F_{\text{табл}} = F(p, m, n-m-1)$, то гипотеза H_0 не отвергается, уравнение регрессии считается статистически незначимым.

Критическое значение $F_{\text{табл}} = F(p, n-m-1)$, при заданном уровне значимости p и числе степеней свободы $(m, n-m-1)$ находится по таблицам Приложения В.

Способ 2

Используется эквивалентная форма критерия. Необходимо сравнить p -значение для статистики Фишера $\text{Prob}(F\text{-statistic})$ с выбранным уровнем значимости ϵ . Если выполняется условие $p < \epsilon$, то нулевая гипотеза отвергается на уровне значимости ϵ , иначе нулевая гипотеза не может быть отвергнута для данного уровня значимости ϵ .

Оценка значимости уравнения регрессии обычно дается в виде таблицы дисперсионного анализа.

4.2.3 Оценка адекватности модели

Остатки (возмущения, отклонения наблюдаемых значений от теоретических) являются оценками случайного члена ϵ уравнения регрессии.

Исследования остатков ϵ_i предполагают проверку наличия следующих предпосылок МНК (т.е. при выполнении их получаются несмещенные эффективные и состоятельные оценки):

1. Нулевая средняя величина ϵ_i , не зависящая от x_i (математическое ожидание возмущения ϵ_i равно нулю).
2. Гомоскедастичность – дисперсия каждого ϵ_i (ошибки, возмущения) одинакова для всех наблюдений, т.е. дисперсия случайного члена в каждом наблюдении имеет только одно значение.

3. Отсутствие автокорреляции остатков. Значения остатков ε_i распределены независимо друг от друга (возмущения ε_i и $\varepsilon_j (i \neq j)$ являются независимыми).
4. Остатки подчиняются нормальному распределению.

Адекватность модели оценивается по результатам анализа остатков (ошибок, возмущений) регрессионного уравнения. Проверяется четыре условия. Если хотя бы одно из условий не выполняется, то модель не может быть признана адекватной.

4.2.4 Нормальность распределения остатков

Анализируя качество модели, необходимо проверить ряд статистических гипотез, использующих критерий Стьюдента, которым можно воспользоваться в случае, когда остатки распределены по нормальному закону.

Кривая плотности нормального распределения задается функцией

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (29)$$

где μ – математическое ожидание; σ – среднее квадратическое отклонение.

Визуально нормальность распределения остатков можно определить, сравнивая кривую плотности нормального распределения с гистограммой частот (частостей) остатков, т.е. со ступенчатой фигурой, состоящей из прямоугольников, основаниями которых служат интервалы одинаковой длины на оси Ox , а высоты равны сумме частот значений остатков, попадающих в интервал. Если линия, соединяющая середины верхних сторон прямоугольников, близка к кривой плотности нормального распределения, то предполагают, что распределение остатков приближено к нормальному.

Асимметрия и эксцесс как числовые характеристики нормально распределенной случайной величины равны 0. При асимметричном распределении вершина кривой сдвинута относительно ординаты выборочной средней. Если асимметрия больше 0, то вершина сдвинута вправо (положительная асимметрия), если меньше 0, то – влево (отрицательная асимметрия).

Эксцесс характеризует относительную остроконечность или сглаженность распределения по сравнению с нормальным распределением. Положительный эксцесс обозначает относительно остроконечное распределение, отрицательный эксцесс – относительно сглаженное распределение.

4.2.5 Гомоскедастичность (гетероскедастичность) остатков

Предположение о постоянстве и конечности дисперсии остатков называется свойством гомоскедастичности остатков. Если оно не выполняется, то такое явление называется гетероскедастичностью. Гетероскедастичность часто вызывается ошибками спецификации, когда не учитывается в модели существенная переменная. Гетероскедастичность приводит к тому, что оценки коэффициентов регрессии не являются эффективными, увеличиваются дисперсии распределений оценок коэффициентов, появляется вероятность неверного вычисления оценок стандартных ошибок коэффициентов регрессии. В результате можно сделать неверный вывод о значимости коэффициента.

4.2.6 Результаты регрессионной статистики и их оценка

После построения регрессионных моделей была получена группа результатов, которая включает в свой состав значения коэффициентов регрессии, а также статистики, на основании которых проверяется значимость влияния фактора для каждого коэффициента, включенного в модель.

Коэффициенты – значения коэффициентов;

Стандартная ошибка – стандартная ошибка коэффициентов;

t-статистика – значение статистики критерия;

P-значение – уровень значимости отклонения гипотезы равенства коэффициентов нулю;

Нижние 95% - нижняя граница доверительного интервала, в котором находится значение коэффициента генеральной совокупности;

Верхние 95% - верхняя граница доверительного интервала, в котором находится значение коэффициента генеральной совокупности.

При необходимости имеется возможность вывести таблицу стандартных и простых остатков, где для каждого значения ряда выводится предсказанное значение, с которым сопоставляется остаток, представляющий разность между прогнозным и реальным значением ряда.

Регрессионная статистика включает в свой состав:

Множественный R – коэффициент множественной корреляции;

R- квадрат – множественный коэффициент детерминации;

Нормированный R- квадрат – скорректированный коэффициент детерминации;

Стандартная ошибка – стандартная ошибка регрессии;

Наблюдения – количество наблюдений.

Дисперсионный анализ; здесь использован ряд общепринятых сокращений:

df – степени свободы (*degree of freedom*);

SS – сумма квадратов отклонений (*Sum of squares*);

MS=S – средний квадрат отклонения (*Mean square*);

F – отношение дисперсий;

Значимость F – критическое значение квантиля распределения Фишера, на котором отвергается нулевая гипотеза отсутствия влияния фактора.

4.2.7 Технологии вычислений в MS Excel

Пример 2.1 Парная регрессия

Результаты построения модели (рисунок 3.9)

Таблица 4.1 - Парная регрессия

Значение коэффициента b_1 Стандартная ошибка коэффициента $b_1 - S_{b_1} = \frac{S}{\sqrt{\sum (x_i - \bar{x})^2}}$	Значение коэффициента b_0 Стандартная ошибка коэффициента $b_0 - S_{b_0} = \frac{S \sqrt{\sum x_i^2}}{\sqrt{n \sum (x_i - \bar{x})^2}}$
Коэффициент детерминации R^2 $R^2 = 1 - \frac{SS_{ост}}{SS_{общ}} = \frac{SS_R}{SS_{общ}}$	Стандартная ошибка регрессии – $S^2 = \frac{\sum e_i^2}{n - 2}$
F - статистика $F = \frac{MS_R}{MS_{ост}} = \frac{\sum (\hat{y} - \bar{y})^2 / 1}{\sum (y - \hat{y})^2 / (n - 2)}$	Число степеней свободы $df = n - m - 1$, m – количество независимых переменных
Регрессионная сумма квадратов – $SS_R = \sum (\hat{y}_i - \bar{y})^2$	Остаточная сумма квадратов – $SS_{ост} = \sum (y_i - \hat{y}_i)^2$

Получили следующие результаты для рассмотренного выше примера. Откуда выписываем, округляя до 4 знаков после запятой и переходим к обозначениям:

Уравнение регрессии:

$$\widehat{y}_x = 0,2593x - 73,247$$

r-значения: (0,0000) (0,0002)

t-статистика: (25,9317) (-4,5682)

Коэффициент корреляции:

$$r_{xy} = 0,9869.$$

Коэффициент детерминации:

$$R^2 = 0,9739.$$

Нормированный (исправленный, скорректированный) коэффициент детерминации:

$$R^2_{норм} = 0,9725$$

$$F_{набл} = F_{факт} = 672,4519$$

p-значение = 0,0000

Количество наблюдений $n=20$, количество независимых переменных $m=1$

Пример 2.1 Парная регрессия.

Расчет статистик Стьюдента и Фишера

Для нахождения табличного (критического) значения критерия Стьюдента определяют число степеней свободы df , которое определяется по формуле $n - m - 1$, и находят его значение при определенном уровне значимости (например, 0,10; 0,05; 0,01).

Таблица 4.2 - Расчет критического значения t-статистики Стьюдента

$t_{табл} = t(\varepsilon; n-m-1) = (20-1-1) = t(18)$	Значение
=СТЮДРАСПОБР(0,05;18)	2,1009
=СТЮДЕНТ.ОБР.2X(0,05;18)	2,1009

Расчет фактического и критического значения

$$F = \frac{R^2}{1-R^2} \cdot \frac{(n-m-1)}{m}$$

n – число наблюдений
m – число факторов

Рисунок 4.1 - Расчет фактического значение критерия Фишера

Табличное (критическое) значение критерия Фишера вычисляют следующим образом:

Таблица 4.3 - Расчет критического значения статистики Фишера F

$F_{табл} = F_{крит} = F(\varepsilon; m, n-m-1)$	Значение
=ФРАСПОБР(p;k1;k2)	
=F.ОБР.ПХ(p;k1;k2)	
=ФРАСПОБР(0,05; 1; 20-1-1)	4,413873419
=F.ОБР.ПХ(0,05; 1;20-1-1)	4,413873419

1. $p = \varepsilon = 0,05$.
2. Определяют $k1$, которое равно количеству факторов (x). Например, в однофакторной модели (модели парной регрессии) $k1 = 1$, в двухфакторной $k1 = 2$.
3. Определяют $k2$ по формуле $n - m - 1$, где n - число наблюдений, m - количество факторов. Например, в однофакторной модели $k1 = n - 2$.

Дисперсионный анализ

	df	
Регрессия	1	k1
Остаток	18	k2
Итого	19	

Построение доверительного интервала (самостоятельно, используя формулы).

Оценка адекватности модели (парная регрессия)

1. Случайные ошибки должны быть в совокупности гауссовскими, то есть должны подчиняться нормальному закону распределения – условие адекватности.

Нормальность распределения остатков: *устанавливается визуально по гистограмме, по асимметрии и эксцессу, с помощью критериев согласия*.

Оценки вида кривой Гаусса и значений асимметрии и эксцесса являются качественными характеристиками распределения. Для надежности вывода (с

вероятностью 0,95) произведем проверку статистической гипотезы о нормальности распределения с помощью критерия согласия хи-квадрат-Пирсона. (<http://helpiks.org/1-3516.html>)

Нулевая гипотеза: остатки подчиняются нормальному закону распределения.

Для проверки гипотезы о нормальности распределения остатков с помощью критерия хи-квадрат Пирсона необходимо сравнить наблюдаемое и критическое значения статистики хи-квадрат.

$$\chi^2_{\text{набл}} \quad \chi^2 = \sum_{i=1}^k \frac{(m_i - p_i)^2}{np_i} \quad (30)$$

где n – объем выборки;
 k – число интервалов,
 m_i – число появления значений вариант, принадлежащих i -му интервалу, $i = 0, 1, \dots, k - 1$;
 p_i – вероятности попадания значений случайной величины в каждый из этих интервалов.

$$\chi^2_{\text{крит}} \quad \chi^2(\varepsilon; k - r - 1)$$

где r – число параметров распределения $F(x)$ (для нормального распределения $r=2$, так как оцениваются два параметра: выборочное среднее квадратическое отклонение и среднее значение), которые оцениваются по выборке,
 k – число непересекающихся интервалов наблюдаемых значений (вариант),
 ε – выбранный уровень значимости.

Гипотеза о нормальном распределении случайного члена принимается на заданном уровне значимости, если $\chi^2_{\text{набл}} < \chi^2_{\text{крит}}$. Если же $\chi^2_{\text{набл}} \geq \chi^2_{\text{крит}}$, то гипотеза отвергается.

Число параметров для основных видов распределений:

Распределение	Параметры	Количество
Нормальное	μ, σ	2
Пуассона	λ	1
Биноминальное	p	1
Показательное	λ	1

Для проверки гипотезы о принадлежности генеральной совокупности нормальному виду распределений необходимо строить группированный статистический ряд, т.к. нормальное распределение является непрерывным. Для этого нужно знать размах выборки, который равен разнице между максимальным и минимальным элементами выборки. Кроме того, нужно рассчитать точечные оценки математического ожидания и среднее квадратическое отклонения (СКО).

- Скопируйте значения ячеек M53:M73 (значения остатков и названием «Остатки») с текущего листа на новый лист в ячейки C12:C32 (C12 находится название столбца).

- Вычислите основные статистические характеристики остатков.

Воспользуйтесь инструментом *Описательная статистика* Пакета анализа данных. Значения параметров в диалоговом окне установите следующим образом:

- *Входной интервал* – введите ссылки на ячейки C13:C32.
- *Группирование* – установите флажок *по столбцам*.
- *Метки в первой строке* – не устанавливайте.

- *Параметры вывода* – установите переключатель на *Выходной интервал*, курсор установите в поле напротив, выделите ячейку A12.
- Установите флажки *Итоговая статистика*, *Уровень надежности (95%)*. Нажмите кнопку *OK*.

- Постройте таблицу в соответствии с рисунком:

	A	B	C	D	E	F	G	H	I	J
12	<i>Столбец1</i>		<i>Остатки</i>		Параметры	Интервалы	Частота	Вероятность	Теоретическая частот	Критерий
13			3,948513094		Объем					
14	Среднее	-1,7E-14	10,61162894		20					
15	Стандартная ошибка	3,055616	15,05370617		Максимум					
16	Медиана	-1,45539	-5,690190857		26,50954283					
17	Мода	#Н/Д	-3,841100755		Минимум					
18	Стандартное отклонение	13,66513	-29,39902352		-29,39902352					
19	Дисперсия выборки	186,7358	-9,770972037		Среднее					
20	Эксцесс	0,076676	-17,92188194		-1,7053E-14					
21	Асимметричность	-0,29745	-5,70084332		СКО					
22	Интервал	55,90857	13,55525965		13,66513133					
23	Минимум	-29,399	-1,630714603		Эксцесс					
24	Максимум	26,50954	10,59032401		0,076675886					
25	Сумма	-3,4E-13	12,43941411		Асситричность					
26	Счет	20	11,69551709		-0,297452462					
27	Уровень надежности(95,0%)	6,395478	26,50954283							
28			-6,45539281							

Рисунок 4.2 – Статистический ряд

- Определите количество интервалов по формуле Стерджесса: $=1+3,322*\log_{10}(20)$. Например, значение, вычисленное по формуле Стерджесса, приблизительно равно 5,32. Значит, количество интервалов можно взять равным $k=5$.
- Определите ширину каждого интервала: $=(26,509542283-(-29,39902352))/5$. Получится значение равное 11,18171. Следовательно, ближайшее целое к полученному – это число 11, $d=11$.
- Произведите расчет массива правых границ интервалов (значений вариант) и абсолютных частот.

- Найти правые границы интервалов группировки (для каждого из 5 интервалов). Для этого в соответствующей ячейке вычислить верхнюю границу первого интервала, введя формулу $=E18+11$ (минимальное в выборке)+(длина интервала); в следующей ячейке - верхнюю границу второго интервала, введя формулу $=(\text{верхняя граница первого интервала})+(\text{длина интервала})$. $=F13+11$

- Скопировать содержимое ячейки (содержащей верхнюю границу второго интервала) в диапазон из трех ячеек. Значение верхней границы последнего интервала равно вычисленному ранее максимальному значению в выборке. В ячейке F17 указать максимальное значение (E16).

Интервалы	-18,399	-7,399023524	3,600976476	14,600976	26,50954
-----------	---------	--------------	-------------	-----------	----------

- Для каждого интервала группировки найти число элементов выборки (частоты наблюдения), попавших в интервал. Важно, чтобы каждый элемент выборки был отнесен к одному и только к одному интервалу, а если значение элемента попадает на границу интервала, то будем относить его к интервалу с младшим номером. Минимальный элемент всегда относим к первому интервалу, максимальный - к последнему. Убедиться, что сумма всех частот наблюдения (абсолютных частот) равна объему выборки ($n=20$).

$=\text{ЧАСТОТА}(C13:C32;F13:F16)$

Онлайн калькулятор: (<http://planetcalc.ru/936/>)

Интервал	Число наблюдений	Процент
-29.40 - -18.22	2	9.52
-18.22 - -7.04	2	9.52
-7.04 - 4.15	10	47.62
4.15 - 15.33	6	28.57
15.33 - 26.51	0	0.00

- *Постройте гистограмму:*

Воспользоваться инструментом *Гистограмма* Пакета анализа данных. Появится окно *Гистограмма*. Значения параметров в появившемся диалоговом окне установите следующим образом:

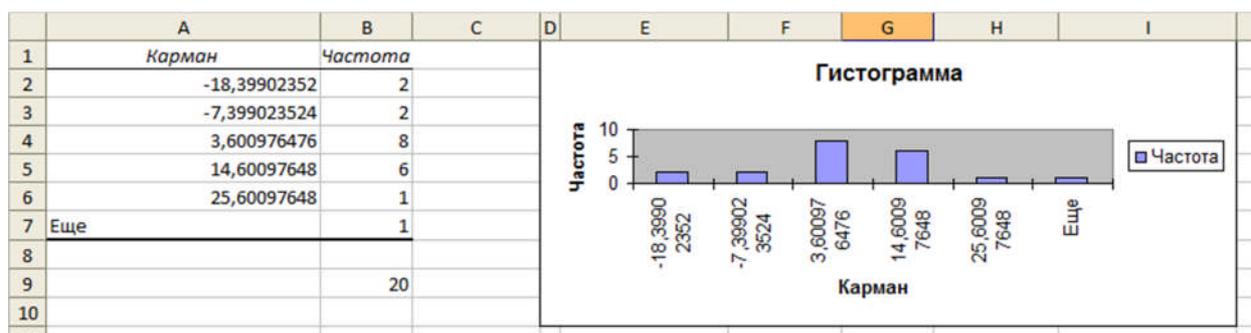


Рисунок 4.3 - Гистограмма

- *Входной интервал* – введите ссылки на ячейки C13:C32.
 - *Интервал карманов* – F13:F17.
 - *Метки* – не устанавливайте.
 - *Параметры вывода* – установите переключатель на *Выходной интервал*, курсор установите в поле напротив, выделите ячейку A1.
 - *Парето* – оставьте пустым.
 - *Интегральный процент* – оставьте пустым.
 - *Вывод графика* – установите флажок. Нажмите кнопку *ОК*. Растяните гистограмму вниз. При необходимости ее можно перемещать.
- Для расчета теоретической вероятности $p_i = F(b_i) - F(a_i)$ вычислите разницу между функциями нормального распределения (функция НОРМРАСП категории «Статистические») с параметрами: «X» – значение границы интервала, «Среднее» - ссылка на ячейку E20, «Стандартное_откл» - ссылка на E22, «Интегральная» - 1.

В ячейку H12 введите значение 0 – значение функции распределения.

В ячейки H13:P17 введите формулу массива:

- 1) выделить ячейки H13:P17;
- 2) нажать функциональную F2;

3) ввести формулу =НОРМРАСП(F13:F17;E20;E22;1);

4) нажать комбинацию клавиш Ctrl+Shift+Enter.

Если появилось только одно значение, то нажать F2 и снова Ctrl+Shift+Enter. В строке формул при активизации любой ячейки диапазона H13:P17 появится формула в фигурных скобках: {=НОРМРАСП(F13:F17;E20;E22;1)}.

В ячейки I13:I17 введите формулу массива {= I14:I17 – I13:I6} (обратите внимание: во второй массив ячейку I17 не включать).

Для расчета теоретической частоты (np_i) ввести в ячейку I13 формулу: =H13*\$E\$14, выполнить распространение на диапазон ячеек I14:I17.

- Произвести расчет элементов суммы $\sum_{i=1}^k \frac{(m_i - p_i)^2}{np_i}$ критерия согласия Пирсона: ввести формулу в ячейку J13 =(H13-I13)^2/I13 и выполнить распространение ее на диапазон J14:J17.

Найдите фактическое (наблюдаемое) значение статистики критерия хи-квадрат и критическое значение статистики критерия хи-квадрат. Для этого вводим в F19 «Сумма», а в J21 - подпись «Критич.», J22 - подпись «Набл.». Ввести в соседние ячейки формулы – в G19: =СУММ(G13:G17), в I19: =СУММ(I13:I17), J19: =СУММ(J13:J17), K19: =СУММ(K13:K17), а в K20: =ХИ2ОБР(0,05;5-2-1), здесь первый параметр – уровень значимости равен 0,05 взят из условия, а степень свободы (k-г-1)=(5-2-1)=3, так как k=5 – число интервалов группировки, а г=2, т.к. были оценены два параметра нормального распределения: математическое ожидание и среднеквадратическое отклонение (СКО).

Так как критическое значение статистики Пирсона больше наблюдаемого значения статистики Пирсона, то нулевая гипотеза не отклоняется.

Остатки имеют нормальное распределение. Условие адекватности выполняется.

	A	B	C	D	E	F	G	H	I	J	K
11							Частоты наблюдений, m_i	Вероятности, p_i		Теоретические частоты, np_i	
12	Столбец1		Остатки		Параметры	Интервалы	Частота	0	Вероятность	Теоретическая частота	Критерий
13			3,948513094		Объем	-18,39902	2	0,089083347	0,089083347	1,781666941	0,026755
14	Среднее	-1,71E-14	10,61162894		20	-7,399024	2	0,294097764	0,205014417	4,100288333	1,075829
15	Стандартная ошибка	3,0556163	15,05370617		Максимум	3,6009765	8	0,603923433	0,309825669	6,196513387	0,524902
16	Медиана	-1,455393	-5,690190857		26,50954283	14,600976	6	0,857348916	0,253425483	5,068509656	0,171189
17	Мода	#Н/Д	-3,841100755		Минимум	26,509543	2	0,973806544	0,116457628	2,32915257	0,046515
18	Стандартное отклонение	13,665131	-29,39902352		-29,39902352						
19	Дисперсия выборки	186,73581	-9,770972037		Среднее	Сумма	20		0,973806544	19,47613089	1,845192
20	Экссесс	0,0766759	-17,92188194		-1,7053E-14						
21	Асимметричность	-0,297452	-5,70084332		СКО					Критич.	5,9914645
22	Интервал	55,908566	13,55525965		13,66513133					Набл.	1,8451918
23	Минимум	-29,39902	-1,630714603		Экссесс						
24	Максимум	26,509543	10,59032401		0,076675886						
25	Сумма	-3,41E-13	12,43941411		Асситричность						
26	Счет	20	11,69551709		-0,297452462						
27	Уровень надежности(95,0%)	6,3954783	26,50954283		Размах						
28			-6,45539281		55,90856636						
29			-22,01331558		Число групп						
30			-2,757212605		5	5,32192809	5,189316586				
31			2,056813138		Ширина						
32			-1,280071016		11						

Рисунок 4.3 – Результаты вычислений

Источники:

Эконометрика и экономико-математические методы и модели : пособие для студентов экономических специальностей / авт.-сост. : Л. П. Авдашкова [и др.]. – Гомель : учреждение образования «Белорусский торгово-экономический университет потребительской кооперации», 2012. – 116 с. ISBN 978-985-461-946-0

Блинов Ю.Ф., Иванцов В.В., Серба П.В. Методы математического моделирования. Ч.2. Электронное учебное пособие. Таганрог, ТТИ ЮФУ, 2012. – 47 с.

Мамонтов, Г.Я. Анализ данных в MS Excel и OO Calc [Текст] : учебно-методическое пособие / Г.Я. Мамонтов, И.А. Икон-Томск : Изд-во Том. гос. архит.-строит. ун-та, 2010. —никова. ISBN 978-5-93057-376-3.–60 с.

Валеев С. Г. Практикум по прикладной статистике : учебное пособие / С. Г. Валеев, В. Н. Клячкин. – Ульяновск: УлГТУ, 2008. – 129 с.: ил. ISBN 978-5-9795-0318-9

<http://www.moiseev.su/tasks/primExcel.pdf>

http://www.sibstrin.ru/files/kis/Voskoboinikov_Balanchuk_TV_i_MS_2013.pdf

2. Оценить математическое ожидание остатков (Mean). Условие адекватности – математическое ожидание остатков равно нулю (симметричность).

Среднее	-1,7E-14
---------	----------

Сформулировать вывод.

3. Оценить постоянство дисперсии остатков. Условие адекватности – дисперсия остатков должна быть постоянной (гомоскедастичность).

Гомоскедастичность – дисперсия каждого ε_i (ошибки) одинакова для всех наблюдений, т.е. дисперсия случайного члена в каждом наблюдении имеет только одно значение.

Для оценки нарушения гомоскедастичности наиболее часто используются графический анализ отклонений, тест ранговой корреляции Спирмена и тест Голдфелда–Квандта.

где D_i – разность между рангом x_i и рангом модуля остатка $e_i = y_i - \hat{y}_i$.

Тест проводится по следующей схеме:

1. Строится линейная модель регрессии.

2. Определяются ранги значений x_i независимой переменной и соответствующие ранги модулей остатков $|e_i|$ (ранг – это порядковый номер значения в ранжированном ряду).

3. Находится коэффициент ранговой корреляции Спирмена.

4. Осуществляется проверка гипотезы об отсутствии гетероскедастичности с помощью t -статистики, наблюдаемое значение которой определяется равенством.

Если $t_{набл} > t_{крит}$, то гетероскедастичность присутствует, значит, МНК-оценки неэффективны.

Так как $M(\varepsilon)$ равно нулю, то МНК-оценки параметров являются несмещенными и состоятельными, поэтому их позволительно использовать, например, для точечного прогнозирования даже в случае гетероскедастичности. Однако в этом случае МНК-оценки не являются эффективными, а, следовательно, результаты (доверительные интервалы для коэффициентов и прогнозных значений), основанные на анализе дисперсии, неверны.

При применении теста Спирмена предполагается, что абсолютные величины остатков и значения объясняющей переменной коррелированы. Эту корреляцию можно измерять с помощью коэффициента ранговой корреляции Спирмена:

$$r = 1 - \frac{6 \sum D_i^2}{n(n^2 - 1)} \quad (31)$$

Для проверки гипотезы о гомоскедастичности случайной переменной сравниваются наблюдаемое и критическое значения t -статистики.

Найдите наблюдаемое значение t -статистики, где r - коэффициент ранговой корреляции Спирмена; D_i – разность между рангом x_i и рангом модуля остатка e_i .

- На лист с исходными данными скопируйте значения независимой переменной (y) в столбец А (начиная с ячейки А1) нового листа и назовите лист «Условие 3». В столбец В (начиная с ячейки В1) скопируйте значения остатков. Копировать вместе с названиями столбцов. В ячейку С1 введите название «Модуль ост». В ячейки С2:С21 введите формулу массива $\{=ABS(B2:B21)\}$
- Воспользуйтесь инструментом *Ранг и перцентиль* Пакета анализа данных. Значения параметров в диалоговом окне установите следующим образом:
Входной интервал – введите ссылки на ячейки А2:А21; *Метки в первой строке* – не устанавливайте; *Выходной интервал* – установите курсор в поле и выделите ячейку D1. Нажмите кнопку *OK*
- Воспользуйтесь инструментом *Ранг и перцентиль* Пакета анализа данных. Значения параметров в диалоговом окне установите следующим образом:
Входной интервал – введите ссылки на ячейки С2:С21; *Метки в первой строке* – не устанавливайте; *Выходной интервал* – установите курсор в поле и выделите ячейку H1. Нажмите кнопку *OK*
- Выделите ячейки D2:G21 и нажмите кнопку *Сортировка по возрастанию (от минимального к максимальному)* на вкладке Данные
- Выделите ячейки H2:K21 и нажмите кнопку *Сортировка по возрастанию*
- Выделите ячейки L2:L21 и введите формулу массива $\{=(F2:F21 - J2:J21)^2\}$. В ячейку L1 введите название «Квадрат разности рангов»
- В ячейку K22 введите название «Коэффициент ранговой корреляции». В ячейку L22 введите формулу $= 1 - 6 * СУММ(L2: L21) / (20 * (20^2 - 1))$
- В ячейку K23 введите обозначение $t_{набл}$
- В ячейку L23 введите формулу для вычисления $t_{набл}$
 $= L22 * КОРЕНЬ(20 - 1)$.

Найдите критическое значение t -статистики.

- В ячейку K24 введите обозначение $t_{крит}$
- В ячейку L24 введите формулу для вычисления $t_{крит}$
 $= СТЬЮДРАСПОБР(0,05;20 - 2)$.

	К	Л
22	Коэффициент ранговой корреляции	-0,1699248
23	$t_{набл}$	-0,7406851
24	$t_{крит}$	2,09302405

Рисунок 4.4 – t -статистика

4. Оценить наличие автокорреляции остатков. Условие адекватности – остатки должны быть взаимно независимыми (некоррелированность). Проводится тест Дарбина-Уотсона.

Для проверки гипотезы об отсутствии автокорреляции случайной переменной сравниваются наблюдаемое и критические значения статистики Дарбина-Уотсона.

Найдите наблюдаемое значение статистики d

$$d = \frac{\sum_{t=2}^n (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^n \varepsilon_t^2}$$

используя в качестве оценок значений случайной переменной соответствующие значения остатков, выполнив действия, приведенные ниже.

- На листе, где произведено построение регрессии, объедините любые две ячейки (например, D23 и E23) и введите название «Условие 4».

В ячейку D24 введите название «Числитель $d_{набл}$ ».

В ячейки D25:D43 введите формулу массива $\{=(C25:C43 - C26:C44)^2\}$.

В ячейку C45 введите слово *Суммы*.

В ячейку D45 введите формулу = СУММ(D25:D43).

В ячейку E24 введите название «Знаменатель $d_{набл}$ ».

В ячейки E25:E44 введите формулу массива $\{=(C25:C44)^2\}$.

В ячейку E45 введите формулу = СУММ(E25:E44).

В ячейку D47 введите название « $d_{набл}$ ».

В ячейку E47 введите формулу для вычисления $d_{набл} = D45/E45$.

	A	B	C	D	E
23				Условие 4	
24	<i>Наблюдение</i>	<i>Предсказанное Y</i>	<i>Остатки</i>	<i>Числитель $d_{набл}$</i>	<i>Знаменатель $d_{набл}$</i>
25	1	186,0514869	3,94851309	44,39711277	15,59075565
26	2	209,3883711	10,6116289	19,73205012	112,6066687
27	3	224,9462938	15,0537062	430,3092638	226,6140694
28	4	245,6901909	-5,6901909	3,419134206	32,37827198
29	5	263,8411008	-3,8411008	653,2074163	14,75405501
30	6	279,3990235	-29,399024	385,2604052	864,3025842
31	7	289,770972	-9,770972	66,43733216	95,47189455
32	8	307,9218819	-17,921882	149,3537848	321,1938521
33	9	315,7008433	-5,7008433	370,7975017	32,49961456
34	10	336,4447403	13,5552597	230,6138141	183,7450643
35	11	341,6307146	-1,6307146	149,3537848	2,659230116
36	12	349,409676	10,590324	3,419134206	112,1549627
37	13	367,5605859	12,4394141	0,553382786	154,7390235
38	14	388,3044829	11,6955171	219,4553587	136,78512
39	15	393,4904572	26,5095428	1086,686982	702,7558611
40	16	406,4553928	-6,4553928	242,0489609	41,67209633
41	17	422,0133156	-22,013316	370,7975017	484,5860628
42	18	442,7572126	-2,7572126	23,17484386	7,602221351
43	19	447,9431869	2,05681314	11,13479586	4,230480285
44	20	471,280071	-1,280071	#Н/Д	1,638581807
45			<i>Суммы</i>	4460,15256	3546,341889
46					
47				<i>$d_{набл}$</i>	1,257676981

Рисунок 4.5 – Статистика Дарбина-Уотсона

4.2.8 Технологии вычислений в EViews

Пример 2.1. Парная регрессия.

Расчет статистик

- Обратите внимание, что фактические значения t -статистики Стьюдента указаны в колонке t -Statistic для коэффициентов $C(1)=a$ и $C(2)=b$
- Определить критическое значение t -статистики Стьюдента для выбранного уровня значимости ϵ можно с использованием встроенной функции @qtdist
- Ввести команду show @qtdist(v,p) в окне ввода команд, $v=1-\epsilon/2$, $p=n-k$ – количество степеней свободы ($k=m+1$ - количество параметров или регрессоров)

Команда show показывает на экран значение функции @qtdist.

Количество наблюдений n равно 20, количество регрессоров k равно 2, $v=0.95$, количество степеней свободы $p=20-(2+1)=18$. Критическое значение t -статистики Стьюдента равно @qtdist(0.975,18)=2.100922040= $t_{кр.}$ (show @qtdist(0.975,18))

- Фактическое значение статистики Фишера указано внизу справа объекта Equation
- Определить критическое значение статистики Фишера F для выбранного уровня значимости ϵ можно с использованием встроенной функции @qfdist

Количество наблюдений n равно 20, количество регрессоров k равно 2. Введем в окне ввода команд строку show @qfdist(v,p1,p2), где $v=1-\epsilon$, $p1=k-1$, $p2= n-k$.

Команда show показывает на экране значение функции @qfdist. $F_{крит}=4,41387341917056$.
show @qfdist(1-0.05,2-1,20-2).

Произвести оценку адекватности модели

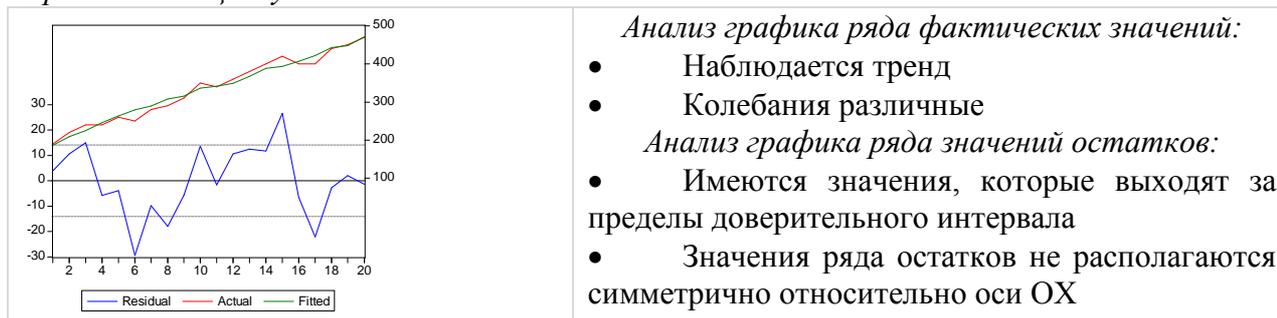


Рисунок 4.6 – Совмещенный график

1. Оценить математическое ожидание остатков (Mean). Условие адекватности – математическое ожидание остатков равно нулю (симметричность).

Тест проводится в окне объекта Equation последовательным выполнением команд View / Residual Tests / Histogram / Normality Test.

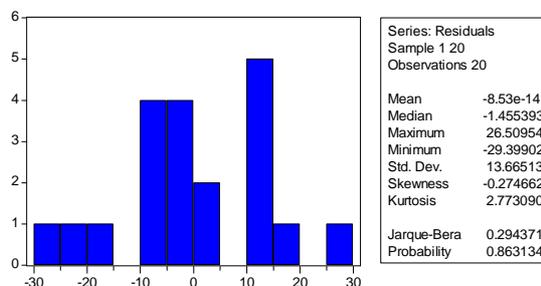


Рисунок 4.7 – Гистограмма остатков

2. Оценить наличие автокорреляции остатков. Условие адекватности – остатки должны быть взаимно независимыми (некоррелированность). Проводится три теста:

- *Тест множителей Лагранжа (Breusch–Godfrey LM test)*
- *Q-статистики Льюнга–Бокса и визуальный анализ графиков остатков, а также ВАКФ и ВЧАКФ*
- *Статистика и тест Дарбина–Уотсона.*

• Тест множителей Лагранжа (*Breusch–Godfrey LM test*). Формулировка нулевой гипотезы: *Остатки являются взаимно независимыми*. Тест проводится в окне объекта Equation последовательным выполнением команд *View / Residual Tests / Serial Correlation LM Test...* Далее указывается, сколько лагов включить для оценивания. Для принятия решения использовать *эквивалентную форму критерия*, который представляет собой сравнение уровня значимости ε и р-значения, которое указывается в строке *Probability* в окне оценивания теста.

Breusch-Godfrey Serial Correlation LM Test:			
F-statistic	1.271088	Probability	0.307379
Obs*R-squared	2.742047	Probability	0.253847

Рисунок 4.8 – Результаты теста множителей Лагранжа

Probability =0,253847.

• Q-статистики *Льюнга–Бокса* и визуальный анализ графиков остатков, а также ВАКФ и ВЧАКФ. Формулировка нулевой гипотезы: *Остатки являются взаимно независимыми*. Тест проводится в окне объекта Equation последовательным выполнением команд *View / Residual Tests / Correlogram / Q-statistics...* Далее указывается, сколько лагов включить для оценивания. Для принятия решения использовать *эквивалентную форму критерия*, которая представляет собой сравнение уровня значимости ε и р-значения, указанного в колонке *Prob* в окне оценивания теста.

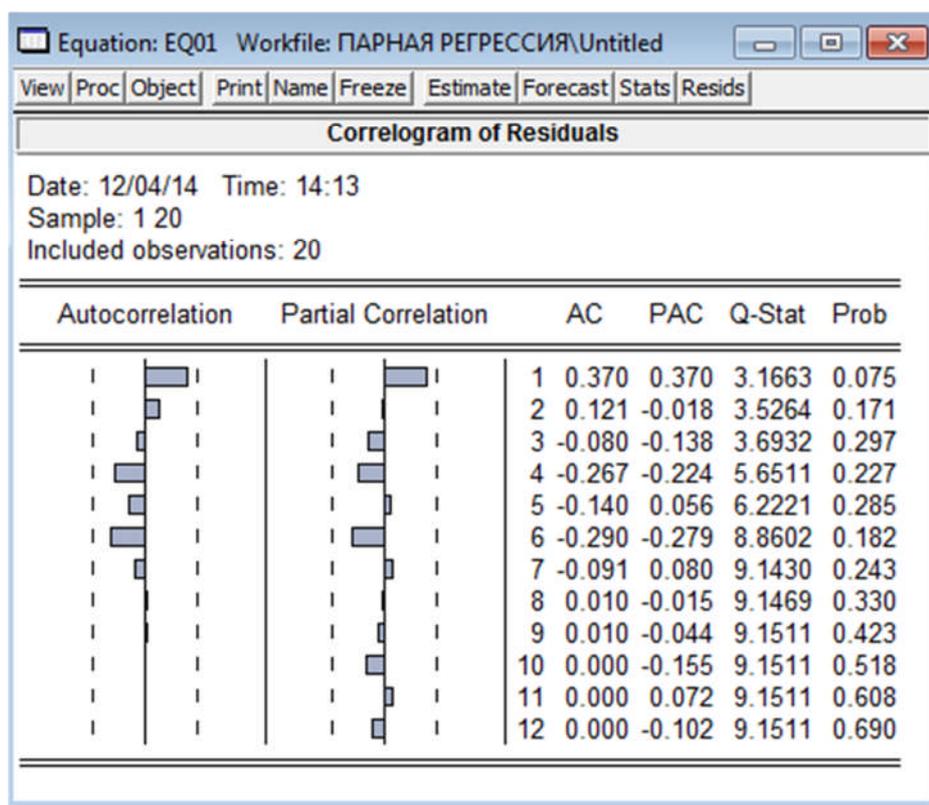


Рисунок 4.9 – Результаты теста Q-статистики Льюнга–Бокса

1 и 2 столбцы содержат информацию об автокорреляции,
3 и 4 – частной автокорреляции.

Все Prob > 0.05 для Q-статистики указывают на принятие нулевой гипотезы. Значит, имеются основания считать, что автокорреляция остатков отсутствует. Значения гистограммы лагов (ВАКФ и ВЧАКФ) не выходят за рамки доверительного интервала.

- Статистика и тест *Дарбина–Уотсона*. Необходимо в меню объекта Equation выполнить пункт *View / Estimation Output*, где будет вычислена фактическая статистика *Durbin-Watson stat*. Решение об отсутствии автокорреляции принимается путем ее сравнения с критическими значениями (d_l и d_u), которые берутся из статистических таблиц критических точек статистики Дарбина-Уотсона (смотри приложение А). Если фактическое значение статистики попадает в зону между d_u и $(4-d_u)$, то говорят, что автокорреляция 1 порядка отсутствует. Если попадает в зону между 0 и d_l , то имеет место положительная автокорреляция, если в зону между $(4-d_l)$ и 4, то отрицательная автокорреляция. В остальных случаях с помощью теста нельзя ни принять, ни отклонить гипотезу об отсутствии автокорреляции.

Проверяется гипотеза

H_0 : случайные ошибки не коррелированы.

- Открыть объект *Equation*.

- View/ Estimation Output.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X	0.259299	0.009999	25.93168	0.0000
C	-73.24723	16.03423	-4.568177	0.0002

R-squared	0.973930	Mean dependent var	334.5000
Adjusted R-squared	0.972482	S.D. dependent var	84.63389
S.E. of regression	14.03959	Akaike info criterion	8.216279
Sum squared resid	3547.980	Schwarz criterion	8.315852
Log likelihood	-80.16279	F-statistic	672.4519
Durbin-Watson stat	1.257096	Prob(F-statistic)	0.000000

- Найти критические значения статистик Дарбина-Уотсона для заданного числа наблюдений (20), числа объясняющих переменных (1) и заданного уровня значимости 0,05. В зависимости от того, в какую область попадает наблюдаемое значение статистики Дарбина-Уотсона, принимают или отвергают нулевую гипотезу.

Однако существуют области неопределенности, связанные с тем, что распределение статистики Дарбина-Уотсона зависит не только от числа наблюдений и числа объясняющих переменных, но и от значений объясняющих переменных. В этом случае используют другие способы проверки.

Если автокорреляция отсутствует, то значение статистики $d \approx 2$. При положительной автокорреляции $0 \leq d < 2$, а при отрицательной $2 < d \leq 4$. Следовательно, $0 \leq d \leq 4$.

Визуальный способ определения наличия автокорреляции остатков – построение графиков зависимостей остатков от номера наблюдений (смотри стр. 13 - 15).

Примечание. Выводы об отсутствии автокорреляции делаются по результатам трех тестов.

Сделать выводы об отсутствии автокорреляции.

3. Оценить постоянство дисперсии остатков. Условие адекватности – дисперсия остатков должна быть постоянной (гомоскедастичность).

Гомоскедастичность: дисперсия случайных ошибок постоянна для всех наблюдений.

Гетероскедастичность: дисперсия случайных ошибок не обязательно одинакова для всех наблюдений.

Для оценки постоянства дисперсии остатков проводится тест Уайта (White Heteroscedasticity test). Формулировка нулевой гипотезы: Дисперсия остатков

является постоянной. Альтернативная гипотеза: дисперсия случайных ошибок в модели непостоянна для всех наблюдений.

Тест проводится в окне объекта Equation последовательным выполнением команд View / Residual Tests / White Heteroskedasticity (no cross terms). Для принятия решения использовать эквивалентную форму критерия, который представляет собой сравнение уровня значимости ϵ и р-значения, которое указывается в строке *Probability* в окне оценивания теста.

White Heteroskedasticity Test:			
F-statistic	0.650591	Probability	0.534248
Obs*R-squared	1.421965	Probability	0.491161

Рисунок 4.4 – Окно с результатами теста Уайта

	Std. Error	t-Statistic	Prob.
	0.009999	25.93168	0.0000
	16.02422	1.568177	0.0002

Probability=0,491161

Сделать вывод самостоятельно (проверка гипотезы о гомоскедастичности остатков).

4. Оценить случайные ошибки: они должны быть в совокупности гауссовскими, то есть должны подчиняться нормальному закону распределения.

Предварительно необходимо сделать копию служебного ряда resid, введя в окне ввода команд команду *series r1=resid*. После нажатия клавиши *Enter* в окне Workfile появится новый временной ряд – ряд остатков *r1*, который необходимо протестировать и сделать выводы.

По гистограмме определить, насколько остатки нормальны. Критерий Жака-Бера используется для проверки гипотезы о том, что исследуемая выборка является нормально распределенной. Для этого смотрим статистику Jarque-Bera. Значение Skewness должно стремиться к 0. Значение Kurtosis должно стремиться к 3.

Критерий Жака-Бера

Проверяется гипотеза H_0 : ряд остатков имеет нормальное распределение.

Решение принимается на основе эквивалентной формы критерия.

Таблица 4.3 - Характеристики закона распределения

Ассиметрия (<i>Skewness</i>)	характеризует степень смещения вариационного ряда относительно среднего значения по величине и направлению.	В симметричной кривой коэффициент асимметрии равен нулю. Если правая ветвь кривой, начиная от вершины, больше левой (<i>правосторонняя асимметрия</i>), то коэффициент асимметрии больше нуля. Если левая ветвь кривой больше правой (<i>левосторонняя асимметрия</i>), то коэффициент асимметрии меньше нуля. Асимметрия менее 0,5 считается малой.
Эксцесс (<i>Kurtosis</i>)	характеризует крутизну кривой распределения - ее заостренность или пологость по сравнению с нормальной кривой.	Коэффициент эксцесса нормального распределения равен 3. Коэффициент островершинного распределения превышает 3, а плосковершинного – меньше 3.

Таблица 4.4 - Descriptive Statistics – Stats Table (характеристики остатков)

Mean	-8.53E-14	Среднее
Median	-1.455393	Медиана
Maximum	26.50954	Максимум
Minimum	-29.39902	Минимум
Std.Dev.	13.66513	Стандартное отклонение (корень из дисперсии)
Skewness	-0.274662	Асимметрия (0 у нормального распределения)
Kurtosis	2.773090	Эксцесс (3 у нормального распределения)
Jarque-Bera	0.294371	Статистика Жака-Бера (для проверки выборки на нормальность распределения)
Probability	0.863134	Вероятность ошибиться, отвергнув гипотезу о том, что выборка из нормального распределения

Для возврата к отображению данных в виде таблицы значений выберите *View/SpreadSheet*.

Исходя из четырех условий, сделать общие выводы об адекватности построенной модели.

	Коэффициенты	Стандартная ошибка	t-статистика	p-значение
Свободный коэффициент	-73.24723	16.03423	-4.568177	0.0002
Коэффициент перед переменной x	0.259299	0.009999	25.93168	0.0000

4.2.9 Анализ результатов вычисления

Пример 2.1. Парная регрессия

Расчет статистик

Нулевая гипотеза (H_0): b_1 ($=0.2593$) равен 0

1 способ	2 способ
$t_f = -4.5682 > t_{kr} = 2.1009$	$P = 0 < 0.05$
нулевая гипотеза отвергается для данного уровня значимости ε	нулевая гипотеза отвергается для данного уровня значимости ε
коэффициент b_1 регрессионного уравнения <i>значим</i> на уровне значимости ε (или, говорят, что оценка коэффициента значимо отличается от нуля)	Коэффициент регрессионного уравнения <i>значим</i> на уровне значимости ε (или, говорят, что оценка коэффициента значимо отличается от нуля)

H_0 : b_0 ($=-73.2472$) равен 0

1 способ	2 способ
$t_f = -4.5682 > t_{kr} = 2.1009$	$P = 0.0002 < 0.05$
нулевая гипотеза отвергается для данного уровня значимости ε	нулевая гипотеза отвергается для данного уровня значимости ε
Коэффициент b_0 регрессионного уравнения <i>значим</i> на уровне значимости ε (или, говорят, что оценка коэффициента b_0 значимо отличается от нуля)	коэффициент b_0 регрессионного уравнения <i>значим</i> на уровне значимости ε (или, говорят, что оценка коэффициента b_0) значимо отличается от нуля)

F (F-statistic)	Значимость F Prob(F-statistic))
672.4519	0.0000

H_0 : $b_1 = 0$

1 способ	2 способ
$F_\Phi = 672.4519 > F_{kr} = 4,41$	$0.0000 < 0.05$
нулевая гипотеза отвергается для данного уровня значимости ε	нулевая гипотеза отвергается для данного уровня значимости ε
регрессионное уравнение значимо в целом на уровне значимости ε и вариация независимых переменных объясняет вариацию зависимой переменной в регрессионном уравнении	регрессионное уравнение значимо в целом на уровне значимости ε и вариация независимых переменных объясняет вариацию зависимой переменной в регрессионном уравнении

Произвести оценку адекватности модели

После того, как модель построена, необходимо удостовериться в её качестве. С этой целью выполняют проверку адекватности модели процессу, объекту или явлению, для которых она построена.

Проверить адекватность модели – значит установить, насколько хорошо модель описывает реальные процессы, происходящие в системе, насколько качественно она будет прогнозировать развитие данных процессов. Проверка адекватности проводится на основании некоторой экспериментальной информации, полученной на этапе функционирования системы или при проведении специального эксперимента, в ходе которого наблюдаются интересные процессы.

Проверка адекватности заключается в доказательстве факта, что точность результатов, полученных по модели, будет не хуже точности расчетов, произведенных на основании экспериментальных данных.

Если рассматривать с точки зрения целевого предназначения моделируемого объекта, то под адекватностью модели понимают степень её соответствия этому предназначению.

1. Условие адекватности: математическое ожидание случайной ошибки (возмущения, остатков) ε_i модели регрессии равно нулю во всех наблюдениях:

$$E(\varepsilon_i) = 0 \quad (i = \overline{1, n})$$

Оценкой математического ожидания случайной переменной является среднее остатков.

EViews

Среднее значение	Значение	Вывод:
Mean	$-8.53 \cdot 10^{-14}$	Числовое значение очень близко к нулю Условие адекватности выполняется

2. Гомоскедастичность – дисперсия случайной ошибки ε_i (возмущения, остатков) постоянна для всех наблюдений, т.е. дисперсия случайного члена в каждом наблюдении имеет только одно значение:

$$D(\varepsilon_i) = E(\varepsilon_i^2) = G^2 = const$$

EViews

Название теста	p-значение (вероятность)	уровень значимости
Уайта	Probability=0,491161	0,05
Нулевая гипотеза	дисперсия случайных ошибок в модели постоянна для всех наблюдений	
Вывод:	Принимается нулевая гипотеза Условие адекватности выполняется	

Гетероскедастичность: дисперсия случайных ошибок не обязательно одинакова для всех наблюдений.

3. Условие адекватности: отсутствие автокорреляции остатков. Между значениями случайных ошибок модели регрессии в любых двух наблюдениях отсутствует систематическая взаимосвязь, т. е. значения остатков ε_i распределены независимо друг от друга (возмущения ε_i и ε_j ($i \neq j$)) являются независимыми/некоррелированными).

Или

Ковариация случайных ошибок любых двух разных наблюдений равна нулю:

$$Cov(\varepsilon_i, \varepsilon_j) = E(\varepsilon_i \varepsilon_j) = 0 \quad (i \neq j).$$

EViews

Название теста	p-значение (вероятность)	уровень значимости
множителей Лагранжа	Probability=0,3	0,05
Нулевая гипотеза	случайные ошибки не коррелированы	
Вывод:	Принимается нулевая гипотеза Условие адекватности выполняется	

Название теста	Льюнга – Бокса (Q-статистика)	
Нулевая гипотеза	случайные ошибки не коррелированы	
Коррелограмма	Смотрите рис. 4.9 1 и 2 столбец содержит информацию об автокорреляции, 3 и 4 – частной автокорреляции.	
Вывод	Для Q-статистики все вероятности больше	уровня значимости 0,05
	Принимается нулевая гипотеза	
	Условие адекватности выполняется	

Название теста	Фактическое значение статистики, d	n	Число независимых переменных	Уровень значимости	Критические значения статистик	
Дарбина – Уотсона	Durbin-Watson stat=1.257096	20	1	0.05	d _l 1.20	d _u 1.41
Нулевая гипотеза	случайные ошибки не коррелированы					
Вывод	Результат неопределенный					

Статистика Дарбина-Уотсона

Статистика Дарбина-Уотсона предназначена для обнаружения автокорреляции первого порядка. Кроме того, уравнение регрессии должно иметь постоянный член и не содержать лаговую зависимую переменную в качестве факторной переменной

Тест основан на расчете d -статистики: $D = d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$

Расчетная величина d сравнивается с двумя табличными уровнями d_l и d_u

Возможные случаи:

Величина статистики D	Результат
$0 \leq D < d_l$	Присутствует положительная автокорреляция
$d_l \leq D < d_u$	Результат неопределенный
$d_u \leq D \leq 4 - d_u$	Автокорреляция отсутствует
$4 - d_u \leq D < 4 - d_l$	Результат неопределенный
$4 - d_l \leq D \leq 4_l$	Присутствует отрицательная автокорреляция

4. Условия адекватности: случайная ошибка модели регрессии – это случайная величина, подчиняющейся нормальному закону распределения с нулевым математическим ожиданием и дисперсией G^2 : $\varepsilon_i \sim N(0, G^2)$.

Или

Остатки подчиняются нормальному закону распределения.

Название теста	р-значение (вероятность)	JB	уровень значимости
Жака-Бера	Probability=0,863134	0,294371	0,05
Нулевая гипотеза	Остатки подчиняются нормальному закону распределения		
Вывод	Принимается нулевая гипотеза		
	Условие адекватности выполняется		

MS Excel

Название теста	Фактическое значение статистики	Критическое значение статистики
Жака-Бера	1,85	5,99
Нулевая гипотеза	Остатки подчиняются нормальному закону распределения	
Вывод	Принимается нулевая гипотеза	
	Условие адекватности выполняется	

Исходя из четырех условий, сделать общие выводы об адекватности построенной модели.

Если выдвинутые предположения справедливы, то оценки неизвестных параметров модели парной регрессии, полученные методом наименьших квадратов, можно считать эффективными, несмещенными и состоятельными оценками неизвестных параметров.

4.3 Задания для самостоятельной работы

Задание 3.1 (для построенных моделей в заданиях 6, 9)

1. *Построить:*
 - доверительные интервалы для коэффициентов регрессии.
2. *Рассчитать:*
 - значения *t*-статистик: фактические и критические
 - значения статистик Фишера *F*: фактические и критические.
3. *Произвести:*
 - Анализ адекватности построенной модели (проверить 4 предпосылки МНК).

5. Лабораторная работа №4

Тема: Прогнозирование и оценка точности прогнозов на основе регрессионных моделей.

Цель: Освоить основные принципы работы с указанными приложениями при вычислении прогнозных значений и показателей прогнозных свойств регрессионной модели.

5.1 Задания

Задание 4.1. Скопировать исходные данные (файлы, созданные в Блокноте) из материалов для студентов (\\ForStudent\КИС...).

Задание 4.2. Вычислить прогнозные значения и показатели прогнозных свойств регрессионных моделей.

Рассмотрим пример 2.1

5.2 Методические указания

5.2.1 Показатели оценок точности прогноза

Чтобы вынести суждение о качестве статистической модели, сначала нужно ознакомиться с табл. 5.1 (Критерии для оценки точности прогноза). Причем в первую очередь следует обратить внимание на раздел этой таблицы «Идеальное значение параметра». Из него можно сделать вывод: *чем ближе стремятся к нулю параметры, представленные в табл. 2, тем выше прогностическая ценность статистической модели.* Единственным исключением из этого правила является параметр Covariance Proportion (доля ковариации, т.е. доля несистематической ошибки), для которого идеальным значением является единица.

Таблица 5.1 – Критерии для оценки точности прогноза

Англоязычный термин	Перевод	Диапазон изменений	Идеальное значение параметра
Root Mean Squared Error	Квадратный корень средней ошибки предсказания	От 0 до ∞	$\Rightarrow 0$
Mean Absolute Error	Средняя ошибка по модулю	От 0 до ∞	$\Rightarrow 0$
Mean Absolute Percentage Error	Средняя ошибка по модулю, %	От 0 до ∞	$\Rightarrow 0$
Theil Inequality Coefficient	Коэффициент неравенства Тейла	От 0 до 1	$\Rightarrow 0$
Bias Proportion	Доля систематической ошибки		$\Rightarrow 0$
Variance Proportion	Доля вариации, т.е. разницы между колебанием фактических и расчетных значений	От 0 до 1	$\Rightarrow 0$
Covariance Proportion	Доля ковариации, т.е. несистематической ошибки	От 0 до 1	$\Rightarrow 1$

Ошибка аппроксимации в пределах 5–7 % свидетельствует о хорошем соответствии статистической модели исходным данным.

5.2.2 Доверительный интервал прогноза

Очевидно, что точность прогноза тем выше, чем меньше величина ошибки, которая представляет собой разность между прогнозируемым и фактическим значением исследуемой величины.

Проблема состоит в том, чтобы вычислить ошибку прогноза, так как фактическое значение прогнозируемой величины станет известно лишь в будущем. Следовательно, методы оценки точности по уже свершившимся событиям не имеют практической ценности, так как являются лишь констатацией факта. При разработке прогноза оценку его точности требуется производить заранее (априорно), когда истинное значение прогнозируемой величины еще не известно. Как же поступить в этих случаях? Дискуссии в специальной литературе отмечают эти трудности, и в итоге все предложения так или иначе связаны с определением доверительного интервала на основе статистического выборочного метода. При этом точность прогноза оценивается величиной доверительного интервала для заданной вероятности его осуществления, а под достоверностью понимают оценку вероятности осуществления прогноза в заданном доверительном интервале. Таким образом, точность прогноза выражается с помощью вероятностных пределов фактической величины от прогнозируемого значения.

Точное совпадение фактических данных и прогностических точечных оценок, полученных путем экстраполяции кривых, характеризующих тенденцию, явление маловероятное. И этому виной следующие источники погрешностей:

1) выбор формы кривой (порядка полинома и т. д.), характеризующей тренд, содержит элемент субъективизма. Во всяком случае, часто нет твердой основы для того чтобы утверждать, что выбранная форма кривой является единственно возможной или тем более наилучшей для экстраполяции в данных конкретных условиях;

2) оценивание параметров кривых (оценивание тренда) производится на основе ограниченной совокупности наблюдений, каждое из которых содержит случайную компоненту. В силу этого параметрам кривой, а, следовательно, и ее положению в пространстве свойственна некоторая неопределенность;

3) тренд характеризует некоторый средний уровень ряда на каждый момент времени. Отдельные наблюдения, как правило, отклонялись от него в прошлом. Естественно ожидать, что подобного рода отклонения будут происходить и в будущем.

Погрешность, связанная со вторым и третьим ее источниками, и может быть отражена в виде доверительного интервала прогноза при принятии некоторых допущений о свойстве ряда.

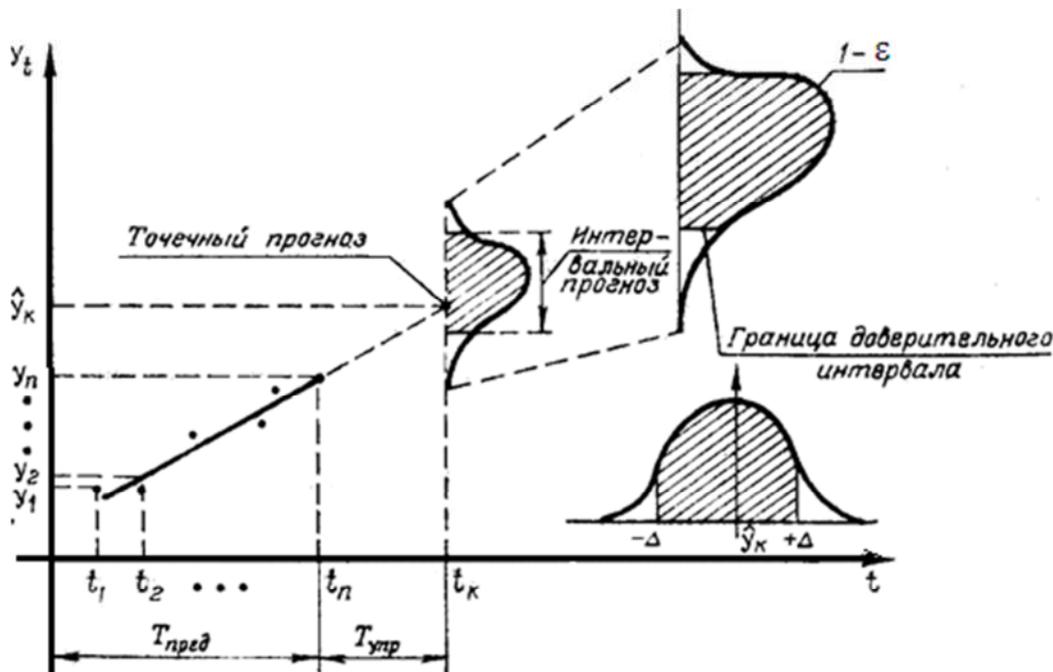


Рисунок 5.1 - Доверительный интервал прогноза

Доверительный интервал для прогноза:

$$y_n \pm \Delta = y_n \pm t_{\text{крит}} S_{\Pi}$$

где S_{Π} - стандартная ошибка прогноза,
 $t_{\text{крит}}$ - t -статистика Стьюдента,
 y_n - прогнозное значение.

При определении стандартной ошибки прогноза, основанного на линейной модели, необходимо учитывать, по крайней мере, два источника неопределенностей. Во-первых, действительные значения зависимой переменной не будут совпадать с расчетными (прогнозными), так как сама линия регрессии описывает взаимосвязь лишь в среднем. Отдельные наблюдения рассеяны вокруг нее. Таким образом, наиболее очевидным фактором, во многом определяющим надежность получаемых по уравнению регрессии прогностических оценок, является рассеяние наблюдений вокруг линии регрессии. Во-вторых, в силу того, что оценивание параметров модели осуществляется по выборочным данным, оценки a и b сами содержат некоторую погрешность. Причем, погрешность в значении a приводит к вертикальному сдвигу линии регрессии. В свою очередь, колебимость оценки b , связанная с ее выборочным происхождением, приводит к «покачиванию» линии регрессии.

Если нанести доверительные границы на график, то они расположены выше и ниже линии регрессии в виде ветвей гиперболы, ограничивая доверительную область.

Доверительный интервал уменьшается при увеличении продолжительности наблюдения (периода основания прогноза) и растет с увеличением периода упреждения прогноза.

5.2.3 Интерпретация полученного уравнения регрессии

Существуют два этапа интерпретации уравнения регрессии. Первый этап состоит в словесном истолковании уравнения так, чтобы это было понятно человеку, не являющемуся специалистом в области статистики. На втором этапе необходимо решить, следует ли ограничиться этим или провести более детальное исследование зависимости.

Интерпретация линейного уравнения регрессии.

Можно сказать, что увеличение x на одну единицу (в единицах измерения переменной x) приведёт к увеличению значения y на b_1 единиц (в единицах измерения переменной y).

Постоянная b_0 дает прогнозируемое значение y (в единицах y), если $x=0$. Это может иметь или не иметь ясного смысла в зависимости от конкретной ситуации.

5.2.4 Технологии вычислений в MS Excel

Можно рассчитать значения линейного тренда с помощью стандартных функций Excel:

=ТЕНДЕНЦИЯ(известные значения y ; известные значения x ; новые значения x ; константа),

где **константа** - ставится 1, если необходимо, чтобы значения тренда рассчитывались с учетом коэффициента (a) для линейного тренда $y=a+bx$;

Для того чтобы рассчитать значения тренда для всего временного диапазона, в поле **новые значения x** вводим диапазон значений X , выделяем диапазон ячеек, равный диапазону со значениями X с формулой в первой ячейке, и нажимаем клавишу F2, а затем — клавиши CTRL + SHIFT + ВВОД.

=ПРЕДСКАЗ(x ; известные значения y ; известные значения x),

где вместо X проставляем номер периода, для которого рассчитываем значение тренда.

Известные значения y фиксируют диапазон в формуле: выделить ссылку и нажать F4).

Средняя ошибка аппроксимации:

$$\bar{A} = \frac{\sum |y_i - y(x_i)| : y_i}{n} \cdot 100(\%) = 3,43\%$$

Качество модели, исходя из относительных отклонений по каждому наблюдению, признается хорошим, если средняя ошибка аппроксимации не превышает 10%.

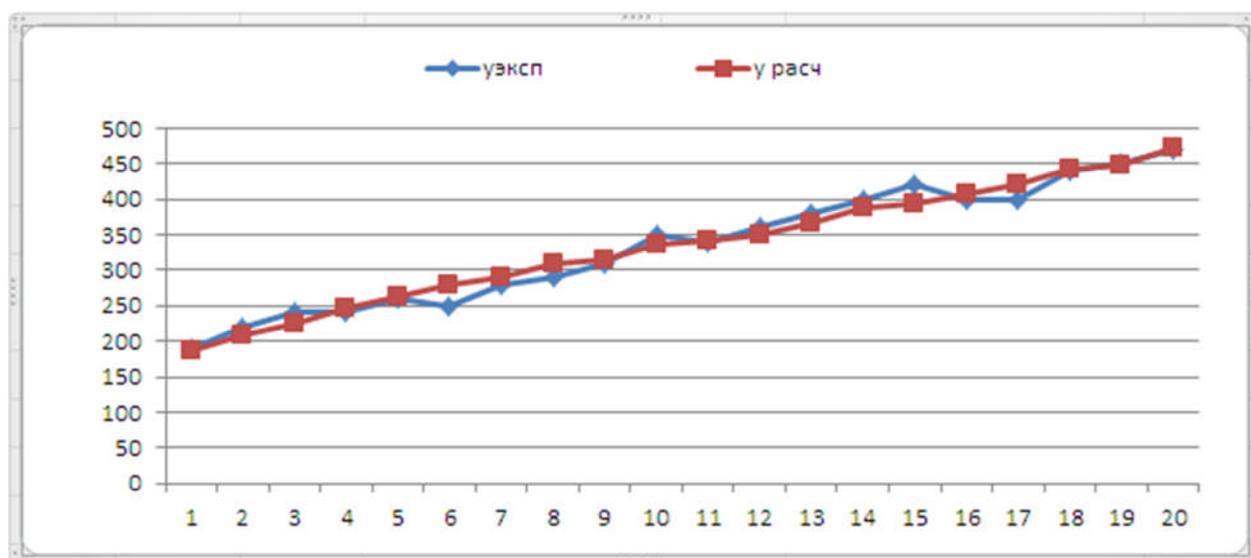
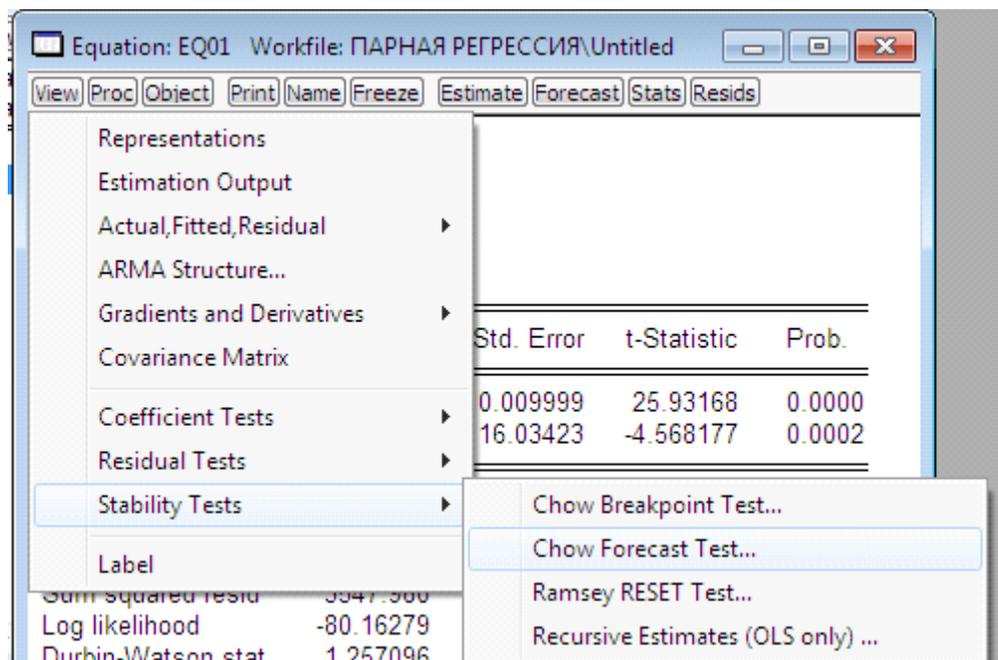


Рисунок 5.2 – Фактические и прогнозные значения

5.2.5 Технологии вычислений в EViews

1. Провести тест Чоу (Chow Forecast Test) на стабильность модели по выборочной совокупности (за **весь период**) и сделать выводы.

Для проверки гипотезы необходимо в окне объекта Equation выполнить команду **View / Stability Tests / Chow Forecast test...**, указать заданный выше диапазон и нажать ОК.



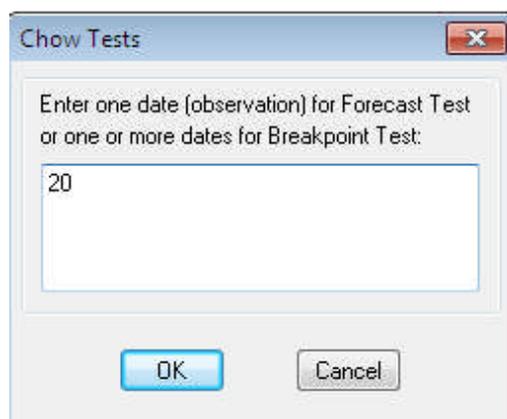


Рисунок 5.3 – Тест Чоу

Chow Forecast Test: Forecast from 20 to 20

F-statistic	0.009712	Probability	0.922648
Log likelihood ratio	0.011424	Probability	0.914881

Нулевая гипотеза: *Модель является стабильной и может использоваться для прогнозирования.*

ВЫВОД

Для принятия решения использовать эквивалентную форму критерия, который представляет собой сравнение уровня значимости ϵ и p -значения, которое указывается в строке Probability в окне оценивания теста. Если $p > \epsilon$, гипотеза принимается, модель стабильна и может использоваться для прогнозирования.

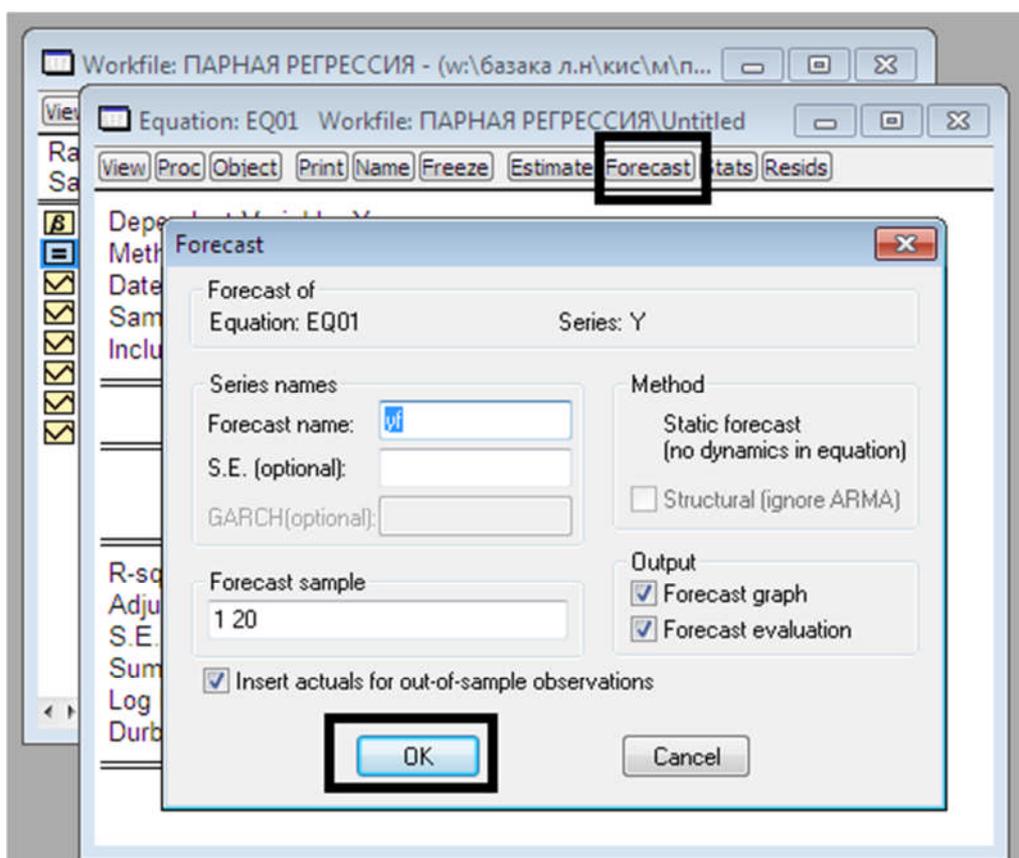
2. Оценить прогнозные свойства построенной модели за весь период. Построить таблицу, содержащую прогнозные свойства модели.

Таблица 5.2 – Прогнозные свойства модели

Показатель	Значение
<i>Root Mean Squared Error</i> (квадратный корень средней ошибки предсказания)	?
<i>Mean Absolute Error</i> (средняя ошибка по модулю)	?
<i>Mean Absolute Percentage Error</i> (средняя ошибка по модулю, %)	?
<i>Theil Inequality Coefficient</i> (коэффициент неравенства Тейла)	?

Проверить, выполняется ли равенство?

$$\text{Bias Proportion} + \text{Variance Proportion} + \text{Covariance Proportion} = 1$$



Открыть объект Equation и нажать закладку Forecast.

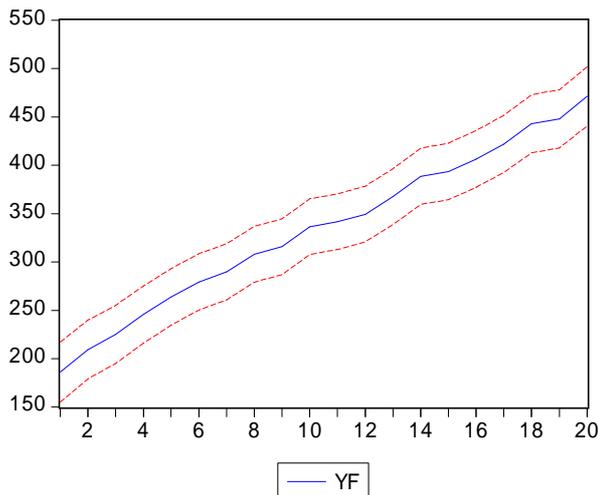
В результате откроется мини-окно FORECAST.

По умолчанию в опции FORECAST NAME (название файла с прогнозом) задается название файла с точечным прогнозом путем прибавления к исходному файлу латинской буквы *f*. Например, если у нас исходный файл — *y*, то название файла с прогнозом будет задано программой как *yf*.

В опции FORECAST SAMPLE (*выборка для прогноза*) по умолчанию задается исходная выборка данных для прогноза, которую при необходимости можно изменить.

В опции METHOD (*метод прогноза*) можно выбрать STATIC FORECAST (статический прогноз), т. е. будет оцениваться точность прогноза только на один следующий период. Если в опции METHOD выбрать вариант DYNAMIC FORECAST (динамический прогноз), то это увеличило бы временной горизонт для прогноза, но его точность существенно снизилась бы. Дело в том, что при динамическом прогнозировании предсказание на следующий период составляется так же, как и при статическом, но прогнозы на более длительные сроки составляются на основе расчетных, т. е. предсказанных, а не фактических значений независимой переменной.

В опции OUTPUT (*вывод итогов*) задается вариант FORECAST EVALUATION (оценка прогноза) и получили таблицу с оценкой точности прогноза этой статистической модели. При необходимости в последней опции можно задать еще и вариант FORECAST GRAPH (график прогнозов), после чего можно получить и график с прогнозами.



Forecast: YF	
Actual: Y	
Forecast sample: 1 20	
Included observations: 20	
Root Mean Squared Error	13.31912
Mean Absolute Error	10.64607
Mean Abs. Percent Error	3.428360
Theil Inequality Coefficient	0.019337
Bias Proportion	0.000000
Variance Proportion	0.006604
Covariance Proportion	0.993396

Показатель	Значение
Forecast (прогноз):	yf
Actual (фактическое значение):	y
Forecast sample (выборка для прогноза):	1 20
Included observations (включено наблюдений):	20
Root Mean Squared Error (квадратный корень средней ошибки предсказания)	13.31912
Mean Absolute Error (средняя ошибка по модулю)	10.64607
Mean Absolute Percentage Error (средняя ошибка по модулю, %)	3.428360
Theil Inequality Coefficient (коэффициент неравенства Тейла)	0.019337
Bias Proportion (доля систематической ошибки прогноза)	0.000000
Variance Proportion (доля вариации)	0.006604
Covariance Proportion (доля ковариации)	0.993396

Ошибка аппроксимации в пределах 5–7 % свидетельствует о хорошем соответствии статистической модели исходным данным.

3. Произвести оценку точности прогнозных значений yf . Для реализации поставленной задачи необходимо вычислить границы доверительного интервала прогноза (yf , количество значений ряда $n=20$):

- Вычислить среднеквадратическое отклонение ошибки прогноза:
scalar sef=@stdev(yf)

Scalar SEF = 83.5234034424

- Вычислить статистику Стьюдента $t_\varepsilon = t_{\text{крит}}(\text{show @qtdist}(0.975, n-k))$; $n=20$, $k=2$.
- Вычислить верхнюю (**series yf_u= yf+sef* t_ε**) и нижнюю границы доверительного интервала прогноза (**series yf_l= yf -sef* t_ε**).

series yf_u= yf +sef*2.10092204024104

series yf_l= yf -sef*2.10092204024104

	yf_l	yf	yf_u
1			
2			
3			
...			
20			

5.2.6 Анализ результатов вычисления

Пример 2.1 Парная регрессия

Введем обозначения, пусть $b_1 = a$, $b_0 = b$.

Таблица 5.3 - Парная регрессия

	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение
b	-73,24722592	16,03423416	-4,568177389	0,000238365
a	0,259298713	0,009999303	25,93167832	1,04396E-15

Таблица 5.4 - Регрессионная статистика

Множественный R	0,986878979
R-квадрат	0,973930119
Нормированный R-квадрат	0,972481792
Стандартная ошибка регрессии	14,03958782
Наблюдения	20

Фактические значения t-статистики превосходят табличное (критическое) значение:

Коэффициенты	$t_{\text{крит}}$	Сравнение	$t_{\text{набл}}$
$b = -73,24722592$	2,10092204	<	$ -4,568177389 $
$a = 0,259298713$	2,10092204	<	25,93167832

Поэтому параметры a , b отличаются от нуля и статистически значимы.

Определим предельную ошибку для каждого показателя:

Предельная ошибка	Расчет	Значение
Δ_b	$t_{\text{крит}} * S_b$	33,68667594
Δ_a	$t_{\text{крит}} * S_a$	0,021007756

Доверительные интервалы:

Коэффициенты	Нижняя граница	Верхняя граница
B	$b - \Delta_b = -106,9339019$	$b + \Delta_b = -39,56055$
A	$a - \Delta_a = 0,238290956$	$a + \Delta_a = 0,280306469$

Анализ верхней и нижней границ доверительных интервалов приводит к выводу о том, что с вероятностью $p = 1 - \varepsilon = 0,95$ параметры a и b , находясь в указанных границах, не принимают нулевых значений, т.е. являются статистически значимыми и существенно отличны от нуля.

Поэтому любое значение:

- из интервала $(a - \Delta_a; a + \Delta_a)$ может служить оценкой параметра a ,
- из интервала $(b - \Delta_b; b + \Delta_b)$ может служить оценкой параметра b .

Таким образом, уравнение регрессии имеет вид

$$y = -73,25 + 0,26x.$$

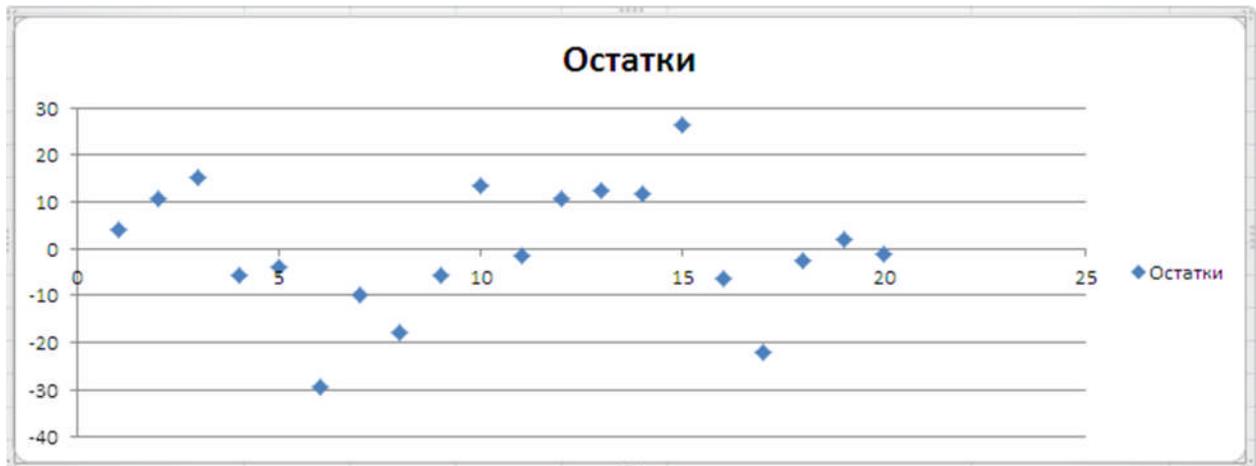
Случайная переменная отсутствует в уравнении, так как коэффициенты регрессии имеют случайный характер, т. е. неучтенные факторы повлияли на их значение при применении МНК.

Качество модели определяет средняя ошибка аппроксимации.

$$\bar{A} = \frac{\sum |y_i - y(x_i)| : y_i}{n} \cdot 100(\%) = 3,43\%$$

Качество построенной модели оценивается как хорошее, так как средняя ошибка аппроксимации не превышает 10%. Полученные оценки уравнения регрессии позволяют использовать его для прогноза.

Построенная модель является адекватной. Так как выполняются все четыре предпосылки МНК.



Параметр регрессии позволяет сделать вывод, что с увеличением значения переменной x на 1 (ед. изм.), то значение переменной возрастает в среднем 0,259298713 (ед.изм.).

Так как на графике остатков точки разбросаны вдоль оси Ox хаотично без видимой закономерности, то зависимости между остатками не наблюдается.

Под прогнозированием в эконометрике понимается нахождение оценки зависимой переменной для любого допустимого значения независимой переменной. Различают точечное и интервальное прогнозирование.

Точечная оценка y^* зависимой переменной определяется путем подстановки в найденное уравнение регрессии заданного допустимого значения независимой переменной x^* .

Интервальный прогноз, или доверительный интервал прогноза, имеет вид:

№	yf_l	yf	yf_u
1	10.5753277382132	186.051486906348	361.527646074482
2	33.9122118926738	209.388371060808	384.864530228943
3	49.4701346623142	224.946293830449	400.422452998583
4	70.2140316885014	245.690190856636	421.166350024771
5	88.3649415864153	263.84110075455	439.317259922684
6	103.922864356056	279.39902352419	454.875182692325
7	114.294812869149	289.770972037284	465.247131205419
8	132.445722767063	307.921881935198	483.398041103332
9	140.224684151883	315.700843320018	491.177002488153
10	160.968581178071	336.444740346205	511.92089951434
11	166.154555434617	341.630714602752	517.106873770886
12	173.933516819438	349.409675987572	524.885835155707
13	192.084426717351	367.560585885486	543.036745053621
14	212.828323743539	388.304482911673	563.780642079808
15	218.014298000085	393.49045716822	568.966616336355
16	230.979233641452	406.455392809587	581.931551977722
17	246.537156411093	422.013315579227	597.489474747362
18	267.28105343728	442.757212605415	618.233371773549
19	272.467027693827	447.943186861961	623.419346030096
20	295.803911848287	471.280071016422	646.756230184557

Sp

83,52340344

средняя квадратическая
ошибка прогноза

5.3 Задания для самостоятельной работы

Задание 4.1 (для построенных моделей в заданиях 6,9)

1. *Рассчитать:*
 - прогнозные значения.
2. *Построить:*
 - график, отображающий прогнозные и фактические значения.
3. *Произвести:*
 - расчет показателей прогнозных свойств регрессионных моделей.

ПРИЛОЖЕНИЕ А
Статистика Дарбина-Уотсона: d_L и d_U , уровень значимости 5%

N	k = 1		k = 2		k = 3		k = 4		k = 5	
	d_L	d_U								
15	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21
16	1,10	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,67	2,10
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,40	1,08	1,53	0,97	1,68	0,86	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,90
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
26	1,30	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,88
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83
31	1,36	1,50	1,30	1,57	1,23	1,65	1,16	1,74	1,09	1,83
32	1,37	1,50	1,31	1,57	1,24	1,65	1,18	1,73	1,11	1,82
33	1,38	1,51	1,32	1,58	1,26	1,65	1,19	1,73	1,13	1,81
34	1,39	1,51	1,33	1,58	1,27	1,65	1,21	1,73	1,15	1,81
35	1,40	1,52	1,34	1,58	1,28	1,65	1,22	1,73	1,16	1,80
36	1,41	1,52	1,35	1,59	1,29	1,65	1,24	1,73	1,18	1,80
37	1,42	1,53	1,36	1,59	1,31	1,66	1,25	1,72	1,19	1,80
38	1,43	1,54	1,37	1,59	1,32	1,66	1,26	1,72	1,21	1,79
39	1,43	1,54	1,38	1,60	1,33	1,66	1,27	1,72	1,22	1,79
40	1,44	1,54	1,39	1,60	1,34	1,66	1,29	1,72	1,23	1,79
45	1,48	1,57	1,43	1,62	1,38	1,67	1,34	1,72	1,29	1,78
50	1,50	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,34	1,77
55	1,53	1,60	1,49	1,64	1,45	1,68	1,41	1,72	1,38	1,77
60	1,55	1,62	1,51	1,65	1,48	1,69	1,44	1,73	1,41	1,77
65	1,57	1,63	1,54	1,66	1,50	1,70	1,47	1,73	1,44	1,77
70	1,58	1,64	1,55	1,67	1,52	1,70	1,49	1,74	1,46	1,77
75	1,60	1,65	1,57	1,68	1,54	1,71	1,51	1,74	1,49	1,77
80	1,61	1,66	1,59	1,69	1,56	1,72	1,53	1,74	1,51	1,77
85	1,62	1,67	1,60	1,70	1,57	1,72	1,55	1,75	1,52	1,77
90	1,63	1,68	1,61	1,70	1,59	1,73	1,57	1,75	1,54	1,78
95	1,64	1,69	1,62	1,71	1,60	1,73	1,58	1,75	1,56	1,78
100	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,76	1,57	1,78

ПРИЛОЖЕНИЕ Б

Значение критерия Стьюдента (t-критерия) Стьюдента (n – число наблюдений , m – число факторов)

df	P													
	0,01	0,02	0,03	0,04	0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,45	0,5	0,55
1	63,6567412	31,8205159	21,2049488	15,8945448	12,7062047	6,3137515	4,1652998	3,0776835	2,4142136	1,9626105	1,6318517	1,1708496	1,0000000	0,8540807
2	9,9248432	6,9645567	5,6427784	4,8487322	4,3026527	2,9199856	2,2819306	1,8856181	1,6035674	1,3862066	1,2096295	0,9313343	0,8164966	0,7126268
3	5,8409093	4,5407029	3,8960459	3,4819088	3,1824463	2,3533634	1,9243197	1,6377444	1,4226253	1,2497781	1,1045199	0,8664184	0,7648923	0,6714714
4	4,6040949	3,7469474	3,2976297	2,9985279	2,7764451	2,1318468	1,7781922	1,5332063	1,3443976	1,1895669	1,0572994	0,8363711	0,7406971	0,6519466
5	4,0321430	3,3649300	3,0028750	2,7565085	2,5705818	2,0150484	1,6993626	1,4758840	1,3009490	1,1557673	1,0305484	0,8190861	0,7266868	0,6405729
6	3,7074280	3,1426684	2,8289279	2,6122418	2,4469118	1,9431803	1,6501731	1,4397557	1,2733493	1,1341569	1,0133489	0,8078686	0,7175582	0,6331353
7	3,4994833	2,9979516	2,7145730	2,5167524	2,3646243	1,8945786	1,6165917	1,4149239	1,2542787	1,1191591	1,0013672	0,8000050	0,7111418	0,6278948
8	3,3553873	2,8964594	2,6338144	2,4489850	2,3060041	1,8595480	1,5922214	1,3968153	1,2403183	1,1081454	0,9925445	0,7941885	0,7063866	0,6240043
9	3,2498355	2,8214379	2,5738040	2,3984410	2,2621572	1,8331129	1,5737358	1,3830287	1,2296592	1,0997162	0,9857784	0,7897126	0,7027221	0,6210023
10	3,1692727	2,7637695	2,5274842	2,3593146	2,2281388	1,8124611	1,5592359	1,3721836	1,2212554	1,0930581	0,9804254	0,7861621	0,6998121	0,6186159
11	3,1058065	2,7180792	2,4906639	2,3281398	2,2009852	1,7958848	1,5475598	1,3634303	1,2144602	1,0876664	0,9760850	0,7832771	0,6974453	0,6166734
12	3,0545396	2,6809980	2,4607002	2,3027217	2,1788128	1,7822875	1,5379565	1,3562173	1,2088525	1,0832114	0,9724949	0,7808867	0,6954829	0,6150617
13	3,0122758	2,6503088	2,4358452	2,2816036	2,1603687	1,7709334	1,5299196	1,3501713	1,2041462	1,0794687	0,9694762	0,7788738	0,6938293	0,6137030
14	2,9768427	2,6244941	2,4148977	2,2637813	2,1447867	1,7613101	1,5230951	1,3450304	1,2001403	1,0762802	0,9669026	0,7771556	0,6924171	0,6125419
15	2,9467129	2,6024803	2,3970050	2,2485403	2,1314495	1,7530503	1,5172280	1,3406056	1,1966893	1,0735314	0,9646824	0,7756718	0,6911969	0,6115385
16	2,9207816	2,5834872	2,3815454	2,2353584	2,1199053	1,7458837	1,5121302	1,3367572	1,1936854	1,0711372	0,9627475	0,7743775	0,6901323	0,6106625
17	2,8982305	2,5669340	2,3680548	2,2238453	2,1098156	1,7396067	1,5076598	1,3333794	1,1910471	1,0690331	0,9610464	0,7732386	0,6891951	0,6098912
18	2,8784405	2,5523796	2,3561800	2,2137032	2,1009220	1,7340636	1,5037077	1,3303909	1,1887115	1,0671695	0,9595390	0,7722287	0,6883638	0,6092069
19	2,8609346	2,5394832	2,3456475	2,2047013	2,0930240	1,7291328	1,5001888	1,3277282	1,1866293	1,0655074	0,9581940	0,7713270	0,6876215	0,6085956
20	2,8453397	2,5279770	2,3362422	2,1966577	2,0859634	1,7247182	1,4970355	1,3253407	1,1847614	1,0640158	0,9569866	0,7705172	0,6869545	0,6080463

ПРИЛОЖЕНИЕ В

	A	B	C	D	E	F	G
n		0,01	0,02	0,03	0,04	0,05	0,1
1		63,65674	31,82052	21,20495	15,89454	12,7062	6,313752
2		9,924843	6,964557	5,642778	4,848732	4,302653	2,919986

Рисунок 1 – Функция СТЬЮДРАСПОБР

СТЮДРАСПОБР(вероятность; степени_свободы)

Вероятность — вероятность, соответствующая двустороннему распределению Стьюдента.
 Степени_свободы — число степеней свободы, характеризующее распределение.

Значение F-критерия Фишера при уровне значимости P=0.05

k1=количеству факторов

k2=n-m-1	1	2	3	4	5	6	7
1	161,4476	199,5	215,7073	224,5832	230,1619	233,986	236,7684
5	6,607891	5,786135	5,409451	5,192168	5,050329	4,950288	4,875872
10	4,964603	4,102821	3,708265	3,47805	3,325835	3,217175	3,135465
15	4,543077	3,68232	3,287382	3,055568	2,901295	2,790465	2,706627
20	4,351244	3,492828	3,098391	2,866081	2,71089	2,598978	2,514011
18	4,413873	3,554557	3,159908	2,927744	2,772853	2,661305	2,576722
30	4,170877	3,31583	2,922277	2,689628	2,533555	2,420523	2,334344
35	4,121338	3,267424	2,874187	2,641465	2,485143	2,371781	2,285235
45	4,056612	3,204317	2,811544	2,578739	2,422085	2,308273	2,221221
50	4,03431	3,18261	2,790008	2,557179	2,400409	2,286436	2,199202
55	4,016195	3,164993	2,772537	2,539689	2,382823	2,268717	2,181333
60	4,001191	3,150411	2,758078	2,525215	2,36827	2,254053	2,166541
65	3,98856	3,138142	2,745915	2,51304	2,356028	2,241716	2,154095

n - число наблюдений

m - количество факторов

	A	B	C	D	E	F	G
1							
2	k2=n-m-1	1	2	3	4	5	6
3	1	161,4476	199,5	215,7073	224,5832	230,1619	233,986

Рисунок 2 – Функция ФРАСПОБР

ФРАСПОБР(вероятность;степени_свободы1;степени_свободы2)

Вероятность — это вероятность, связанная с F-распределением.

Степени_свободы 1 — количество факторов.

Степени_свободы 2 — $(n - k - 1)$,

где m – количество факторов, включенных в модель

ПРИЛОЖЕНИЕ Г

Автокорреляция

Отсутствие автокорреляции между значениями ряда оценивается по коррелограмме ряда.

Экономический пакет EViews

Построение коррелограммы временного ряда (выборочных автокорреляционной и частной автокорреляционной функций). Для построения коррелограммы временного ряда выберите опцию **Correlogram...** в меню **View** окна временного ряда – **View/Correlogram...**:

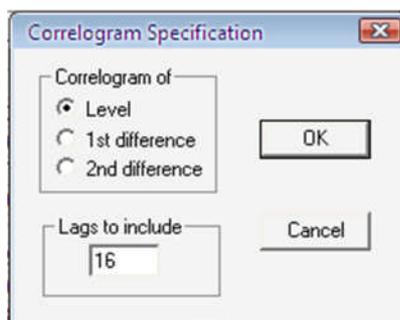
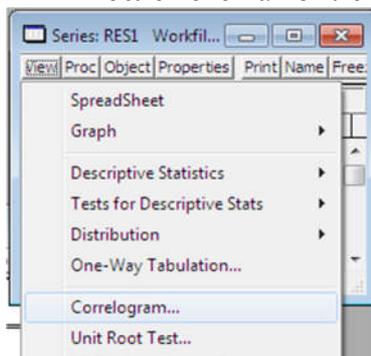


Рисунок 1 – Коррелограмма (спецификация)

В появившемся окне необходимо определить:

- Использование какого ряда (раздел – Correlogram of): самого временного ряда (Level), либо его первых (1st difference) или вторых разностей (2d difference) необходимо для построения коррелограммы.
- Количество лагов*, для которых необходимо построить коррелограмму** (Lags to include).

После того как было отмечено все в необходимых полях необходимо нажать ОК.



* Простое эмпирическое правило говорит, что достаточно взять $T/4$ запаздываний, где T – длина временного ряда (количество наблюдений).

** При большом количестве наблюдений, как правило, по умолчанию предлагается построить коррелограмму для 36 запаздываний.

Рисунок 2 – Способ построения коррелограммы

Замечание. График выборочной автокорреляционной функции (ВАКФ) называется коррелограммой. Коррелограмма является быстро убывающей функцией. Если формально построенная коррелограмма не удовлетворяет этому свойству, это, скорее всего, означает, что ряд на самом деле нестационарный.

Замечание. Таким образом, вы можете определить нужное количество AR-членов, просто глядя на PACF (ВЧАКФ) ряда: если значения PACF значимы до лага k , а после него незначимы, следует начинать с авторегрессионной модели порядка k .

ПРИЛОЖЕНИЕ Д

Спецификация уравнения (регрессия)

Построить модель можно аналитическим способом.

Для построения моделей экономических временных рядов использовать термин регрессия – общепринятый термин в математической статистике и эконометрике.

Для построения модели необходимо задать:

- Перечень независимых переменных, по которым вычисляется индикатор;
- Коэффициенты при независимых переменных;
- Формулу вычисления индикатора, по которой будет вычисляться зависимая переменная.

Следует подогнать регрессию к ряду. В эконометрике слово «подгонка – fit, fitting» является стандартной процедурой, в ходе которой с помощью одного из многочисленных методов оценки вычисляется соответствие регрессии (модели) рядам. Наиболее известным из методов оценки является метод наименьших квадратов (МНК).

В результате оценки интерес проявляется к двум вопросам:

- Соответствие модели ряда – величина остающейся ошибки;
- Стабильность вычисленных параметров регрессии в будущем.

Ответы на эти вопросы даются в ходе диагностики моделей.

Диагностика моделей

Диагностика индикаторов (регрессий) разделена на три группы:

- Диагностика коэффициентов;
- Диагностика остатков;
- Диагностика стабильности.

Каждая процедура проверки включает спецификацию нулевой гипотезы, которая является гипотезой при тесте. Результат теста состоит из выборки значений одной или более статистик и их присоединенных p -значений. Последние указывают на вероятность выполнения условия нулевой гипотезы, на которой построена тестовая статистика.

Таким образом, малые p -значения приводят к отклонению нулевой гипотезы. Например, если p -значение лежит между 0.05 и 0.01, то нулевая гипотеза отклонена на пяти процентном уровне, но не на однопроцентном уровне.

Следует учесть, что имеются различные предположения и результаты распределения, связанные с каждым тестом. Например, у некоторых из статистик есть точные, конечные тестовые распределения (обычно t или F -распределения). Другие являются большими выборками тестовой статистики с асимптотическими χ^2 распределениями.

ПРИЛОЖЕНИЕ Е

Статистические пакеты

Пакет SPSS для Windows

Пакет SPSS предназначен в первую очередь для статистиков-профессионалов. Он включает развитый аппарат статистического анализа. Программу SPSS считают одним из лидеров среди универсальных статистических пакетов. Научиться писать алгоритмы в виде скриптов на языке SPSS можно самостоятельно без начального программистского образования.

SPSS имеет удобные графические средства (более 50 типов диаграмм), а также развитые средства подгонки ответов. Аналитические параметры отображаются на экране в виде простых и понятных меню и диалоговых окон. Контекстно-ориентированная справочная система содержит пошаговые инструкции для наиболее важных операций.

Пакет Statistica 6.0

Пакет Statistica 6.0 предназначен для тех, кто хорошо владеет статистической терминологией. Пакет пользуется популярностью благодаря высокой активности фирмы-разработчика Statsoft и дилера в России – Softline, способствующих популяризации пакета (см. сайт www.statsoft.ru).

Ряд авторов считает, что пакет Statistica является хорошо сбалансированным по соотношению «мощность/удобство». Наличие достаточно широкого спектра функциональных алгоритмов делает его достаточно привлекательным для статистиков-профессионалов. Пакет включает в себя ряд непараметрических методов анализа, методы многомерного анализа: дискриминантного, факторного кластерного, логлинейного и др.

Средства манипуляции исходными данными в пакете Statistica хорошо развиты. Данные относительно легко отредактировать, можно создавать новые переменные, выбирать отдельные наблюдения или «вырезать» подмножества данных по строкам и/или по столбцам таблицы «объект-признак». Благодаря обширной панели инструментов, для выделения большинства манипуляций достаточно нескольких щелчков мыши, так как почти для всех функций пакета здесь имеются пиктограммы.

Сильной стороной пакета является графика и средства редактирования графических материалов. В пакете представлены сотни типов графиков 2D или 3D, матрицы и пиктограммы. Предоставляется возможность разработки собственного дизайна графика.

<http://www.quality-life.ru/metodologiya01.php>

ПРИЛОЖЕНИЕ Ж

Приложение *EViews*

Основы

Начало работы с EViews. Познакомится с элементами главного окна EViews.

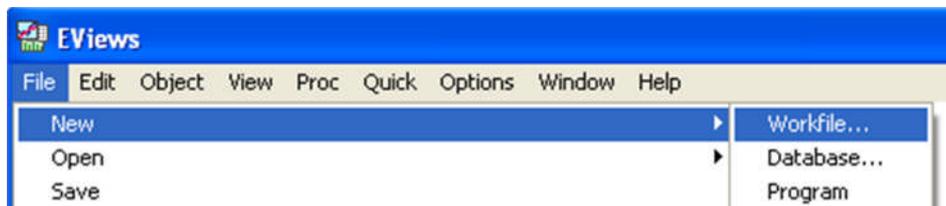


Создайте новый рабочий файл Workfile: **File/New/Workfile**. Затем он заполняется данными, взятыми из файла *.txt: **File/Import/Read Text-Lotus-Excel**. Сохраните файл: сделайте активным окно Workfile и выберите в главном меню **File/Save As...**

ПРИЛОЖЕНИЕ 3

EViews 5. Создание рабочего файла (Workfile)

- Осуществить запуск EViews 5.
- Создать новый Workfile: EViews/File/New/Workfile.



Появится окно «Workfiles Create» (Создание рабочего файла).

- Определить структурный тип рабочего файла.

Workfile structure type (рабочего файла структурный тип):	Рабочие workfiles, имеющие нерегулярные данные по дате, могут позже преобразованы в workfiles с датой другой спецификации и / или содержащие другие серии (ряды).
<ul style="list-style-type: none"> • датировано - регулярная частота представления данных, • балансная группа, • неструктурированно/недатированно (observations – указать количество наблюдений). 	

- При регулярной частоте представления данных необходимо выбрать частоту для представления данных (date specification), ввести начальную и конечную дату.

Представление данных	Формат даты	
	начальную дату (Start date)	конечную дату (End date)
Frequency		
ГОДОВЫЕ (Annual)	YYYY	
МЕСЯЧНЫЕ (Monthly)	MM/DD/YYYY	
КВАРТАЛЬНЫЕ (Quarterly)	YYYYq(номер квартала)	

- При недатированных данных необходимо указать количество наблюдений (observations – наблюдения).
- Ввести имя рабочего файла в поле WF раздела Names (Optional).

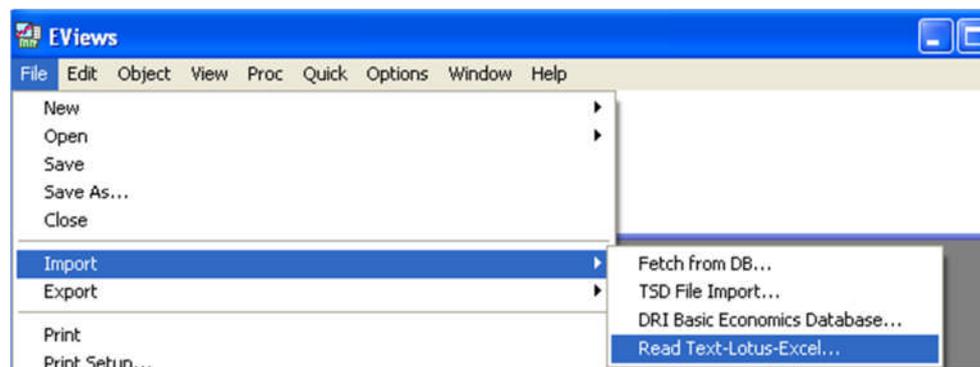
В окне Workfile: ИМЯ РАБОЧЕГО ФАЙЛА будет указано количество наблюдений (№№ obs) и будут сформированы 2 служебных объекта: первый – с – служебный ряд, в котором в дальнейшем будут сохраняться вычисленные значения параметров построенной модели; второй - resid - служебный ряд, в котором будут сохраняться остатки последнего построенного эконометрического уравнения.

ПРИЛОЖЕНИЕ И

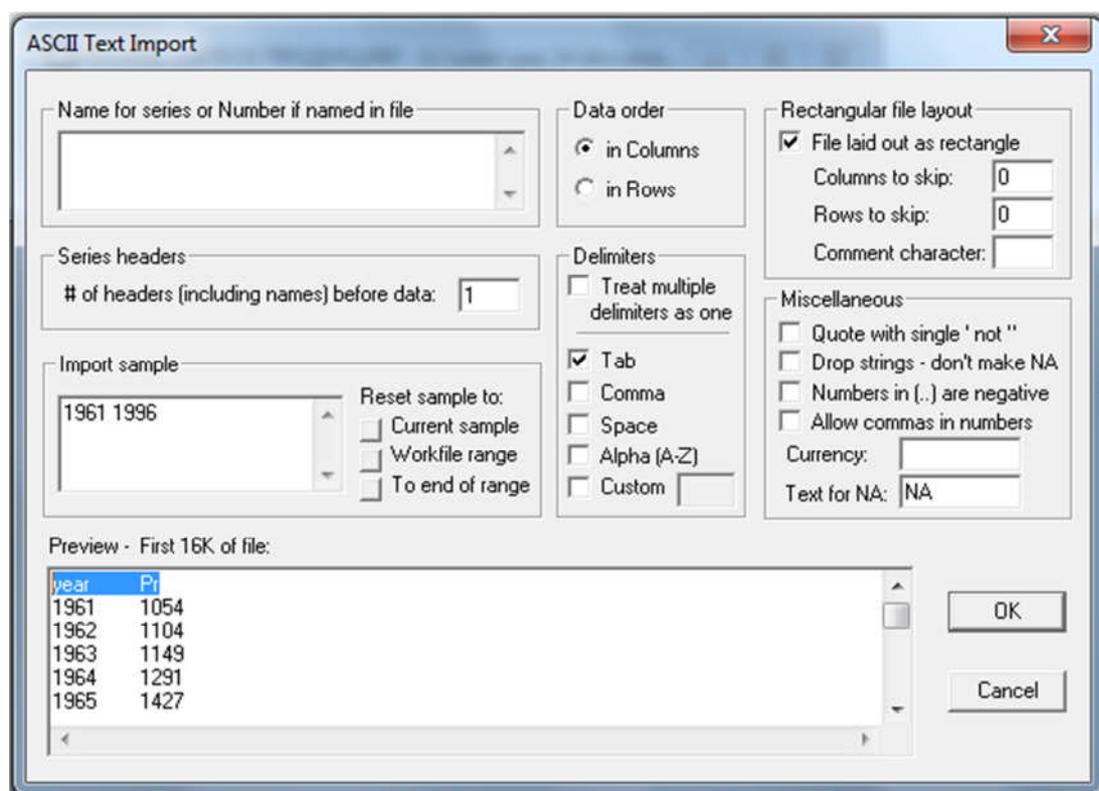
EViews 5. Импорт данных из файла *.txt

Замечание. В файле *.txt для разделения целой и дробной частей должна содержаться десятичная точка, а не запятая.

Импортировать исходные данные из файла *.txt: File/Import/Read Text-Lotus-Excel, в окне Open выбрать имя файла и нажать Открыть, в окне ASCII Text Import ввести необходимые параметры.



Через окно проводника открыть файл *.txt. Откроется окно импорта данных.



Замечание. Имена серий (временных рядов) можно скопировать из поля предпросмотра (Preview ...) в поле **Name for series**. Нажать **ОК**.

Обратить внимание, появятся новые серии значений в окне **Workfile: ИМЯ РАБОЧЕГО ФАЙЛА**. Сохранить созданный рабочий файл в индивидуальной папке **File/Save As...**

ПРИЛОЖЕНИЕ К

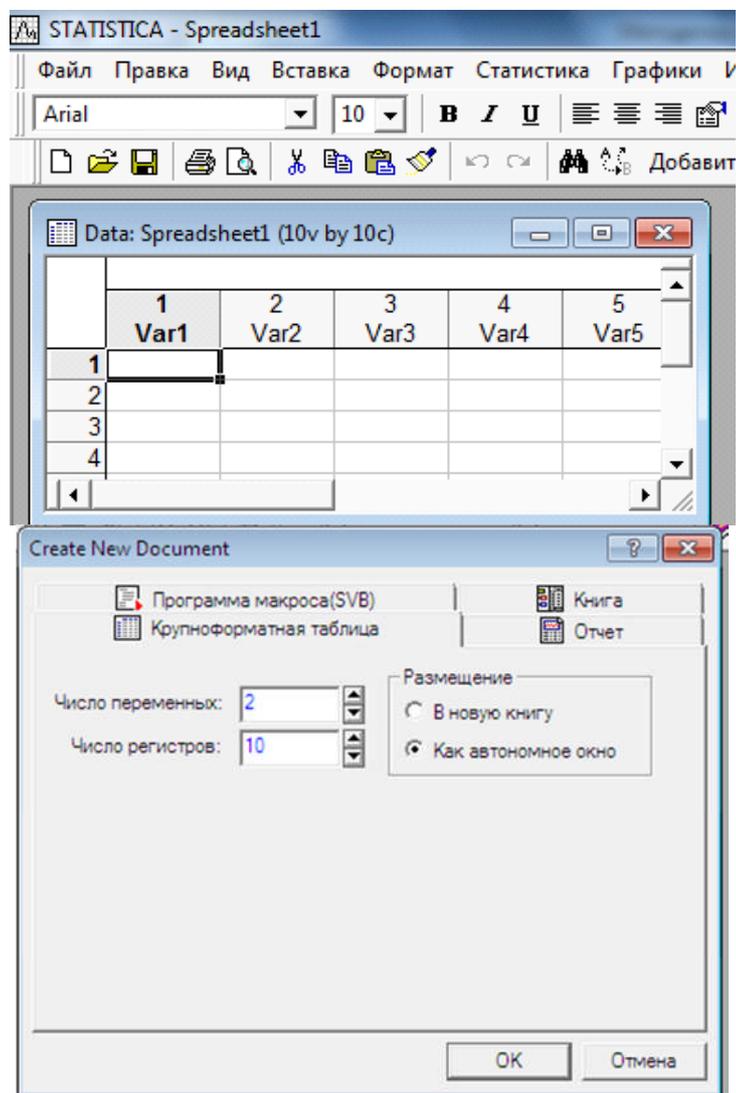
Список обозначений (EViews)

Наименование Рус./англ.	Общепринятые обозначения	Обозначение в Eviews, принятые по умолчанию
Оценки коэффициентов регрессии/ Estimated coefficients	a, b - для парной регрессии b ₀ , b ₁ , b ₂ , ... - для множественной регрессии	c(1), c(2), ...
Стандартные ошибки коэффициентов регрессии / Standard errors	с.о., se	<i>Std. Error,</i> <i>se</i>
Расчетное значение зависимой переменной (y)/ Fitted value	y _p , \hat{y}	<i>yf, yhat</i>
Статистика Фишера / F- statistics	F _{стат}	<i>F-statistic</i>
p-ЗНАЧЕНИЕ - величина, применяемая при проверке гипотез	P-значение, P-value	<i>Prob</i>
Коэффициент детерминации / Coefficient of determination	R ²	R-squared
Скорректированный коэффициент детерминации/ Adjusted coefficient of determination	R ² _{adj}	Adjusted R- squared
Сумма квадратов остатков/ Residual Sum of Squares	RSS, ESS (например, в учебнике Магнуса Я.Р.)	Sum squared resid
Остатки/Residuals	e	Resid
Метод наименьших квадратов / Least squares	МНК	LS

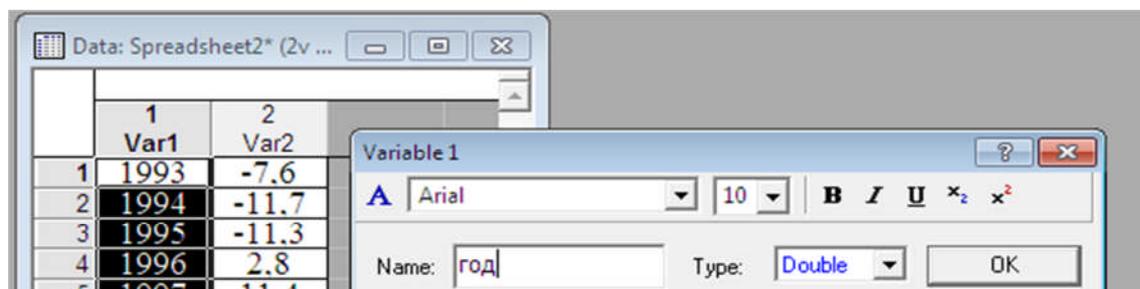
ПРИЛОЖЕНИЕ Л

STATISTICA. Создание файла, содержащего таблицу наблюдений

1. Запустить пакет статистической обработки данных STATISTICA.
2. Создать новую таблицу для необходимого количества переменных и наблюдений: закрыть таблицу *Date: Spreadsheet1...* (нажать [x]), [Файл – Новый] или [Ctrl+N].



3. Переименовать столбцы.



4. Сохранить таблицу: **Файл/Сохранить как.**

ПРИЛОЖЕНИЕ М

Методы определения распределения

Вид, расположение и разброс — три основных характеристики распределения, которые необходимо определить, работая с данными.

Вид распределения важно определить, чтобы знать, какие статистические модели пригодны для обработки данных.

Для контроля качества используют нормальное, биномиальное, гипергеометрическое экспоненциальное распределения, распределения Пуассона, Вейбулла.

Одной из задач статистического анализа является оценка меры соответствия (расхождения) полученных эмпирических данных (выборки) известному теоретическому распределению. В то же время применяемые статистические методы в качестве предпосылок часто требуют определенного закона распределения. Наиболее часто проверяется предположение о нормальном распределении генеральной совокупности. Для этих целей применяют: графический метод, выборочные параметры распределения, критерии согласия.

- *графический метод*, позволяющий визуально дать оценку меры соответствия или расхождения; например, график **выборочной функции распределения** на рис. 1 напоминает форму нормальной кривой и при большом объеме (>50) выборки совпадения/расхождения более очевидны;
- *выборочные параметры формы распределения* (числовые характеристики асимметрию и эксцесс); при большом числе наблюдений ($n > 100$) неплохие результаты дает вычисление асимметрии и эксцесса; предположение о нормальности распределения не противоречит имеющимся данным, если асимметрия и эксцесс близки к нулю (MS Excel, Statistica);
- *критерии согласия*



Рисунок 1 - Сопоставление выборочного распределения и кривой нормального распределения

ПРИЛОЖЕНИЕ Н

Принятие статистических решений

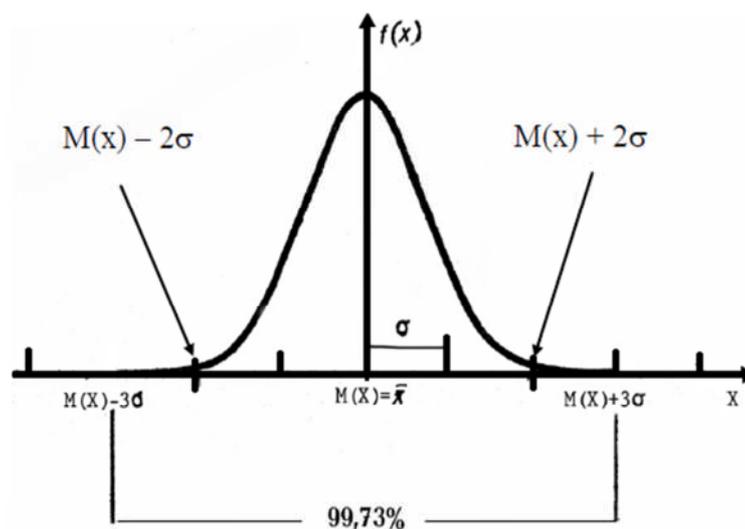
Статистическая гипотеза – это предположение о виде или отдельных параметрах распределения вероятностей, которое подлежит проверке на имеющихся данных.

Проверка статистических гипотез – это процесс формирования решения о возможности принять или отвергнуть утверждение (гипотезу), основанный на информации, полученной из анализа выборки. Методы проверки гипотез называют критериями.

В большинстве случаев рассматривают так называемую нулевую гипотезу (нуль-гипотезу H_0), состоящую в том, что все события произошли случайно, естественным образом. Альтернативная гипотеза (H_1) состоит в том, что события случайным образом произойти не могли, и имело место воздействия некоего фактора.

Обычно нулевая гипотеза формируется таким образом, чтобы на основании эксперимента или наблюдений ее можно было отвергнуть с заранее заданной вероятностью ошибки. Это, заранее заданная вероятность ошибки, называется уровнем значимости. Уровень значимости – максимальное значение вероятности появления события, при котором событие считается практически невозможным. В статистике наибольшее распространение получил уровень значимости, равный $\varepsilon=0,05$. Поэтому если вероятность, с которой интересующее событие может произойти случайным образом $p < 0,05$, то принято считать это событие маловероятным, и если оно все же произошло, то это не было случайным. В наиболее ответственных случаях, когда требуется особая уверенность в достоверности полученных результатов, надежности выводов, уровень значимости принимают равным $\varepsilon=0,01$ или даже $\varepsilon=0,001$.

Величину P , равную $1-\varepsilon$, называют доверительной вероятностью (уровнем надежности), то есть вероятностью, признанной достаточной для того, чтобы уверенно судить о принятом статистическом решении. Соответственно, в качестве доверительных вероятностей выбирают значения 0,95, 0,99, 0,999.



$$P(M(X) - \sigma < X < M(X) + \sigma) = 0,6827 = 68,27 \%$$

$$P(M(X) - 2\sigma < X < M(X) + 2\sigma) = 0,9545 = 95,45 \%$$

$$P(M(X) - 3\sigma < X < M(X) + 3\sigma) = 0,9973 = 99,73 \%$$

ПРИЛОЖЕНИЕ О

Идентификация временного ряда с помощью автокорреляционных функций

Для определения вида и порядка процессов, порождающих стационарный временной ряд, используют аппарат автокорреляционных функций: обычной – **АКФ** и частной – **ЧАКФ**. При этом в целом можно выделить три возможных ситуации:

А. Процесс авторегрессии 1-3 порядка проявляет себя тем, что **АКФ** экспоненциально убывает либо представляет собой смесь синусоиды и убывающей экспоненты (осциллирующее убывание), а **ЧАКФ** имеет одно или несколько (2-3) ненулевых значений, после чего становится неотличимой от **ЧАКФ** белого шума. Количество ненулевых значений **ЧАКФ** дает порядок авторегрессии p .

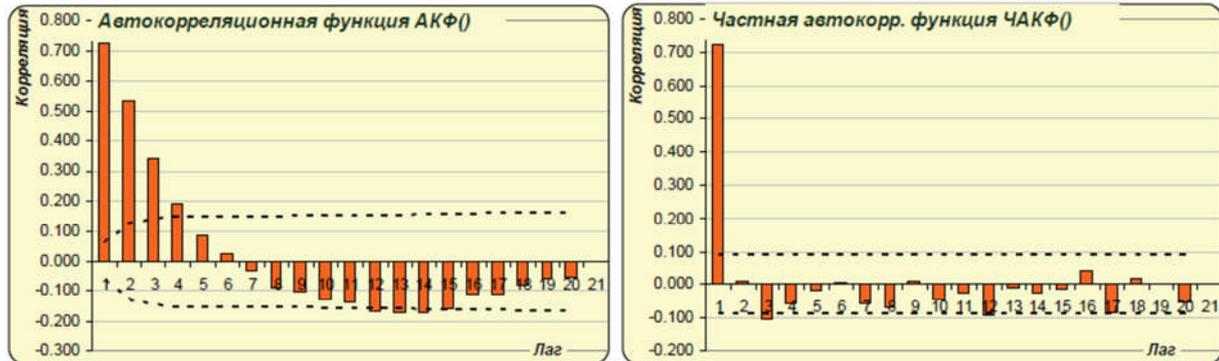


Рисунок 1 - Авторегрессия первого порядка

В. Процесс скользящего среднего 1-3 порядка проявляет себя обратной ситуацией: когда **ЧАКФ** экспоненциально убывает либо представляет собой смесь синусоиды и убывающей экспоненты, а **АКФ**, напротив, имеет одно или несколько ненулевых значений, после чего становится неотличимой от **АКФ** белого шума. В этом случае количество ненулевых значений **АКФ** дает порядок скользящего среднего q .

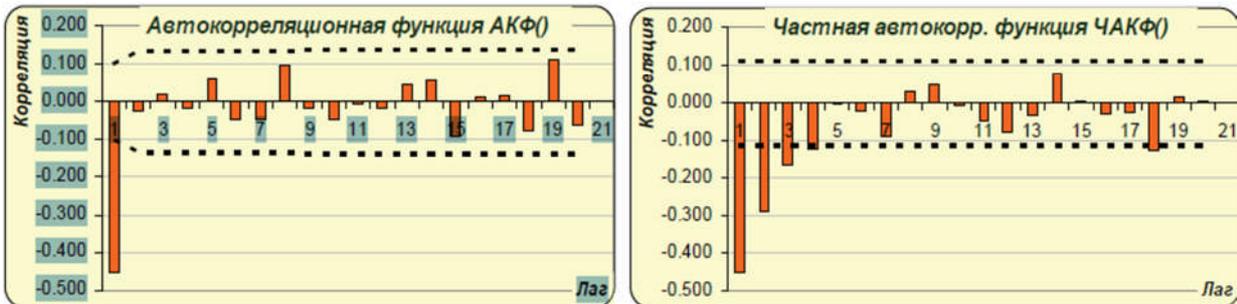


Рисунок 2 – Скользящее среднее третьего порядка

ПРИЛОЖЕНИЕ П

Смесь двух процессов: процесса авторегрессии p -порядка, процесса скользящего среднего q -порядка

Смесь двух процессов приводит к появлению экспонент и синусоид в обеих функциях, однако если вы видите две чистые синусоиды, то это свидетельствует о наличии процесса первого порядка и для авторегрессии, и для скользящего среднего.

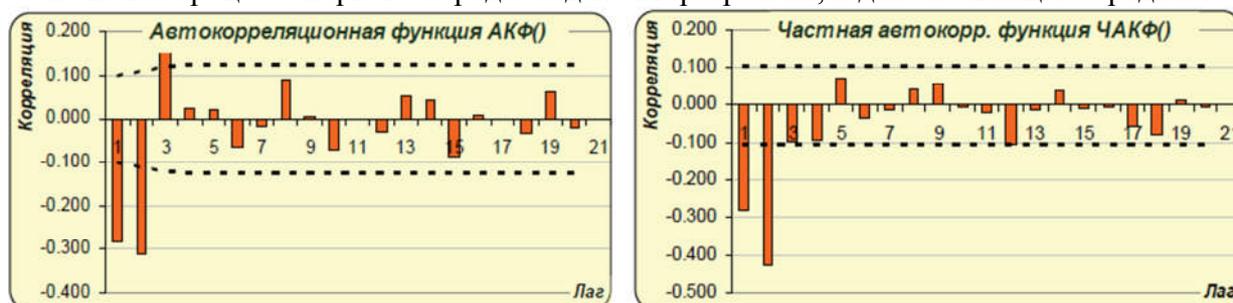


Рисунок 1 – Смесь двух процессов

Другие формы АКФ и ЧАКФ свидетельствуют о том, что во временном ряде имеется тренд (практически неубывающая на отрезке 15-20 значений лага АКФ), либо сезонные колебания (АКФ в виде практически неубывающей синусоиды), от которых следует избавиться, для того чтобы приступить к идентификации внутренних параметров ряда.

Следует еще заметить, что вид АКФ и ЧАКФ зависит от длины временного ряда. Если ряд длинный (несколько сотен наблюдений), форма автокорреляционных функций оказывается более четко выраженной, модель ряда и параметры модели определяются достаточно четко. Если же ряд короткий (несколько десятков наблюдений), коррелограммы становятся смазанными, а идентификация более сложной, т.к. сами функции АКФ и ЧАКФ будут вычислены с гораздо меньшей точностью.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 Эконометрика: Учебник / Под редакцией чл.-корреспондента Российской академии наук И.И. Елисеевой / И.И. Елисеева. – М.;: Финансы и статистика, 2003. – 344 с., ил.
- 2 Практикум по эконометрике: Учеб. Пособие / И.И. Елисеева, С.В. Курышева, Н.М. Гордеенко и др.; Под ред. И.И Елисеевой. – М.;: Финансы и статистика, 2003. – 192 с.: ил.
- 3 Овсянникова М.М. Компьютерный практикум по эконометрике: Для студентов специальности 080109 «Бухгалтерский учет, анализ и аудит», 080105 «Финансы и кредит». – Глазов: Глазовский инженерно-экономический институт, 2011. – 64 с.
- 4 Величко А.С. Изучаем эконометрику. Начальный курс : учебное пособие / А. С. Величко. – Владивосток : Изд-во Дальневост. ун-та, 2007. – 72 с.
- 5 Янковский, И.А. Прикладная эконометрика: методические указания. Ч.1 / И.А. Янковский. – Пинск: ПолесГУ, 2013. – 44 с.
- 6 Эконометрика и экономико-математические методы и модели : пособие для студентов экономических специальностей / авт.-сост. : Л.П. Авдашкова [и др.]. – Гомель : учреждение образования «Бело-русский торгово-экономический университет потребительской кооперации», 2012. – 116 с.
- 7 Бородич С.А. Вводный курс эконометрики: Учебное пособие – Мн.: БГУ, 2000. – 354 с.
- 8 Борздова Т.В. Основы статистического анализа и обработка данных с применением Microsoft Excel : учеб. пособие / Т.В. Борздова. – Минск : ГИУСТ БГУ, 2011. – 75 с.
- 9 Брюков Владимир Как предсказать курс доллара. Эффективные методы прогнозирования с использованием Excel и EViews. - Издательский дом: КНОРУС; ЦИПСИР, 2011.
- 10 Валеев С.Г. Практикум по прикладной статистике : учебное пособие / С.Г. Валеев, В.Н. Клячкин. – Ульяновск : УлГТУ, 2008. – 129 с.: ил.
- 11 Воскобойников Ю.Е. Математическая статистика (с примерами в Excel): учеб. пособие / Ю.Е. Воскобойников, Е.И. Тимошенко ; Новосиб. гос. архитектур.-строит. ун-т (Сибстрин). – 2-е изд., перераб. и доп. – Новосибирск : НГАСУ (Сибстрин), 2006. – 152 с.
- 12 Мастицкий С.Э. Методическое пособие по использованию программы STATISTICA при обработке данных биологических исследований. - Мн.: РУП "Институт рыбного хозяйства". - 76 с.
- 13 Каракозов, С.Г. Введение в эконометрику. Линейные модели: учебное пособие. – Ульяновск: УлГТУ, 2007. – 107 с.
- 13 Определение параметров выборки и проверка гипотезы о нормальности распределения
// Web-сайт “Термист” [Электронный ресурс]. – Режим доступа: http://www.termist.com/bibliot/publik/projekt/10_08_12/10_08_12_05.htm – Дата доступа: 06.11.2014
- 14 Проверка гипотезы о нормальном распределении генеральной совокупности по критерию Пирсона
// ООО «Новый семестр» 2006-2015 [Электронный ресурс]. – Режим доступа: http://math.semestr.ru/group/task_4.php – Дата доступа: 06.11.2014
- 15 Принцип решения задач проверки статистических гипотез Дата добавления: 2014-02-03 [Электронный ресурс]. – Режим доступа:

//Сайт studopedia.ru http://studopedia.ru/2_92477_printsip-resheniya-zadach-proverki-statisticheskikh-gipotez.html – Дата доступа:

06.11.2014

16 Пак Т.В., Еремеева Я.И. Эконометрика. Учебное пособие. – Владивосток: Изд-во Дальневост. ун-та, 2009. -70 с.

17 Шалабанов А.К., Роганов Д.А. Практикум по эконометрике с применением MS Excel. Линейные модели парной и множественной регрессии. – Казань: Академия управления «ГИСБИ», 2008.

18 Шеламова М.А. Статистический анализ медико-биологических данных с использованием программы Excel : учеб.-метод. пособие / М. А. Шеламова, Н. И. Инсарова, В. Г. Лещенко. – Минск : БГМУ, 2010. – 96 с.

19 Лабораторные работы по курсу «Автоматизированные системы обработки биомедицинской информации» /Сост. А.А. Костюкевич, А.М. Криштапович. — Мн.: БГУИР, 2003. — 48 с.: ил.

20 Алёхина А.Э. Эконометрика : учеб. – метод. Пособие / А.Э. Алёхина, С.А. Поттосина. – Минск : БГУИР, 2013. – 98 с. : ил.

21 Поттосина, С.А. Эконометрика: практикум для студентов спец. Системы и технологии в экономике всех форм обуч. / С.А. Поттосина, А.Э. Алёхина. – Минск: БГУИР, 2007. – 71 с.: ил.

22 Шанченко, Н. И. Эконометрика: лабораторный практикум : учебное пособие /Н. И. Шанченко. – Ульяновск : УлГТУ, 2011. – 117 с.

Рева М.Л. Эконометрика: метод, указания и задания к выполнению контрольной работы / М.Л. Рева. - Архангельск: Арханг. гос. техн. ун-т, 2010.-32 с.

23 Эконометрика: практикум / сост. М.Л. Рева; Сев. (Арктич.) федер. ун-т им. М.В. Ломоносова. - Архангельск: ИПЦ САФУ, 2012. - 48 с.

24 Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс: Учеб. — 6-е изд., перераб.и доп. - М.: Дело, 2004. - 576 с.

25 Матюшок В.М., Балашова С.А., Лазанюк И.В. Основы эконометрического моделирования с использованием Eviews: Учебное пособие. – М.: РУДН, 2010. – 164 с.

Учебное издание

Базака Людмила Николаевна
Разинков Александр Иванович

**Статистическая обработка данных в среде пакетов
Statistica, EViews и MS Excel**

Методические указания по выполнению лабораторных работ

Ответственный за выпуск *П.Б. Пигаль*

Публикуется в авторской редакции

Подписано в печать 30.03.2015 г. Формат 60×84/8
Бумага офсетная. Гарнитура «Таймс». Ризография.
Усл. печ. 16,04 л. Уч.-изд. 5,03 л.
Тираж 86 экз. Заказ № 799

Отпечатано в редакционно-издательском отделе
Полесского государственного университета.
225710, г. Пинск, ул. Днепровской флотилии, 23.