МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
Учреждение образования
«Витебский государственный технологический университет»

ПРИКЛАДНОЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ ЭКОНОМИЧЕСКИХ ДАННЫХ

Методические указания
к практическим занятиям для студентов специальностей
1-28 01 01 «Экономика электронного бизнеса»,
6-05-0611-04 «Электронная экономика»

Витебск
2025

УДК 311.3/.4

Составитель:

К. Р. Быков

Одобрено кафедрой «Экономика и электронный бизнес»
УО «ВГТУ», протокол № 5 от 05.12.2024.

Рекомендовано к изданию редакционно-издательским
советом УО «ВГТУ», протокол № 4 от 23.12.2024.

Прикладной статистический анализ экономических данных : методические указания к практическим занятиям / сост. К. Р. Быков. – Витебск : УО «ВГТУ», 2025. – 111 с.

В методических указаниях содержатся задания к практическим занятиям по всем темам дисциплины, предусмотренным учебной программой курса.

УДК 311.3/.4

© УО «ВГТУ», 2025

ОГЛАВЛЕНИЕ

Практическое занятие 1. «Основы прикладного статистического анализа экономических данных» 4
Практические занятия 2–4. «Множественный корреляционно-регрессионный анализ» 19
Практические занятия 5–7. «Снижение размерности признакового пространства с использованием статистических программных пакетов» 52
Практическое занятие 8. «Кластерный анализ многомерных наблюдений с использованием статистических программных пакетов» 67
Список использованных источников 88
Приложение А. Пошаговый алгоритм выполнения корреляционно-регрессионного анализа в MS Excel 92
Приложение Б. Показатели регрессионной статистики в программе MS Excel 93
Приложение В. Основы прикладного статистического анализа экономических данных 94
Приложение Г. Матричная алгебра 95
Приложение Д. Распределение Фишера (*F*-распределение) 97
Приложение Е. Распределение Дарбина – Уотсона (Durbin – Watson) 102
Приложение Ж. Исходные данные организаций для выполнения прикладного статистического анализа 108
Приложение И. Распределение *t*-критерия Стьюдента при уровне значимости 0,10, 0,05 и 0,01 109
Приложение К. Исходные данные коммерческой организации для выполнения линейного множественного корреляционно-регрессионного анализа 110

ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 1
«Основы прикладного статистического анализа экономических данных»

Цель занятия: получить навыки предварительного анализа данных; изучить характеристики генеральной и выборочной совокупности; получить навыки выполнения корреляционного и регрессионного анализа и оценки их результатов с использованием MS Excel.

Задания по теме

Задание 1.1. По 20 предприятиям имеются следующие данные для линейного корреляционно-регрессионного анализа в MS Excel (табл. 1.1).

Таблица 1.1 – Исходные данные для корреляционно-регрессионного анализа

№ предприятия	Производительность труда на одного рабочего, тыс. руб.	Энерговооруженность труда на одного рабочего, кВт. ч	% рабочих ручного труда к общей численности
n	Y	X_1	X_2
1	9,8	4,8	40
2	6,7	2,8	50
3	12,4	7,0	38
4	6,9	3,8	57
5	11,8	5,5	31
6	7,3	3,0	56
7	8,4	3,4	45
8	10,7	5,2	35
9	11,1	5,4	32
10	7,3	2,9	54
11	8,8	3,9	43
12	10,0	5,0	36
13	12,1	6,2	26
14	11,0	5,1	32
15	6,8	2,7	55
16	7,0	2,9	50
17	12,0	6,0	30
18	7,5	3,3	48
19	8,0	3,4	45
20	9,2	3,8	49

1. Рассчитайте линейные парные коэффициенты корреляции всеми возможными способами и постройте матрицу парных коэффициентов корреляции.

2. Оцените значимость коэффициента корреляции по t-критерию Стьюдента (*t*-статистика).

3. Выберите фактор, наиболее тесно связанный с результативным

признаком, и постройте линейное парное уравнение регрессии.

4. Оцените качество модели парной линейной регрессии на основе коэффициента детерминации.

5. Проверьте значимость линейной парной модели регрессии по *F*-критерию Фишера.

6. Рассчитайте коэффициент эластичности.

7. Поясните экономический смысл полученных результатов статистического анализа.

8. Выполните статистический анализ на рабочих листах в программе MS Excel по формулам с помощью методических указаний к заданию.

Методические указания к выполнению задания 1.1

ШАГ 1. Рассчитайте линейные парные коэффициенты корреляции и постройте матрицу парных коэффициентов корреляции.

Для расчета парного коэффициента корреляции наиболее удобной являются следующие формулы:

$$r_{xy} = \frac{\sum xy - \frac{\sum y \sum x}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]\left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}} \quad (1.1)$$

или

$$r_{xy} = \frac{\overline{xy} - \bar{x}\,\bar{y}}{\sigma_x\,\sigma_y} = a_1 \frac{\sigma_x}{\sigma_y}, \quad (1.2)$$

где r_{xy} – коэффициент корреляции изменяется от -1 до +1 и показывает тесноту и направление корреляционной связи; значения Σy, Σx, Σxy, Σx^2 и Σy^2 рассчитывают по данным наблюдения; σ_x – среднее квадратическое (стандартное) отклонение факторного признака; σ_y – среднее квадратическое (стандартное) отклонение результативного признака.

Чтобы рассчитать парные коэффициенты корреляции, постройте расчетную таблицу 1.2.

Таблица 1.2 – Вспомогательные данные для поиска решения

№ предприятия	Исходные данные			Расчетные данные					
n	*Y*	X_1	X_2	Y^2	X_1^2	X_2^2	YX_1	YX_2	X_1X_2
1									
...									
20									
Итого									
Среднее									
Дисперсия σ^2									
СКО σ									

Подставьте данные из таблицы 1.2 в формулы коэффициентов корреляции (1.1).

Результаты расчетов парных коэффициентов корреляции представьте в виде следующей корреляционной матрицы (табл. 1.3).

Таблица 1.3 – Корреляционная матрица

Переменная	y	x_1	x_2
y			
x_1			
x_2			

Сделайте выводы.

ШАГ 2. Оцените значимость коэффициента корреляции.

Оценка значимости (существенности) линейного парного коэффициента корреляции основана на сопоставлении значения r с его средней квадратической ошибкой (σ_r), таблица 1.4.

Таблица 1.4 – Оценка значимости (существенности) линейного коэффициента корреляции на основе средней квадратической ошибки

Объем выборки (число единиц наблюдения)	Формула средней ошибки коэффициента корреляции	Характеристика значений
при $n > 50$	$\sigma_r = \frac{1 - r^2}{\sqrt{n}}$	Если при этом коэффициент корреляции r превышает свою среднюю ошибку σ_r больше чем в 3 раза, т. е. если $\frac{\lvert r \rvert}{\sigma_r} > 3$, то он считается значимым, а связь – реальной.
при $n < 30$	$\sigma_r = \frac{\sqrt{1 - r^2}}{\sqrt{n - 2}}$	Значимость коэффициента корреляции проверяется на основе t-критерия Стьюдента. Для этого рассчитывается фактическое (расчетное) значение критерия: $t_{r\ \text{факт}} = \lvert r \rvert \sqrt{\frac{n-2}{1-r^2}}$, которое сопоставляется с критическим (табличным) значением $t_{табл}$ определяемым по приложению для числа степеней свободы ν = n – 2 и заданного уровня значимости (обычно α = 0,05). Если $t_{расч} > t_{табл}$ коэффициент парной корреляции считается значимым, а связь – реальной.

ШАГ 3. Выберите фактор, наиболее тесно связанный с результативным признаком, и постройте линейное парное уравнение регрессии. Сделайте выводы.

Для разработки линейного парного уравнения регрессии используйте систему нормальных уравнений, составленных в результате применения метода наименьших квадратов (МНК), формула (1.3):

$$\begin{cases} a_0 n + a_1 \sum x = \sum y \\ a_0 \sum x + a_1 \sum x^2 = \sum y\, x \end{cases} \quad (1.3)$$

Необходимые для решения данной системы показатели n, Σy, Σx, Σxy, и Σx^2 определяются по расчетным данным таблицы 1.2. Решить такую систему нормальных уравнений можно разными способами.

Для нахождения a_0 и a_1 при линейной зависимости можно воспользоваться формулами:

$$a_1 = \frac{\overline{x\,y} - \bar{x}\bar{y}}{\sigma_x^2} \quad \text{или} \quad a_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}, \quad (1.4)$$

где a_1 – коэффициент (параметр) уравнения регрессии.

Параметр a_0 уравнения регрессии определяется как

$$a_0 = \bar{y} - a_1 \bar{x} \quad (1.5)$$

Линейное парное уравнение регрессии примет вид:

$$y_x = a_0 + a_1 x \quad (1.6)$$

Основной смысл регрессионного анализа состоит в том, чтобы по полученному уравнению регрессии найти теоретические уровни $\bar{y}_x$, которые могут служить прогнозируемыми показателями на будущий период.

Замечание. Для парной линейной регрессии проверка гипотезы о значимости параметра a_1 и коэффициента корреляции r_{xy} равносильна проверке гипотезы о существенности уравнения регрессии в целом, т. е. $t_{a1} = t_r = \sqrt{F}$.

ШАГ 4. Оцените качество модели парной линейной регрессии на основе коэффициента детерминации. Сделайте вывод.

Общее качество уравнения регрессии оценивается по тому, как хорошо эмпирическое уравнение регрессии согласуется со статистическими данными. Суммарной мерой общего качества уравнения регрессии (соответствия уравнения регрессии статистическим данным) является коэффициент детерминации.

Для парной регрессии $R^2 = r^2_{xy}$.

Модель с коэффициентом детерминации выше 80 % можно признать достаточно хорошей (коэффициент корреляции превышает 90 %).

ШАГ 5. Проверьте значимость модели парной линейной регрессии по *F*-критерию Фишера (*F-тест*). Сделайте вывод.

Значение F-критерия Фишера для парной линейной регрессии можно рассчитать по формуле:

$$F = \frac{r^2}{1 - r^2}(n-2) \quad \text{или} \quad F = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 / k}{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 / (n-k-1)}, \quad (1.7)$$

где n – число (количество) данных в выборке; k – количество регрессоров (для парной регрессии k = 1).

Величина F имеет распределение Фишера с $\nu_1 = 1$, $\nu_2 = n - 2$ степенями свободы. Вычисленное значение F-критерия признается достоверным (отличающимся от единицы), если оно больше табличного (критического). В этом случае нулевая гипотеза (H_0) отвергается и делается вывод о существенности изучаемой связи: $F_{факт} > F_{табл}$.

ШАГ 6. Средний коэффициент эластичности для уравнения прямой определяется по формуле:

$$\overline{Э} = a_1 \frac{\bar{x}}{\bar{y}} \qquad (1.8)$$

Коэффициент эластичности показывает, на сколько процентов изменяется результативный признак при изменении факторного признака на 1 %.

ШАГ 7. Поясните экономический смысл полученных результатов статистического анализа.

Задание 1.2. На основе построенного уравнения регрессии в задании 1.1. выполните следующие шаги статистического анализа.

1. Рассчитайте по построенному уравнению регрессии теоретические уровни (значения) $\hat{y}_x$ и сделайте вывод о правильности выполненных расчетов.
2. Оцените качество регрессионной модели, используя среднюю относительную ошибку аппроксимации.
3. Представьте графически корреляционное поле и линию регрессии эмпирических и теоретических (сглаженных) данных.
4. Определите точечный прогноз производительности труда рабочего.
5. Выполните все этапы статистического анализа в Excel на рабочих листах в соответствии с требованиями методических указаний к заданию.

Методические указания к выполнению задания 1.2

ШАГ 1. На основе построенного уравнения регрессии исчислите теоретические уровни и сделайте вывод о правильности выполненных расчетов.

Выполнение данного этапа статистического анализа основано на результатах задания 1.1 (ШАГИ 1 и 2).

Таблица 1.5 – Исходные и расчетные данные для оценки качества уравнения парной регрессии

№ предприятия	Исходные данные	Расчетные данные		
n	y_i	$\hat{y}_i =$	$(y_i - \hat{y}_i)$	$(y_i - \hat{y}_i)^2$
1				
...				
20				
Итого				

$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ – является остаточная колеблемость общей колеблемости y.

ШАГ 2. Для оценки качества (уровень точности) регрессионных моделей используется средняя относительная ошибка аппроксимации, формула:

$$E_{\text{отн}} = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i} \times 100\ \% = \frac{1}{n}\sum_{i=1}^{n}\frac{|e_i|}{y_i} \times 100\ \%, \qquad (1.9)$$

где $|y_i - \hat{y}_i| = |e_i|$ – абсолютные остатки; y_i – зависимые переменные $(i = \overline{1,n})$ или исходные экспериментальные значения; $\hat{y}_i$ – значения y, вычисленные по регрессионной модели.

Если $E_{\text{отн}} < 7\ \%$ свидетельствует о хорошем качестве (точности) регрессионной модели. Допустимый предел значений $E_{\text{отн}}$ не более 8–10 % (или не превышает 10–12 %).

ШАГ 3. Представьте графически корреляционное поле и линию регрессии эмпирических и теоретических (сглаженных) данных. Определите точечный прогноз производительности труда рабочего.

График «поле корреляции» в программе *Excel* строится с использованием точечной диаграммы, выбираем вкладку основного меню «Вставка», на панели инструментов находим область для построения диаграмм и кликаем по наименованию «точечная» (рис. 1.1).

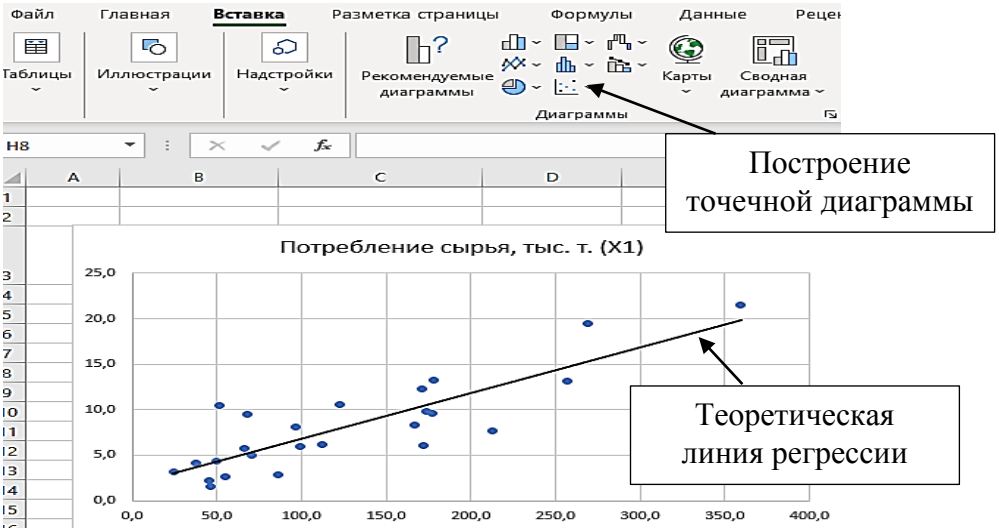


Рисунок 1.1 – Зависимость выпуска продукции (y) от потребления сырья (x)

Для построения теоретической линии регрессии кликните левой кнопкой мыши по точкам на диаграмме, на панели инструментов выберете область «Макеты диаграмм», кликните по названию «Линия тренда», в списке линий выберете «Линейная».

Задание 1.3. На основе полученных результатов статистического анализа данных в задании 1.1 выполните следующие шаги статистического анализа.

1. Рассчитайте стандартные ошибки коэффициентов уравнения регрессии и стандартную ошибку регрессии.

2. Оцените статистическую значимость коэффициентов (параметров) парной линейной регрессии.

3. Постройте интервальные оценки (доверительные интервалы) для параметров уравнения регрессии на уровне значимости 0,05.

4. Выполните все этапы статистического анализа данных в программе Excel на рабочих листах по формулам, представленным в методических указаниях к заданию.

Методические указания к выполнению задания 1.3

ШАГ 1.1. Стандартные ошибки коэффициентов (параметров) уравнения регрессии рассчитайте по формулам:

$$S_{a_0} = S\sqrt{\frac{\overline{x^2}}{ns_x^2}}, \tag{1.10}$$

$$S_{a_1} = \frac{S}{\sqrt{ns_x^2}} = \frac{S}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}, \tag{1.11}$$

где S_{a_0}, S_{a_1} – стандартные ошибки коэффициентов (параметров) линейного уравнения регрессии; S – стандартная ошибка регрессии (см. ШАГ 1.2); $\overline{x^2} = \frac{\sum_{i=1}^{n} x_i^2}{n}$ средний квадрат переменной X; $S_x^2 = \overline{x^2} - \bar{x}^2$ – дисперсия переменной X (выборочная дисперсия переменной X);

ШАГ 1.2. Стандартную ошибку регрессии (S) вычислите по формуле:

$$S = \sqrt{S^2} = \sqrt{\frac{1}{n-2}\sum(y_i - \hat{y}_i)^2} = \sqrt{\frac{\sum e_i^2}{n-2}}, \tag{1.12}$$

где S^2 – остаточная дисперсия; $\sum e_i^2 = \sum(y_i - \hat{y}_i)^2$ – остаточная (необъясненная) сумма квадратов, являющаяся мерой остаточного не объясненного уравнением регрессии разброса (разброса точек вокруг линии регрессии); y_i – наблюдаемые значения зависимой переменной Y; $\hat{y}_i$ – рассчитанные значения по уравнению регрессии.

Для расчета стандартной ошибки регрессии (S) составьте вспомогательную таблицу (см. методические указания к заданию 1.2, ШАГ 1).

ШАГ 2. Оценку значимости отдельных коэффициентов (параметров) парной линейной регрессии рассчитайте с помощью t-статистики Стьюдента по формулам:

$$t_{a_0} = \frac{a_0}{S_{a_0}} \quad \text{и} \tag{1.13}$$

$$t_{a_1} = \frac{a_1}{S_{a_1}} = \frac{a_1}{S}\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \qquad (1.14)$$

где a_0 и a_1 – коэффициенты (параметры) уравнения регрессии.

Для проверки нулевой гипотезы H_0 используется расчетная статистика, имеющая распределение Стьюдента с $\nu = n - 2$ степенями свободы в случае парной линейной регрессии. Из таблиц распределения Стьюдента (*t*-распределения) находят критическое значение $t_{крит}$ при заданном уровне значимости α.

Из сравнения расчетной статистики с критическим значением при заданном уровне значимости α получаем: если $|t_{расч}| < t_{крит}$, то H_0 принимается, т. е. коэффициент a_1 статистически незначим; если $|t_{расч}| > t_{крит}$, то H_0 отклоняется, т. е. коэффициент a_1 статистически значим.

Критическое значение $t_{крит}$ определяется также в Excel с помощью статистической функции $t_{крит}$ = СТЬЮДРАСПОБР (α, ν) (рис. 1.3).

ШАГ 3. Доверительные интервалы для коэффициентов уравнения регрессии определите по формулам:

$$a_0 - t_{крит}S_{a0} < a_0 < a_0 + t_{крит}S_{a0} \text{ или } a_0 - \Delta_{a0} < a_0 < a_0 + \Delta_{a0}, \qquad (1.15)$$

$$a_1 - t_{крит}S_{a1} < a_1 < a_1 + t_{крит}S_{a1} \text{ или } a_1 - \Delta_{a1} < a_1 < a_1 + \Delta_{a1}, \qquad (1.16)$$

где a_0 и a_1 – коэффициенты регрессии; S_{a_0}, S_{a_1} – стандартные ошибки коэффициентов (параметров) линейного уравнения регрессии; $t_{крит} = t\,(1 - \alpha;\ n - 2)$ – критическое значение статистики Стьюдента (см. ШАГ 2); $\Delta_{a0} = t_{крит}S_{a0}$ – предельная ошибка параметра a_0 линейной регрессии; $\Delta_{a1} = t_{крит}S_{a1}$ – предельная ошибка параметра a_1 линейной регрессии.

Замечание. Доверительный интервал оценки для коэффициентов уравнения регрессии определяет границы, в которых истинное значение коэффициентов находится с вероятностью $1 - \alpha$.

Если в границы доверительного интервала попадает ноль, т. е. нижняя граница отрицательна, а верхняя положительна, то оцениваемый параметр принимается равным нулю, так как он одновременно не может принимать и положительное, и отрицательное значения. Тем самым он признается неадекватным.

ШАГ 4. Поясните экономический смысл полученных результатов статистического анализа данных.

Задание 1.4. На основе построенного линейного уравнения регрессии в задании 1.1. выполните следующие шаги статистического анализа.

1. Рассчитайте прогнозное значение результата (производительность труда рабочего), если прогнозное влияние фактора $x_{прогн}$ увеличится на 25 % от его среднего уровня.
2. Оцените точность прогноза, рассчитав ошибку прогноза.

3. Определите 95%-й доверительный интервал для индивидуального прогнозного значения $\hat{y}_{\text{прогн}}$.

4. Выполните все этапы статистического анализа в MS Excel на рабочих листах по формулам, представленным в методических указаниях к заданию.

Методические указания к выполнению задания 1.4

ШАГ 1. Прогноз $\hat{y}_{\text{прогн}}$ для $x_{\text{прогн}}$ постройте, используя полученное ранее уравнение регрессии (см результаты уравнения регрессии в задании 1.1).

$$x_{\text{прогн}} = x_{\text{средн}} \times \text{К}_{\text{роста}}, \tag{1.17}$$

где $\text{К}_{\text{роста}}$ – коэффициент роста (снижения) независимого признака X; $\bar{x} = x_{\text{средн}}$ – среднее значение независимого признака X.

Прогнозное значение $\hat{y}_{\text{прогн}}$ определите на основе уравнения регрессии, полученного ранее (см. результаты уравнения регрессии в задании 1.1) и ШАГА 1. по формуле:

$$\hat{y}_{\text{прогн}} = a_0 + a_1 x_{\text{прогн}}, \tag{1.18}$$

где $\hat{y}_{\text{прогн}}$ – предсказанное значение (точечный прогноз) на основе уравнения регрессии.

ШАГ 2. Стандартную ошибку прогноза (предсказания) для индивидуального значения $\hat{y}_{\text{прогн}}$ рассчитайте по формуле:

$$S_{\hat{y}\ \text{прогн}} = S\sqrt{1 + \frac{1}{n} + \frac{\left(x_{\text{прогн}} - \bar{x}\right)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}, \tag{1.19}$$

где S – стандартная ошибка регрессии (модели) (см. методические указания к заданию 1.3); $x_{\text{прогн}}$ – новое значение признака, для которого дается прогноз; x_i – значения независимого признака X; n – количество наблюдений, использованных для регрессионного анализа.

ШАГ 3. Доверительный интервал для индивидуального прогнозного значения $\hat{y}_{\text{прогн}}$ вычислите на основе выражения:

$$\hat{y}_{\text{прогн}} - t_{\text{крит}} S_{\hat{y}\ \text{прогн}} < \gamma_{\hat{y}_{\text{прогн}}} < \hat{y}_{\text{прогн}} + t_{\text{крит}} S_{\hat{y}\ \text{прогн}}, \tag{1.20}$$

или

$$\hat{y}_{\text{прогн}} - \Delta_{\hat{y}\ \text{прогн}} < \gamma_{\hat{y}_{\text{прогн}}} < \hat{y}_{\text{прогн}} + \Delta_{\hat{y}\ \text{прогн}}, \tag{1.21}$$

где $S_{\hat{y}\ \text{прогн}}$ – стандартная ошибка прогноза для индивидуального значения $\hat{y}_{\text{прогн}}$; $t_{\text{крит}} S_{\hat{y}\ \text{прогн}} = \Delta_{\hat{y}\ \text{прогн}}$ – предельная ошибка прогноза, которая

в 95 % случаев не будет превышена; $t_{крит}$ – критическое значение t-статистики Стьюдента при заданном уровне значимости (1 – α) и числе степеней свободы, которое в случае парной линейной регрессии составляет $\nu = n - 2$. (см. методические указания к заданию 1.3).

Точность выполненного прогноза проверьте на основе диапазона верхней и нижней границ доверительного интервала D_γ по формуле:

$$D_\gamma = \frac{\gamma_{\hat{y}\max}}{\gamma_{\hat{y}\min}} \tag{1.22}$$

ШАГ 4. Поясните экономический смысл полученных результатов статистического анализа и оформите выводы в аналитической записке.

Задание 1.5. На основе результатов в задании 1.1 выполните следующие шаги статистического анализа.

1. Оцените среднюю производительность труда на одного рабочего для предприятий с энерговооруженностью труда на одного рабочего 3,8 кВт. ч.
2. Оцените точность прогноза, рассчитав ошибку прогноза.
3. Определите 95%-й доверительный интервал для среднего значения производительности труда на одного рабочего для таких же предприятий.
4. Выполните все этапы статистического анализа в Excel на рабочих листах по формулам, представленным в методических указаниях к заданию.

Методические указания к выполнению задания 1.5

ШАГ 1. Среднюю производительность труда на одного рабочего для предприятий с энерговооруженностью труда на одного рабочего 3,8 кВт. ч оценим с помощью полученного ранее уравнения регрессии (см. результаты уравнения регрессии в задании 1.1).

Выборочной оценкой является $x = 3{,}8$, которую найдем по уравнению регрессии (см. результаты уравнения регрессии в задании 1.1):

$$\hat{y}_{x=3,8} = a_0 + a_1 x \ , \tag{1.23}$$

где $\hat{y}_{x=3,8}$ – прогноз для $x = 3{,}8$, используя полученное ранее уравнение регрессии.

ШАГ 2. Стандартную ошибку прогноза для среднего значения $\hat{y}_x$ рассчитайте по формуле:

$$S_{\hat{y}_x} = S\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}, \tag{1.24}$$

где S – стандартная ошибка регрессии (модели) (см. методические указания к заданию 1.3); x – значение признака, для которого дается прогноз (см. условие

задания 1.5 и ШАГ 1); $\bar{x}$ – среднее значение независимого признака X; x_i – значения независимого признака X; n – количество наблюдений, использованных для регрессионного анализа.

ШАГ 3. Доверительный интервал для уравнения регрессии Y рассчитайте на основе выражения:

$$\hat{y}_x - t_{\text{крит}} S_{\hat{y}\,x} < Y < \hat{y}_x + t_{\text{крит}} S_{\hat{y}\,x}\,, \tag{1.25}$$

или

$$\hat{y}_x - \Delta_{\hat{y}\,x} < Y < \hat{y}_x + \Delta_{\hat{y}\,x}\,, \tag{1.26}$$

где $S_{\hat{y}\,x}$ – стандартная ошибка прогноза для уравнения регрессии Y; $t_{\text{крит}}$ – критическое значение t-статистики Стьюдента при заданном уровне значимости $(1 - \alpha)$ и числе степеней свободы, которое в случае парной линейной регрессии составляет $\nu = n - 2$ (см. методические указания к заданию 1.3); $t_{\text{крит}} S_{\hat{y}\,x} = \Delta_{\hat{y}\,x}$ – предельная ошибка прогноза.

ШАГ 4. Поясните экономический смысл полученных результатов статистического анализа.

Задание 1.6. По приведенным данным в задании 1.1 выполните линейный парный корреляционно-регрессионный анализ с применением специальных инструментов анализа Excel на рабочих листах и методических указаниях.

Методические указания к выполнению задания 1.6

ШАГ 1. Для проведения линейного парного корреляционно-регрессионного анализа в MS Excel предварительно установите пакет анализа, выполнив последовательно действия: *Файл – Параметры (Excel) – Надстройки – Перейти* (выделим в окне доступных надстроек *Пакет анализа*), после этого на вкладке *Данные* ленты появится команда *Анализ данных*.

ШАГ 2. Для иллюстрации выполнения задания 1.6 в Excel рассмотрим выборку из $n = 50$ наблюдений. Целью является пошаговый статистический анализ зависимости объема реализованной продукции (Y, тыс. руб.) от цены товара (фактор X, руб.).

Пошаговые результаты работы инструментов анализа *Корреляция* и *Регрессия* в MS Excel представлены в приложении А:

Пояснение показателей линейного корреляционно-регрессионного анализа в программе MS Excel представлено в приложении Б.

ШАГ 3. Для рассматриваемого примера уравнение регрессии примет вид:

$$\hat{y}_x = 4059{,}7 - 245{,}4x \tag{1.27}$$

ШАГ 3.1. Проанализируем значения показателей, приведенных в таблице «Регрессионная статистика» и наименованием показателей в MS Excel

из приложения Б:

– множественный R (в нашем примере парный коэффициент корреляции) = 0,840, т. е. наблюдается тесная зависимость объема реализованной продукции от выделенного признака-фактора (цены);

– R-квадрат = 0,706, т. е. на 70,6 % вариация объема реализованной продукции обусловлена влиянием цен;

– нормированный R-квадрат = 0,699, что также свидетельствует об адекватности уравнения регрессии;

– стандартная ошибка регрессии = 52,3 тыс. руб. Чтобы ответить на вопрос о возможности использования полученного уравнения регрессии на практике, сравним стандартную ошибку регрессии с величиной среднего квадратического отклонения результативного признака. Для данных σ_y= 8946,1 тыс. руб., следовательно, использование уравнения регрессии является целесообразным.

ШАГ 3.2. Оценим адекватность модели с помощью дисперсионного анализа. Наблюдаемое значение F и рассчитанный уровень значимости гипотезы H_0 представлены в таблице «Дисперсионный анализ» (приложение А). Так как $F_{факт} > F_{табл}$, ($F_{факт}$ = 115,3; $F_{табл}$ = 115,04) а значимость F = 2,32863E-14 меньше заданного уровня значимости $\alpha = 0,05$, гипотеза H_0 отклоняется, т. е. уравнение регрессии статистически адекватно.

Табличное (критическое) значение *F*-критерия можно найти также с использованием функции FРАСПОБР в MS Excel, (рис. 1.2).

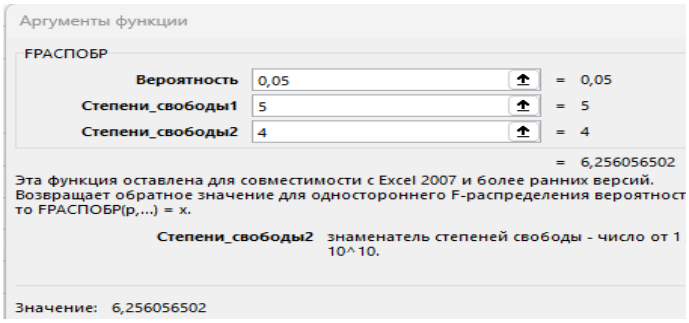


Рисунок 1.2 – Расчет табличного значения *F*-распределения Фишера в MS Excel

Если расчетное значение больше табличного ($F_{расч} > F_{табл}$) при заданном уровне значимости, то уравнение регрессии следует признать значимым, т. е. его можно использовать для анализа и прогнозирования.

ШАГ 3.3. Проверим значимость параметров уравнения регрессии с использованием значений T и рассчитанных уровней значимости гипотезы H_0, представленных в столбце *t*-статистика и *p*-значение (приложение А).

Для нашего примера (прилож. А) все расчетные значения *t*-критерия больше $t_{кр}$= 2,0106 (α = 0,05; ν = 50 – 2 = 48), что свидетельствует о значимости параметров уравнения регрессии. Табличное значение t-статистики вычислим с помощью функции СТЬЮДРАСПОБР в MS Excel, рисунок 1.3.

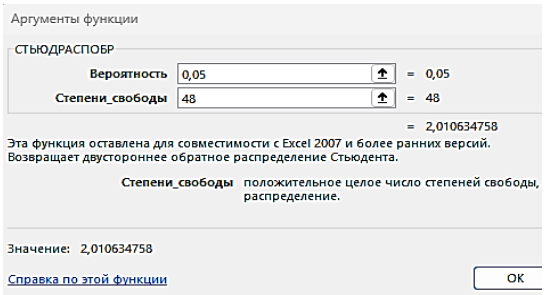


Рисунок 1.3 – Диалоговое окно функции СТЬЮДРАСПОБР

Задания для самостоятельного решения

Задание 1.7. Составьте по статистическому анализу заданий 1.1–1.6 итоговый отчет результатов, полученных по формулам и с помощью применения специальных инструментов Excel. Составьте итоговый отчет и сделайте вывод.

Методические указания к выполнению задания 1.7

ШАГ 1. Итоговый отчет результатов статистического анализа представьте на рабочих листах MS Word в таблицу 1.6

Таблица 1.6 – Итоговый отчет результатов линейного парного корреляционно-регрессионного анализа экономических данных

Название этапа анализа	Значение, полученное по формуле	Значение, полученное по функции в Excel	Вывод

ШАГ 2. Общий вывод по результатам линейного парного корреляционно-регрессионного анализа экономических данных.

Задание 1.8. Приведите типы статистических данных согласно представленным критериям:

Критерий классификации	Типы статистических данных
По способу получения информации	
По числу переменных	
По упорядочению во времени	
По типу шкалы измерения	

Воспользовавшись статистическими данными из официальных источников, найдите каждому типу данных по одному примеру. Пример на каждый тип данных должен сопровождаться ссылкой на источник, таблицей значений и графиком. Результаты представьте на рабочих листах MS Word.

Методические указания к выполнению задания 1.8

В качестве источников для поиска данных можно использовать белорусские статистические и информационные официальные сайты:

– Национальный статистический комитет Республики Беларусь: https://www.belstat.gov.by;

– Национальный банк Республики Беларусь – https://www.nbrb.by;

– Министерство экономики Республики Беларусь: https://economy.gov.by/ru;

– Министерство финансов Республики Беларусь: https://www.minfin.gov.by;

– Министерство связи и информатизации Республики Беларусь: https://www.mpt.gov.by и др.

Задание 1.9. Раскройте суть понятий нормирования (стандартизации) и унификации данных:

Наименование	Сущность понятий
Стандартизация данных	
Центрирование данных	
Нормирование данных	
Унификация данных	

Приведите данные X: 20, 40, 60, 80 к стандартизованному виду с использованием MS Excel.

Методические указания к выполнению задания 1.9

Для стандартизации данных расчет статистических показателей в MS Excel составьте в следующую таблицу:

Показатель	Формула	Расчет	Значение
Среднее арифметическое			
Дисперсия			
Среднеквадратическое отклонение			
Стандартизованные данные			

Задание 1.10. Сделайте развернутые выводы по основам прикладного статистического анализа экономических данных, приложение В:

– для анализа взаимосвязи между переменными *x* и *y*. Сверху каждого рисунка представлены соответствующие коэффициенты корреляции (прилож. В. 1);

– о роли нормального закона распределения эмпирических данных в

экономических исследованиях (прилож. В. 2);

– по диаграммам рассеяния наблюдений и линиям регрессии (прилож. В. 3).

Задание 1.11. Приведите данные X: 20, 40, 60, 80 к стандартизованному виду с использованием MS Excel. Расчет представьте на рабочем листе Excel.

Методические указания к выполнению задания 1.11

Оформление результатов на рабочем листе Excel и пояснение ввода формул в соответствующие ячейки таблицы:

	A	B	C	D
1	Данные			
2	исходные		стандартизированные	
3	X		Z	
4				
5				
6				
7				
8	M	S	M	S
9				

Ячейка	Формула
A9	=СРЗНАЧ (A4:B7)
B9	=СТАНДОТКЛОНА (A4:B7)
C4	=(A4-A9) / B9
Формулу ячейки C4 «протягиваем» до D7 включительно	
C9	=СРЗНАЧ (C4:D7)
B9	=СТАНДОТКЛОНА (C4:D7)

M – среднее арифметическое; S (или σ) – среднеквадратическое отклонение.

ПРАКТИЧЕСКИЕ ЗАНЯТИЯ 2–4
«Множественный корреляционно-регрессионный анализ»

Цель занятия: получить теоретические и практические навыки выполнения линейного множественного корреляционно-регрессионного анализа экономических данных и оценки значимости результатов с применением статистических программных пакетов MS Excel, SPSS и STATISTICA».

Задания по теме «Множественная корреляция»

Задание 2.1. Рассчитайте множественный коэффициент корреляции на основе данных задания 1.1 с применением матричной алгебры и использованием встроенных функций MS Excel. Результаты представьте на рабочих листах Excel.

Методические указания к выполнению задания 2.1

ШАГ 1. Для расчета линейного множественного коэффициента корреляции с применением технологии выполнения операций над матрицами в Excel изучите некоторые аспекты работы над матрицами на примерах в приложении Г.

При выполнении операций над матрицами, решении систем линейных уравнений применяются следующие функции Excel:

МУМНОЖ – умножение матриц;

ТРАНСП – транспонирование матрицы;

МОПРЕД – вычисление определителя матрицы;

МОБР – вычисление обратной матрицы.

Функции для выполнения операций с матрицами находятся в категории Математические (рис. 2.1 и 2.2).

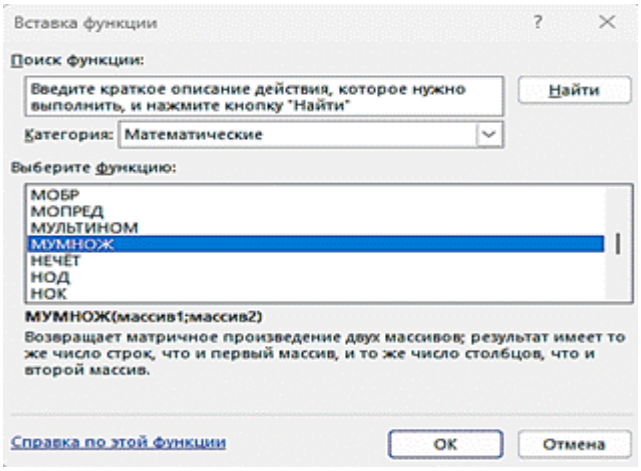


Рисунок 2.1 – Окно вставки функции **МУМНОЖ** в среде Excel

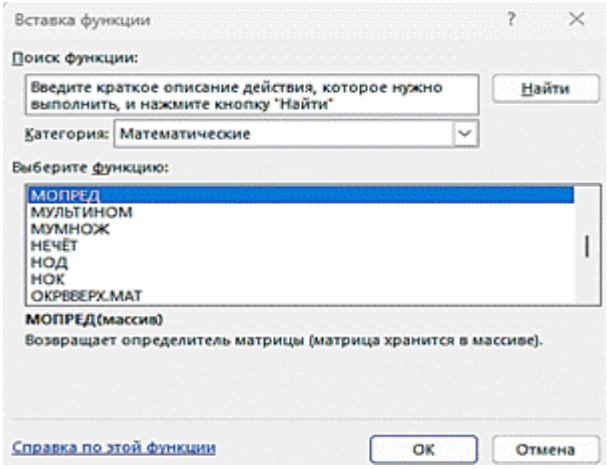


Рисунок 2.2 – Окно вставки функции **МОПРЕД** в среде Excel

Вычислите на основе данных примеров в приложении Г определитель матрицы с помощью функции **МОПРЕД** в Excel (рис. 2.3 и 2.4).

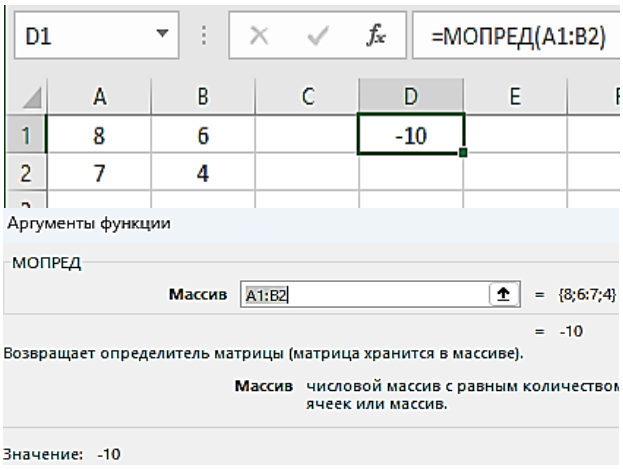


Рисунок 2.3 – В ячейке D1 получен результат вычислений матрицы A

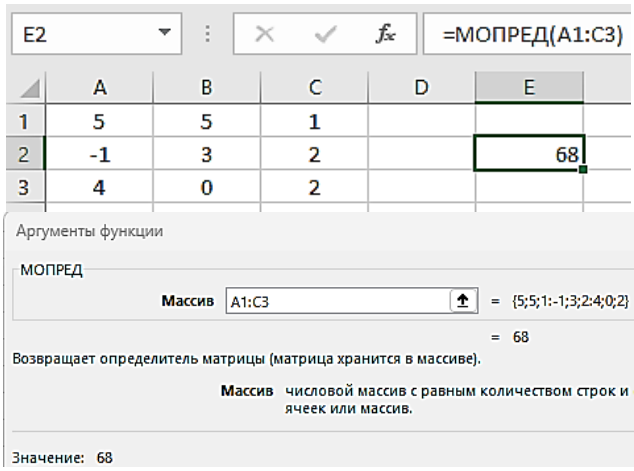


Рисунок 2.4 – В ячейке E2 получен результат вычислений матрицы B

Задание 2.2. По результатам корреляционного анализа вычислена матрица парных коэффициентов корреляции, рис. 2.5.

	A	B	C	D	E
1		Y	X1	X2	
2	Y	1	0,97	-0,93	
3	X1	0,97	1	-0,86	
4	X2	-0,93	-0,86	1	
5					

Рисунок 2.5 – Матрица парных коэффициентов корреляции

1. Рассчитайте частные и множественный коэффициенты корреляции. Для расчета линейного множественного коэффициента корреляции используйте формулу с применением матричной алгебры.
2. Определите множественный коэффициент детерминации.
3. Проведите оценку значимости коэффициентов корреляции.
4. Проанализируйте полученные результаты и сделайте выводы.
5. Представьте на рабочих листах Excel результаты всех этапов задания.

Методические указания к выполнению задания 2.2

ШАГ 1.1. Для определения линейного множественного коэффициента корреляции необходимо выполнить некоторые действия над матрицами:

– рассчитайте определитель корреляционной матрицы (см. примеры приложения Г);

– определите алгебраическое дополнение первого диагонально элемента r_{11} той же матрицы.

ШАГ 1.2. Расчет частного коэффициента корреляции можно выполнить по одному из двух направлений на выбор:

– вычисление по формуле с применением действий матричной алгебры;

– определение по формуле, используя коэффициенты парной корреляции (рис. 2.5).

ШАГ 2. Множественный коэффициент детерминации рассчитайте по формулам:

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{USS}{TSS} \tag{2.1}$$

или

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (y_i - \bar{y})^2}, \tag{2.2}$$

где R^2 – коэффициент множественной детерминации является мерой объясняющего качества уравнения регрессии; $TSS = \sum (y_i - \bar{y})^2$ – общий разброс зависимой переменной; $ESS = \sum (\hat{y}_i - \bar{y})^2$ – разброс, объясненный регрессией; $USS = \sum e_i^2$ – разброс, не объясненный регрессией.

ШАГ 3.1. Значимость линейных парных коэффициентов корреляции проверьте на основе *t*-критерия Стьюдента для совокупностей n < 50.

ШАГ 3.2. Существенность множественного коэффициента корреляции и детерминации проверьте с помощью *F*-критерия Фишера (*F*-тест) и *t*-критерия Стьюдента.

$$F = \frac{R^2}{1 - R^2} \frac{n - k - 1}{k}, \tag{2.3}$$

где R^2 – коэффициент множественной детерминации; n – количество наблюдений; k – число факторов в модели.

Замечание. *F*-тест для проверки значимости R^2 выполняется так же, как и в парном регрессионном анализе с поправкой на число степеней свободы.

Задания по теме «Множественная регрессия»

Задание 3.1. Постройте множественную линейную регрессионную модель производительности труда одного рабочего для задания 1.1.

1. Рассчитайте уравнение множественной линейной регрессии.
2. Оцените тесноту связи между анализируемыми признаками с помощью множественного коэффициента корреляции.
3. Оцените общее качество уравнения регрессии с помощью множественного коэффициента детерминации и статистическую значимость

уравнения множественной регрессии с помощью *F*-критерия Фишера (α = 0,05).

4. Оцените статистическую значимость коэффициентов множественной регрессии с помощью *t*-критерия Стьюдента (α = 0,05).

5. Рассчитайте доверительные интервалы для каждого коэффициента a_j линейного уравнения регрессии.

6. Поясните экономический смысл полученных результатов множественного регрессионного анализа.

Представьте все этапы эконометрического анализа на рабочих листах в MS Excel по формулам в соответствии с требованиями методических указаний.

Методические указания к выполнению задания 3.1

ШАГ 1. Для разработки линейного уравнения множественной регрессии с двумя независимыми переменными используйте систему нормальных уравнений, составленных в результате применения метода наименьших квадратов (МНК), формула (3.1):

$$\begin{cases} a_0 n + a_1 \sum x_1 + a_2 \sum x_2 = \sum y \\ a_0 \sum x_1 + a_1 \sum x_1^2 + a_2 \sum x_1 x_2 = \sum y x_1 \\ a_0 \sum x_2 + a_1 \sum x_1 x_2 + a_2 \sum x_2^2 = \sum y x_2 \end{cases} \quad (3.1)$$

Для определения параметров уравнения множественной регрессии постройте расчетную таблицу 3.1.

Таблица 3.1 – Вспомогательные вычисления для определения параметров уравнения множественной регрессии

№ предприятия	Исходные данные			Расчетные данные					
n	*Y*	X_1	X_2	Y^2	X_1^2	X_2^2	YX_1	YX_2	X_1X_2
1									
...									
20									
Итого									
Среднее									

Подставьте вычисленные значения из таблицы 3.1 в систему нормальных уравнений (формула 3.1).

Выполните ряд арифметических действий, решая систему нормальных уравнений. Получите коэффициенты (параметры) уравнения множественной регрессии и постройте на их основе множественную линейную регрессионную модель производительности труда одного рабочего в целом по группе предприятий.

ШАГ 2. Общую оценку качества полученного уравнения регрессии вычисляют с помощью коэффициента множественной корреляции (индекс корреляции) и множественного коэффициента детерминации. Коэффициент

множественной корреляции рассчитайте по формуле:

$$R = \sqrt{1 - \frac{ESS}{TSS}} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}} = \sqrt{1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}}, \quad (3.2)$$

где ESS = $\sum_{i=1}^{n} e_i^2$ – не объясненная регрессией (остаточная) сумма квадратов отклонений; TSS = $\sum_{i=1}^{n} (y_i - \bar{y})^2$ – общая сумма квадратов отклонений зависимой переменной от ее среднего выборочного значения; $\hat{y}_i$ – предсказанное (рассчитанное по уравнению регрессии) значение зависимой переменной.

Прежде чем рассчитывать коэффициент множественной корреляции определите составляющие формулы 3.2 в таблице 3.2.

Таблица 3.2 – Вспомогательные вычисления для оценки качества уравнения множественной регрессии

№ предприятия	Исходные данные	Расчетные данные		
n	y_i	$\hat{y}_i$	$\hat{y}_i - \bar{y}$	$(\hat{y}_i - \bar{y})^2$
1				
...				
20				
Итого				
Среднее				

Замечание. Коэффициент множественной корреляции универсален, так как он отражает тесноту связи и точность модели, а также может использоваться при любой форме связи переменных.

После того как уравнение регрессии построено, выполняется проверка значимости построенного уравнения в целом и отдельных параметров. Оценка значимости уравнения регрессии позволяет узнать, пригодно уравнение для практического использования (например, для прогноза) или нет.

ШАГ 3. Оценку качества уравнения регрессии с помощью коэффициента множественной детерминации выполните на основе полученного результата по коэффициенту множественной корреляции. Чем ближе к единице значения этих характеристик, тем выше качество модели.

Для проверки значимости (адекватности) всей модели регрессии и правильности выбора формы связи используется *F*-критерий Фишера, вычисляемый по формуле:

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)} = \frac{R^2}{1-R^2} \times \frac{n-k-1}{k}, \quad (3.3)$$

где k – количество факторов (параметров), включенных в модель; n – численность совокупности; R^2 – совокупный коэффициент множественной детерминации.

Величина $F_{табл}$ находится по таблицам (приложение Д) при заданном уровне значимости (P = 1 – α) и числе степеней свободы $\nu_1 = k$, $\nu_2 = n - k - 1$. Если расчетное значение *F*-статистики больше табличного $F_{расч} > F_{табл}$, то модель считается статистически значимой (существенной).

ШАГ 4. Проверка значимости отдельных коэффициентов множественной регрессии связана с определением расчетных значений *t*-критерия Стьюдента (*t*-статистики) для соответствующих коэффициентов регрессии (отношение коэффициента регрессии к его средней ошибке), вычисляемые по формулам:

$$t_{a_j} = \frac{\hat{a}_j}{\sigma_{a_j}} = \frac{\hat{a}_j}{\sigma_e\sqrt{b_{jj}}}; \qquad \sigma_e = \sqrt{\frac{1}{n-k-1}\sum_{i=1}^{n} e_i^2}, \tag{3.4}$$

$$\text{или } t_{a_j} = \frac{a_j}{S_{a_j}}, \qquad S_{a_j} = \sqrt{\frac{SS_{ост}}{n-m-1}[(X^TX)^{-1}]_{jj}}, \tag{3.5}$$

где $\sigma_{a_j} = S_{a_j}$ – среднеквадратическое (стандартное) отклонение коэффициента уравнения регрессии a_j (стандартная ошибка параметра уравнения a_j); σ_e – стандартная ошибка; b_{jj} – диагональный элемент матрицы $(X^TX)^{-1}$; $SS_{ост}$ – остаточная сумма квадратов отклонений, $SS_{ост} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$; $[(X^TX)^{-1}]_{jj}$ – диагональный элемент матрицы $(X^TX)^{-1}$.

Расчетные значения $t_{расч}$ сравниваются с табличными (критическими) $t_{табл}$ по таблице значений t – критерия Стьюдента при уровне значимости α и числе степеней свободы. Если расчетное значение t-критерия с $(n - k - 1)$ степенями свободы больше его табличного значения $|t_{a_j}| > t_{крит}$ при заданном уровне значимости, то коэффициент регрессии a_j считается статистически значимым.

Замечание. *Статистическая значимость коэффициентов регрессии и близкое к единице значение коэффициента детерминации R^2 не гарантируют высокое качество уравнения регрессии.*

ШАГ 5. Доверительные интервалы для коэффициентов a_j линейного уравнения регрессии находятся по формуле:

$$a_j \pm t_{крит} S_{aj} \tag{3.6}$$

ШАГ 6. Поясните экономический смысл полученных результатов линейного множественного корреляционно-регрессионного анализа.

Задание 3.2. Предполагается, что объем предложения некоторого блага (Y) для функционирующей в условиях конкуренции организации зависит линейно от цены (X1) данного блага и заработной платы (X2) сотрудников

организации, производящих данное благо.

Статистические данные, собранные за 16 кварталов, занесены в следующую таблицу:

Y	20	35	30	45	60	69	75	90	105	110	120	130	130	130	135	140
X1	10	15	20	25	40	37	43	35	38	55	50	35	40	55	45	65
X2	12	10	9	9	8	8	6	4	4	5	3	1	2	3	1	2

1. Оцените по МНК коэффициенты уравнения регрессии.
2. Определите интервальные оценки коэффициентов при уровне значимости α = 0,1.
3. Оцените общее качество уравнения регрессии с помощью множественного коэффициента детерминации и статистическую значимость уравнения множественной регрессии с помощью *F*-критерия Фишера при уровне значимости α = 0,05 (можно определить по таблице из приложения Д).
4. Проверьте гипотезу об отсутствии автокорреляции остатков на основе статистики Дарбина – Уотсона *DW*.
5. Сделайте выводы по построенной модели.

Методические указания к выполнению задания 3.2

Выполните **ШАГИ 1–3** задания в MS Excel с помощью задания 3.1.

ШАГ 4. Гипотеза об отсутствии автокорреляции проверяется на основе статистики Дарбина – Уотсона *DW*, которая определяется по формуле:

$$DW = \frac{\sum(e_i - e_{i-1})^2}{\sum e_i^2}, \qquad (3.7)$$

где $e_i = y_i - \hat{y}_i$ – остатки регрессии.

Рассчитайте для критерия *DW* составляющие формулы 3.7 в таблице 3.3.

Таблица 3.3 – Вспомогательные вычисления для расчета критерия *DW*

№	Расчетные данные				
n	$\hat{y}_i$	$e_i = y_i - \hat{y}_i$	$e_i - e_{i-1}$	$(e_i - e_{i-1})^2$	e_i^2
1					
...					
16					
Итого					

Для значений статистики *DW* выполняется неравенство $0 \leq DW \leq 4$.

Необходимым условием независимости случайных отклонений является близость к двойке значения статистики Дарбина – Уотсона (автокорреляция отсутствует). Оценки, получаемые по критерию, являются не точечными, а интервальными. Верхние (d_2) и нижние (d_1) критические значения, позволяющие принять или отвергнуть гипотезу об отсутствии автокорреляции,

зависят от количества уровней динамического ряда и числа независимых переменных модели. Значения этих границ для уровня значимости α = 0,05 даны в специальных таблицах (приложение Е).

Справочно. Автокорреляция может появиться из-за опущения переменных, неверной формы функции в оценочном выражении, например, линейная модель тогда, когда она должна быть нелинейной и др.

Задания по теме «Множественный корреляционно-регрессионный анализ с использованием статистических программных пакетов

Задание 4.1. По данным, представленным в приложении Ж, исследуется зависимость рентабельности основных средств Y (коэф.) 54 коммерческих организаций и следующих факторов:

- X1 – фондоотдача, руб.;
- X2 – фондоотдача машин и оборудования, руб.;
- X3 – амортоемкость продукции, руб.;
- X4 – коэффициент износа основных средств, коэф.;
- X5 – коэффициент износа активной части основных средств, коэф.;
- X6 – удельный вес машин и оборудования в основных средствах, коэф.;
- X7 – удельный вес машин и оборудования в активной части основных средствах, коэф.;
- X8 – удельный вес пассивной части в основных средствах, коэф.;
- X9 – коэффициент поступления активной части основных средств, коэф.;
- X10 – коэффициент обновления активной части основных средств, коэф.;
- X11 – коэффициент выбытия активной части основных средств, коэф.;
- X12 – коэффициент прироста инвестиций в основные средства, коэф.;
- X13 – доля амортизационных отчислений в источниках финансирования инвестиций в основных средствах, коэф.

1. Используйте пошаговую множественную линейную регрессию, применяя метод исключения, постройте модель для зависимой переменной «рентабельность основных средств» за счет значимых факторов.
2. Рассчитайте коэффициент множественной корреляции. Оцените качество полученного множественного уравнения регрессии с помощью коэффициента детерминации и скорректированного коэффициента детерминации.
3. Оцените статистическую значимость уравнения регрессии, используя *F*-критерий Фишера (α = 0,05).
4. Дайте сравнительную оценку силы связи между факторами и результатом с помощью коэффициента эластичности, бета-коэффициента и дельта-коэффициента.
5. Выполните ранжирование факторов по уровню их значимости на рентабельность основных средств.

6. Поясните экономический смысл полученных результатов линейного множественного корреляционно-регрессионного анализа.

Для выполнения задания используйте инструменты *Анализа данных* в MS Excel. Результаты всех этапов задания представьте на рабочих листах Excel.

Методические указания к выполнению задания 4.1

Выполните задание в MS Excel с помощью команды *Анализ данных*. Загрузите (или скопируйте) в Excel данные приложения Ж для выполнения линейного множественного корреляционно-регрессионного анализа рентабельности основных средств по следующим шагам.

ШАГ 1. Выбор факторов для построения линейной регрессионной модели выполните из условия задания 4.1. В статистическом анализе рентабельность основных средств является зависимая переменная Y (коэф.). Выбор объясняющих переменных из X1–X13 выполните с помощью:

– анализа матрицы коэффициентов парной корреляции;
– метода отсева (пошаговый отбор методом исключения).

Для проведения корреляционного анализа используйте команду *Анализ данных* – инструмент анализа *Корреляция* в Excel. С помощью этого инструмента анализа получите матрицу коэффициентов парной корреляции.

Анализ матрицы коэффициентов парной корреляции начните с ее первого столбца, где расположены коэффициенты корреляции, отражающие тесноту связи зависимой переменной *Y* «Рентабельность основных средств» с включенными в анализ факторами. Факторы, имеющие слабую связь с зависимой переменной, не рекомендуется включать в модель регрессии.

Пошаговый отбор методом исключения основан на последовательном исключении факторов с помощью *t*-критерия из модели, (рис. 4.1).

15						
16		*Коэффициенты*	*Стандартная ошибка*	***t*-статистика**	*P-Значение*	*Нижние 95%*
17	Y-пересечение	0,011875064	0,248314522	**0,047822674**	0,962095549	-0,489987306
18	X1	0,087855057	0,102592979	**0,856345707**	0,396907769	-0,119493088
19	X2	0,009243371	0,051086991	**0,180933941**	0,857333096	-0,094007289
20	X3	0,000118935	0,012305723	**0,009665032**	0,992336584	-0,024751859
21	X4	0,019931694	0,15154199	**0,131525881**	0,896018649	-0,286346093
22	X5	-0,051323085	0,11597201	**-0,442547168**	0,660477219	-0,28571126

Рисунок 4.1 – Фрагмент выполнения регрессионного анализа методом исключения с помощью *t*-статистики в MS Excel

После построения уравнения регрессии и оценки значимости всех коэффициентов регрессии из модели исключите тот фактор, коэффициент при

котором незначим и имеет наименьший по абсолютной величине коэффициент t. После этого получают новое уравнение множественной регрессии и снова проводят оценку значимости всех оставшихся коэффициентов регрессии. Процесс исключения факторов останавливается на том шаге, при котором все регрессионные коэффициенты значимы не только при 5%-м, но и при 1%-м уровне значимости.

Постройте модель для зависимой переменной «рентабельность основных средств» за счет значимых факторов и поясните экономический смысл ее параметров. Для проведения регрессионного анализа используйте вкладку *Данные* – команду *Анализ данных* – инструмент анализа *Регрессия* в Excel.

ШАГ 2. Для расчета коэффициента множественной корреляции и оценки качества полученного уравнения регрессии с помощью коэффициента детерминации используйте инструмент анализа *Регрессия* в Excel.

ШАГ 3. Проверку значимости уравнения регрессии проведите на основе *F*-критерия Фишера ($\alpha = 0{,}05$). Его расчетное значение приведено в пятом столбце таблицы «Дисперсионный анализ» отчета регрессионного анализа в Excel (рис. 4.2).

9						
10	Дисперсионный анализ					
11		*df*	*SS*	*MS*	*F*	*Значимость F*
12	Регрессия	13	0,284169321	0,021859179	**12,12837066**	4,597E-10
13	Остаток	40	0,072092713	0,001802318		
14	Итого	53	0,356262034			

Рисунок 4.2 – Результат *F*-критерия Фишера в дисперсионном анализе Excel

F-критерий также можно рассчитать по формуле (задание 2.2).

Табличное значение *F*-критерия (распределение Фишера, *F*-распределение) при доверительной вероятности 0,95 и значения ν_1, ν_2 можно найти по статистическим таблицам (приложение Д).

Табличное (критическое) значение *F*-критерия можно найти также с использованием функции FРАСПОБР в MS Excel, (см. рис. 1.2).

ШАГ 4. Важное значение при оценке влияния факторов играют коэффициенты регрессионной модели. Однако непосредственно с их помощью нельзя сопоставить факторы по степени их влияния на зависимую переменную из-за различия единиц измерения и разной степени колеблемости. Для устранения таких различий при интерпретации применяются коэффициенты эластичности $Э_j$, а также бета- и дельта-коэффициенты β_j , Δ_j.

Для каждого коэффициента регрессии вычислите коэффициент эластичности по формуле:

$$Э_j = \hat{a}_j \frac{\bar{x}_j}{\bar{y}}, \quad (4.1)$$

где $\hat{a}_j$ – коэффициент регрессии при j-м факторе; $\bar{x}_j$ – среднее значение j-го фактора; $\bar{y}$ – среднее значение изучаемого показателя.

Коэффициент эластичности показывает, на сколько процентов изменяется зависимая переменная при изменении j-го фактора на 1 %.

Эластичность не нормирована и может изменяться от -∞ до +∞. Высокий уровень эластичности означает сильное влияние независимой переменной на объясняемую переменную.

Для более точной оценки влияния факторов на зависимую переменную Y применяется бета-коэффициент β_j. В отличие от коэффициента эластичности бета-коэффициент учитывает степень колеблемости факторов.

$$\beta_j = \hat{a}_j S_{x_j}/S_y, \quad (4.2)$$

где $S_{x_j} = \sigma_{x_j}$ – среднеквадратическое отклонение j-го фактора; $S_y = \sigma_y$ – среднеквадратическое отклонение результативного фактора.

Стандартные отклонения вычисляются по формулам:

$$S_{x_j} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}, \quad (4.3)$$

$$S_y = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}. \quad (4.4)$$

Бета-коэффициент β_j показывает, на какую часть своего среднеквадратического отклонения σ_y изменится результативный признак *y* при изменении факторного признака x_j на величину своего σ_{x_j} при фиксированных значениях остальных факторных признаков. Фиксация обычно осуществляется на уровне средних значений $\bar{x}$.

Более конкретную (количественную) информацию о вкладе каждого из факторных признаков в вариацию результативного признака *y* можно получить, рассчитав дельта-коэффициенты Δ_j.

Экономический смысл дельта-коэффициента Δ_j заключается в том, что он показывает на сколько процентов изменение результативного признака *y* зависит от конкретного факторного признака x_j. Дельта-коэффициенты позволяют ранжировать регрессоры по степени их воздействия на зависимую переменную и выделить доминирующий фактор. Формула расчета дельта-коэффициентов имеет вид:

$$\Delta_j = r_j \frac{\beta_j}{R^2}, \quad (4.5)$$

где $r_j = r_{yx_j}$ – парный коэффициент корреляции между j-м фактором (j = 1, k) и зависимой переменной; β_j – стандартизованный коэффициент регрессии факторного признака x_j; R^2 – множественный коэффициент детерминации.

ШАГ 5. Выполните ранжирование факторов по уровню их значимости на рентабельность основных средств. Для более полной интерпретации результатов регрессионной модели рентабельности основных средств представьте рассчитанные коэффициенты регрессии, коэффициенты эластичности $Э_j$, а также бета- и дельта-коэффициенты β_j, Δ_j в сводную таблицу 4.1.

Таблица 4.1 – Оценки коэффициентов регрессионной модели рентабельности основных средств

Фактор	Коэффициент регрессии $\hat{a}_j$		Коэффициент эластичности $Э_j$		Бета-коэффициент βj		Дельта-коэффициент ∆j		Сводный ранг
	значение	ранг	значение	ранг	значение	ранг	значение	ранг	
X_1									
…									
X_n									

Задание 4.2. По исходным данным, представленным в приложении Ж, выполните линейный множественный корреляционно-регрессионный анализ рентабельности основных средств по следующим направлениям.

1. Постройте матрицу парных коэффициентов корреляции. Установите, какие факторы мультиколлинеарны. С помощью разработанных методов устраните или уменьшите мультиколлинеарность.
2. Выполните тест Феррара – Глоубера на мультиколлинеарность всего массива переменных (критерий «хи-квадрат»).
3. Отберите в регрессионную модель информативные факторы с помощью метода включения. Постройте линейную множественную регрессию рентабельности основных средств только со значимыми факторами.
4. Оцените качество полученного уравнения регрессии с помощью средней относительной ошибки аппроксимации и коэффициента детерминации. Проверьте статистическую значимость коэффициента детерминации.
5. Выполните оценку статистической значимости коэффициентов уравнения множественной регрессии, используя *t*-критерий Стьюдента.
6. Оцените силу влияния факторов на зависимую переменную по модели с помощью коэффициента эластичности, бета-коэффициента и дельта-коэффициента.
7. Выполните ранжирование факторов по уровню их значимости на рентабельность основных средств.
8. Дайте экономическую интерпретацию полученных результатов линейного множественного корреляционно-регрессионного анализа.

Для выполнения задания используйте команду *Анализ данных* в MS Excel. Результаты всех этапов задания представьте на рабочих листах MS Excel.

Методические указания к выполнению задания 4.2

Выполните задание в MS Excel с помощью команды *Анализ данных*. Загрузите (или скопируйте) в Excel данные приложения Ж для выполнения линейного множественного корреляционно-регрессионного анализа в MS Excel.

Выполните линейный множественный корреляционно-регрессионный анализ рентабельности основных средств с помощью следующих шагов.

ШАГ 1. Выполните выбор факторов для построения множественной регрессионной модели. Из условия задания 4.1 рентабельность основных средств – это зависимая переменная Y (коэф.). Выбор объясняющих переменных из X1– X13 выполните с помощью:

– анализа матрицы коэффициентов парной корреляции;

– методов устранения или уменьшения мультиколлинеарности.

Для анализа матрицы коэффициентов парной корреляции используйте в Excel команду *Анализ данных* – инструмент *Корреляция*.

Анализ матрицы коэффициентов парной корреляции начните с ее первого столбца, где расположены коэффициенты корреляции, отражающие тесноту связи зависимой переменной *Y* «Рентабельность основных средств» с включенными в анализ факторами. Факторы, имеющие слабую связь с зависимой переменной, не рекомендуется включать в модель регрессии.

Затем выполните анализ остальных столбцов матрицы коэффициентов парной корреляции с целью выявления мультиколлинеарности. Самый простой из методов устранения или уменьшения мультиколлинеарности состоит в том, что из двух объясняющих переменных, имеющих высокий коэффициент корреляции (больше 0,8), одну переменную исключают из рассмотрения. При этом какую переменную оставить, а какую удалить из анализа, руководствуйтесь из экономических соображений.

ШАГ 2. Выполните тест Фаррара – Глоубера на мультиколлинеарность. С помощью этого теста проверяют, насколько значимо определитель матрицы парных коэффициентов корреляции отличается от единицы. Если он равен нулю, то столбцы матрицы *X* линейно зависимы и вычислить оценку коэффициентов множественной регрессии по методу наименьших квадратов становится невозможно.

Для проверки наличия мультиколлинеарности всего массива переменных (критерий «хи-квадрат») выполните следующие этапы:

– постройте корреляционную матрицу *R* и найдите ее определитель det[$\boldsymbol{R}$] с помощью функции МОПРЕД (см. методические указания к заданию 2.1);

– вычислите наблюдаемое значение статистики Фаррара – Глоубера по следующей формуле:

$$FG_{\text{набл}} = -\left[n - 1 - \frac{1}{6}(2k + 5)\right]\ln(\det[R]), \qquad (4.6)$$

где n – количество наблюдений; k – количество факторов.

– фактическое значение критерия $FG_{\text{набл}}$ сравните с табличным значением χ^2 при $0{,}5k$ $(k - 1)$ степенях свободы и уровнем значимости α = 0,05.

Табличное (критическое) значение χ^2 можно найти с помощью функции **ХИ2.ОБР.ПХ** в MS Excel (рис. 4.3).

Если фактическое значение критерия $FG_{\text{набл}}$ больше табличного (критического), то в массиве объясняющих переменных существует мультиколлинеарность.

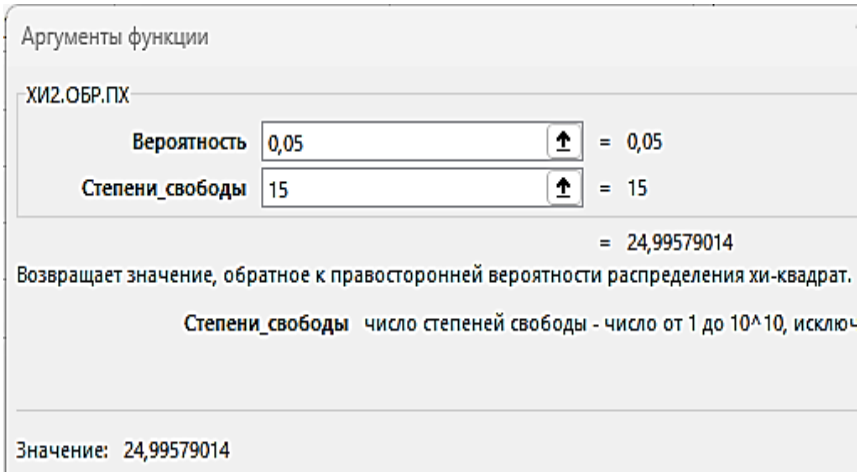


Рисунок 4.3 – Получение табличного (критического) значения χ^2 в MS Excel

ШАГ 3. Пошаговый отбор наиболее информативных факторов с помощью метода включения для построения множественной регрессионной модели основан на последовательном включении факторов в уравнение регрессии и последующей проверке их значимости с помощью множественного коэффициента детерминации R^2 и значения F-критерия.

На первом этапе задания 4.2 в модель вводится та переменная, которая имеет наибольший по абсолютной величине коэффициент корреляции с результативным признаком. Для уточнения модели регрессии на втором этапе в рассмотрение включается следующая переменная с наибольшим частным коэффициентом корреляции. На следующем, третьем этапе процедуры пошаговой регрессии выберите с наибольшим частным коэффициентом корреляции и т. д. Процедура пошаговой регрессии (для данного задания) прекращается тогда, если оставшаяся переменная (или переменные) не обеспечивает заданное минимальное значение F-критерия, а также включение оставшейся переменной (или переменных) в уравнение существенно не увеличивает множественный коэффициент детерминации R^2.

После того как уравнение регрессии построено, выполняется проверка значимости построенного уравнения в целом и отдельных параметров.

ШАГ 4. Для оценки качества полученного уравнения регрессии с помощью коэффициента детерминации используйте команду *Анализ данных* – инструмент анализа *Регрессия* в Excel (таблица *Регрессионная статистика*).

Для оценки качества (уровень точности) регрессионных моделей используется средняя относительная ошибка аппроксимации, формула (1.8).

Для оценки общего качества модели множественной регрессии вычисляют коэффициент детерминации по формуле:

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (y_i - \bar{y})^2}\ , \qquad (4.7)$$

где $\sum e_i^2 = \sum (y_i - \hat{y}_i)^2$ – остаточная (необъясненная) сумма квадратов, являющаяся мерой остаточного не объясненного уравнением регрессии разброса (разброса точек вокруг линии регрессии).

Для проверки гипотезы о статистической значимости коэффициента детерминации на практике используется следующая *F*-статистика:

$$F = \frac{R^2}{1-R^2} \times \frac{n-m-1}{m} \qquad (4.8)$$

Статистика F при выполнении предпосылок МНК и при справедливости H_0 имеет распределение Фишера с числами степеней свободы $\nu_1 = m$, $\nu_2 = n - m - 1$.

ШАГ 5. Оценка статистической значимости коэффициентов уравнения множественной регрессии, выполняется с помощью *t*-критерия Стьюдента по формулам (см. методические указания к заданию 3.1, ШАГ 4) или с помощью инструмента *Регрессия* в Excel (графа *t*-статистика в таблице отчета).

Табличное (критическое) значение *t*-критерия Стьюдента можно найти с помощью функции **СТЬЮДРАСПОБР** в MS Excel, (рис. 4.4).

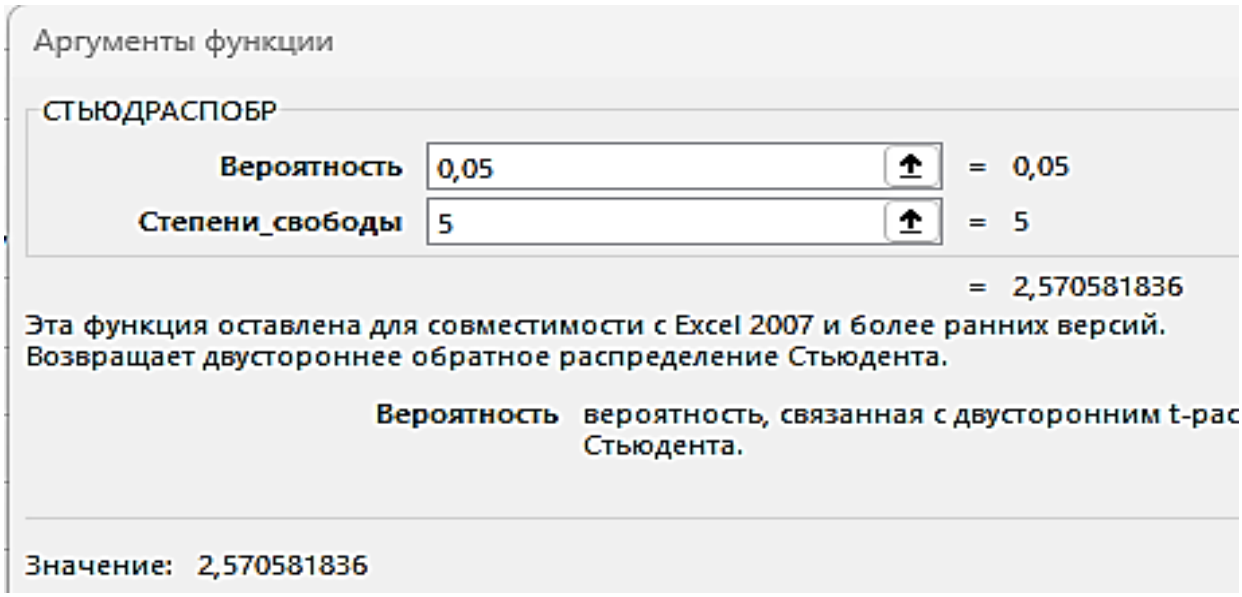


Рисунок 4.4 – Получение табличного (критического) значения *t*-критерия Стьюдента в Excel

Вычисленное значение t_{aj} сравнивается с табличным (критическим) значением *t*-критерия: если $|t_{расч}| > t_{табл}$, то коэффициент модели считается существенным (значимым).

Табличное значение *t*-критерия Стьюдента при уровне значимости 0,10, 0,05 и 0,01 можно найти по статистическим таблицам (приложение И).

ШАГ 6. Оценку силы влияния факторов на зависимую переменную по модели с помощью коэффициента эластичности, бета-коэффициента и дельта-коэффициента выполните по формулам, представленным в методических указаниях к заданию 4.1 (ШАГ 4).

ШАГ 7. Ранжирование факторов по уровню их значимости на рентабельность основных средств выполните по примеру таблицы 4.1 в методических указаниях к заданию 4.1 (ШАГ 5).

ШАГ 8. Экономическую интерпретацию полученных результатов эконометрического анализа представьте в сводной таблице 4.2.

Таблица 4.2 – Модели пошагового регрессионного анализа рентабельности основных средств и характеристики их точности

Шаг	Модель	Параметр модели	*t*-критерий Стьюдента			Средняя ошибка аппроксимации, %	Коэффициент детерминации
			расчетное	критическое	результаты сравнения		
1	$\hat{y}_x =$	a_1 a_2 a_n			знач. незнач.		
…	…	…	…	…	…	…	…
n							

Задание 4.3. По данным 54 коммерческих организаций приложения Ж, выполните статистический анализ зависимости рентабельности основных средств от факторов X1–X13 по следующим направлениям.

1. Определите парные коэффициенты корреляции между всеми переменными.

2. Определите наличие или отсутствие мультиколлинеарности и устраните ее с помощью одного из разработанных методов.

3. Определите коэффициенты (параметры) множественного уравнения регрессии, используя один из методов (включения или исключения).

4. Постройте линейную множественную модель регрессии только со статистически значимыми факторами в стандартизированной форме. Дайте экономическую интерпретацию коэффициентам регрессионной модели.

5. Оцените статистическую значимость параметров множественной регрессионной модели с помощью *t*-критерия Стьюдента, проверьте нулевую гипотезу о значимости уравнения с помощью *F*-критерия ($\alpha = 0{,}01$), оцените

качество полученного уравнения регрессии с помощью коэффициента детерминации.

6. Определите долю неучтенных факторов в дисперсии результативного признака.

7. Определите доверительные интервалы множественных коэффициентов регрессии при уровне доверительной вероятности (p) 0,95.

8. Для выполнения задания используйте команду *Анализ данных* в MS Excel. Результаты всех этапов задания представьте на рабочих листах Excel.

Методические указания к выполнению задания 4.3

Выполните задание в MS Excel с помощью команды *Анализ данных*. Загрузите (или скопируйте) в Excel данные из приложения Ж для выполнения линейного множественного корреляционно-регрессионного анализа.

Линейный множественный корреляционно-регрессионный анализ рентабельности основных средств выполните с помощью следующих шагов.

ШАГ 1. Выполните выбор факторов для построения множественной регрессионной модели. Из условия задания 4.3 рентабельность основных средств – это зависимая переменная Y (коэф.). Выбор объясняющих переменных из X1–X13 выполните с помощью анализа матрицы коэффициентов парной корреляции.

Для анализа матрицы коэффициентов парной корреляции используйте команду *Анализ данных* – инструмент анализа *Корреляция* в Excel.

Пошаговый корреляционный анализ (парный и множественный) выполните с помощью методических указаний к заданиям 3.1–4.2 в MS Excel.

ШАГ 2. Наличие или отсутствие мультиколлинеарности выполните с помощью методических указаний к заданиям 3.1–4.2 в Excel.

ШАГ 3. Выбор информативных факторов для модели регрессии, используя один из методов (включения или исключения), выполните с помощью методических указаний к заданиям 3.1–4.2 в MS Excel.

ШАГ 4. Для устранения влияния различия дисперсий и единиц измерения отдельных переменных на результаты регрессионного анализа в ряде случаев целесообразно вместо исходных значений переменных x_{ij} использовать стандартизованные значения:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j} \qquad \text{или} \tag{4.9}$$

$$t_y = \frac{y_i - \bar{y}}{\sigma_y}; \qquad t_{xj} = \frac{x_{ij} - \bar{x}_j}{\sigma_{xj}}, \tag{4.10}$$

где z_{ij} или t_y , t_{xj} – стандартизованные значения; σ_y и σ_{xj} – средние квадратические отклонения по переменным y и x_j.

Линейное уравнение множественной регрессии в стандартизованном

масштабе имеет следующий вид:

$$\hat{Z}_0 = \beta_1 Z_1 + \beta_2 Z_2 + \cdots + \beta_m Z_m \text{ или} \quad (4.11)$$

$$t_y = \beta_1 t_{x1} + \beta_2 t_{x2} + \cdots + \beta_p t_{xp}, \quad (4.12)$$

где β_j – стандартизованные коэффициенты регрессии, которые вычислены методом наименьших квадратов (МНК) путем составления и решения системы уравнений.

Стандартизованные коэффициенты регрессии для модели с двумя факторами определяются по формулам:

$$\beta_1 = \frac{r_{yx_1} - r_{yx_2} r_{x_1 x_2}}{1 - r_{x_1 x_2}^2}; \quad \beta_2 = \frac{r_{yx_2} - r_{yx_1} r_{x_1 x_2}}{1 - r_{x_1 x_2}^2} \quad (4.13)$$

Единицей измерения признаков в стандартизованном масштабе уравнения регрессии принимается их среднее квадратическое отклонение. Коэффициенты регрессии в стандартизованном масштабе показывают на сколько стандартных отклонений изменится *Y* при изменении каждой из X_j на одно стандартное отклонение.

Выполните ранжирование факторов по силе влияния на результат *y*, исходя из величин стандартизованных коэффициентов. Чем больше по модулю величина коэффициента, тем сильнее соответствующий фактор влияет на результат.

Стандартизованные коэффициенты регрессии для модели можно найти с помощью команды *Анализ данных* – инструмента *Регрессия* в среде Excel.

ШАГ 5. Оценку статистической значимости параметров множественной регрессионной модели с помощью t-критерия Стьюдента и значимость уравнения с помощью *F*-критерия (α = 0,01), а также качество полученного уравнения регрессии с помощью коэффициента детерминации выполните с помощью методических указаний к заданиям 3.1–4.2 в MS Excel.

ШАГ 6. Определите долю неучтенных факторов в дисперсии результативного признака.

На основе совокупного коэффициента детерминации можно определить остаточную дисперсию по формуле:

$$\sigma_{\text{ост}}^2 = 1 - R_{y,x1,x2}^2, \quad (4.14)$$

где $\sigma_{\text{ост}}^2$ – остаточная дисперсия; $R_{y,x1,x2}^2$ – совокупный (множественный) коэффициент детерминации.

ШАГ 7. Доверительные интервалы множественных коэффициентов регрессии при уровне доверительной вероятности (p) 0,95 определите с помощью формулы 3.6 задания 3.1.

Задание 4.4. По исходным данным, представленным в приложении Ж, выполните следующие направления линейного множественного регрессионного анализа рентабельности основных средств организаций.

1. По результатам решения заданий (4.1–4.3) исчислите теоретические уровни на основе построенного уравнения регрессии и сделайте вывод о правильности выполненных расчетов.

2. Сравните эмпирические данные y_i и рассчитанные по уравнению регрессии $\hat{y}_i$ и вычислите остатки ряда e_i.

3. Представьте графически корреляционное поле и линию регрессии эмпирических и теоретических (сглаженных) данных. Определите точечный прогноз рентабельности основных средств.

4. Дайте экономическую интерпретацию полученных результатов линейного множественного корреляционно-регрессионного анализа.

5. Для выполнения задания используйте инструмент *Регрессия* в MS Excel. Результаты всех этапов задания представьте на рабочих листах Excel.

Методические указания к выполнению задания 4.4

Выполните задание в MS Excel с помощью команды *Анализ данных*. Загрузите (или скопируйте) в Excel данные из приложения Ж для выполнения линейного множественного регрессионного анализа.

ШАГ. 1. Для выполнения данного этапа статистического анализа руководствуйтесь результатами заданий (4.1–4.3).

Для получения значений $\hat{y}_i$ последовательно подставьте значения x_i в полученное множественное уравнение регрессии.

ШАГ. 2. Вычислите остатки ряда e_i и сравните эмпирические данные y_i с рассчитанными по уравнению регрессии $\hat{y}_i$ в таблице 4.3.

Таблица 4.3 – Вычисление остатков ряда на основе эмпирических и расчетных данных

№ предприятия	Исходные данные	Расчетные данные	
n	y_i	$\hat{y}_i$	$e_i = y_i - \hat{y}_i$
1			
…			
54			
Итого			

ШАГ. 3. График корреляционного поля и линию регрессии эмпирических и теоретических (сглаженных) данных представьте на основе данных приложения Ж и задания 1.2 ШАГ 3.

Для определения точечного прогноза рентабельности основных средств подставьте в выделенные факторы x_i регрессионной модели новые их значения.

Задание 4.5. Постройте множественную регрессионную модель эффективности использования основных средств по данным, приведенным в приложение Ж с применение программы IBM SPSS Statistics.

Требуется:

1) обосновать выбор факторных признаков для построения регрессионной модели на основе анализа матрицы коэффициентов корреляции;

2) используя пошаговую множественную регрессию (метод исключения или метод включения), построить линейную регрессионную модель, включающую значимые факторы. Дать экономическую интерпретацию коэффициентов модели регрессии;

3) для оценки качества всего уравнения регрессии определить линейный коэффициент множественной корреляции и коэффициент детерминации. Оценить значимость уравнения регрессии;

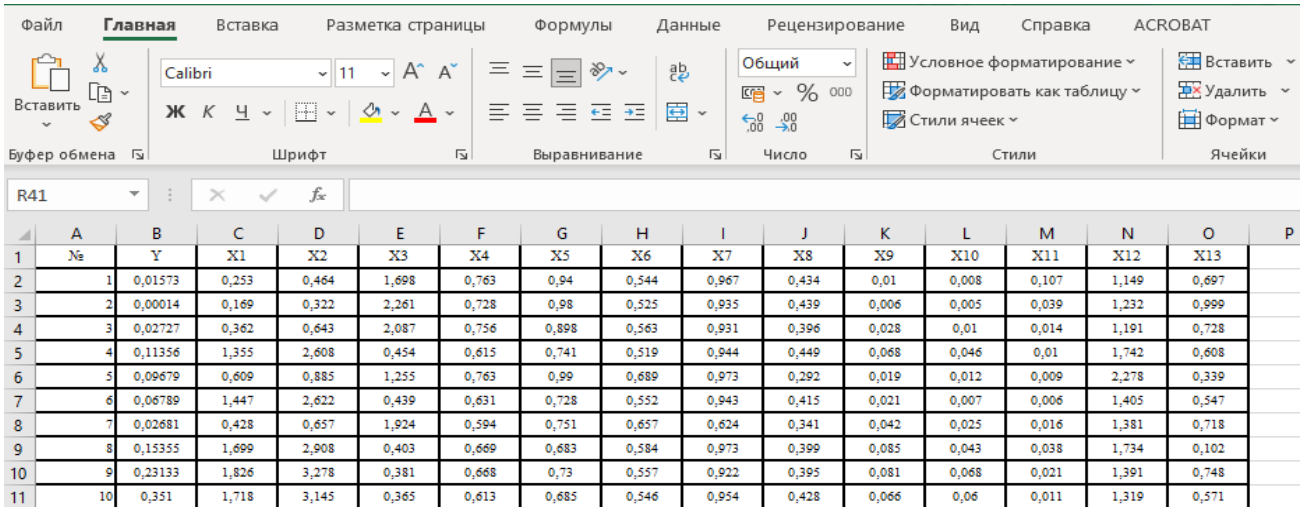
4) построить точечный и интервальный прогноз результирующего показателя (α = 0,1), если известны следующие прогнозные значения объясняющих переменных: X_1 = 1,549; X_9 = 0,406; X_{10} = 0,226.

Методические указания к выполнению задания 4.5

ШАГ 1. Подготовка данных для использования пакета IBM SPSS Statistics. Данные могут быть введены непосредственно в SPSS или импортированы из текстовых файлов, файлов формата Excel, ACCESS и т. п. В нашем случае имеется файл данных в формате Excel (рис. 4.5). Приступим к импорту данных в файл IBM SPSS.

Для импорта данных из файла формата Excel в IBM SPSS выполните следующие действия:

ШАГ 1.1. После запуска программы IBM SPSS и появления на экране пустого файла SPSS, подобного изображенному на рис. 4.6, выберите в строке меню Файл – Открыть – Данные и задайте имя Excel-файла с нашими данными, его тип, после чего нажмите кнопку Открыть (Open) (рис. 4.7).



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	№	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13
2	1	0,01573	0,253	0,464	1,698	0,763	0,94	0,544	0,967	0,434	0,01	0,008	0,107	1,149	0,697
3	2	0,00014	0,169	0,322	2,261	0,728	0,98	0,525	0,935	0,439	0,006	0,005	0,039	1,232	0,999
4	3	0,02727	0,362	0,643	2,087	0,756	0,898	0,563	0,931	0,396	0,028	0,01	0,014	1,191	0,728
5	4	0,11356	1,355	2,608	0,454	0,615	0,741	0,519	0,944	0,449	0,068	0,046	0,01	1,742	0,608
6	5	0,09679	0,609	0,885	1,255	0,763	0,99	0,689	0,973	0,292	0,019	0,012	0,009	2,278	0,339
7	6	0,06789	1,447	2,622	0,439	0,631	0,728	0,552	0,943	0,415	0,021	0,007	0,006	1,405	0,547
8	7	0,02681	0,428	0,657	1,924	0,594	0,751	0,657	0,624	0,341	0,042	0,025	0,016	1,381	0,718
9	8	0,15355	1,699	2,908	0,403	0,669	0,683	0,584	0,973	0,399	0,085	0,043	0,038	1,734	0,102
10	9	0,23133	1,826	3,278	0,381	0,668	0,73	0,557	0,922	0,395	0,081	0,068	0,021	1,391	0,748
11	10	0,351	1,718	3,145	0,365	0,613	0,685	0,546	0,954	0,428	0,066	0,06	0,011	1,319	0,571

Рисунок 4.5 – Фрагмент файла исходных данных в формате MS Excel

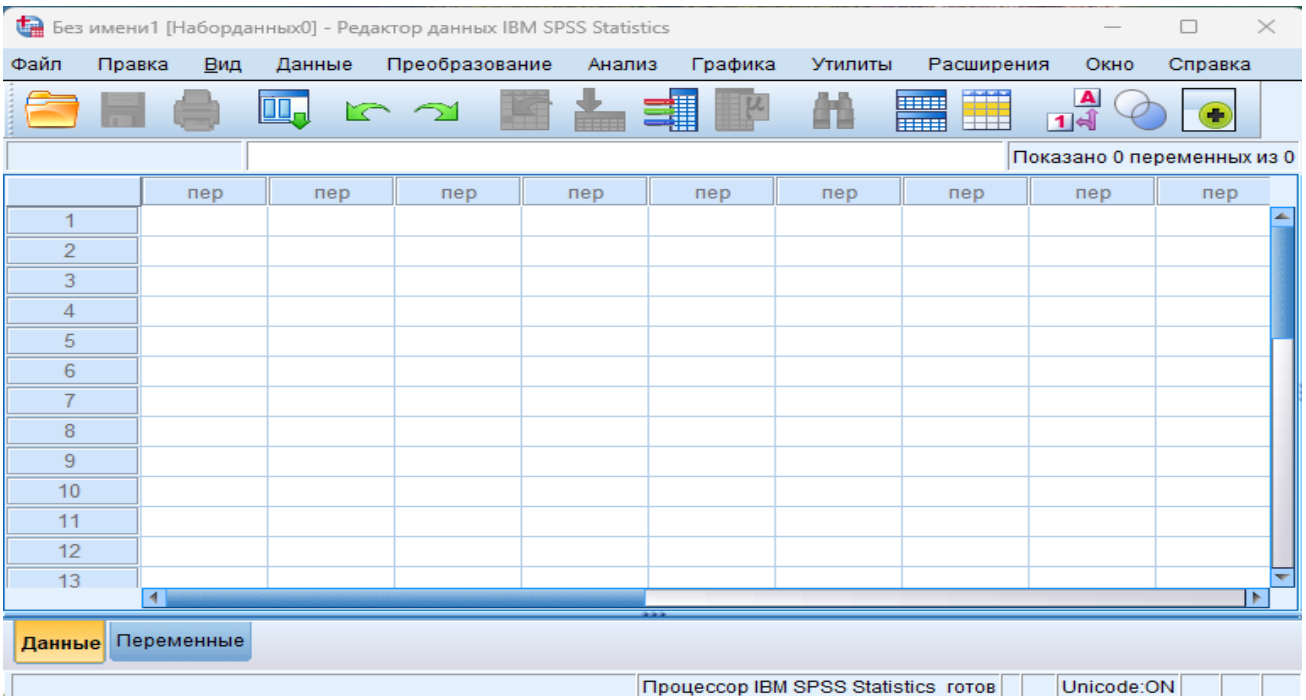


Рисунок 4.6 – Окно редактора данных в IBM SPSS Statistics

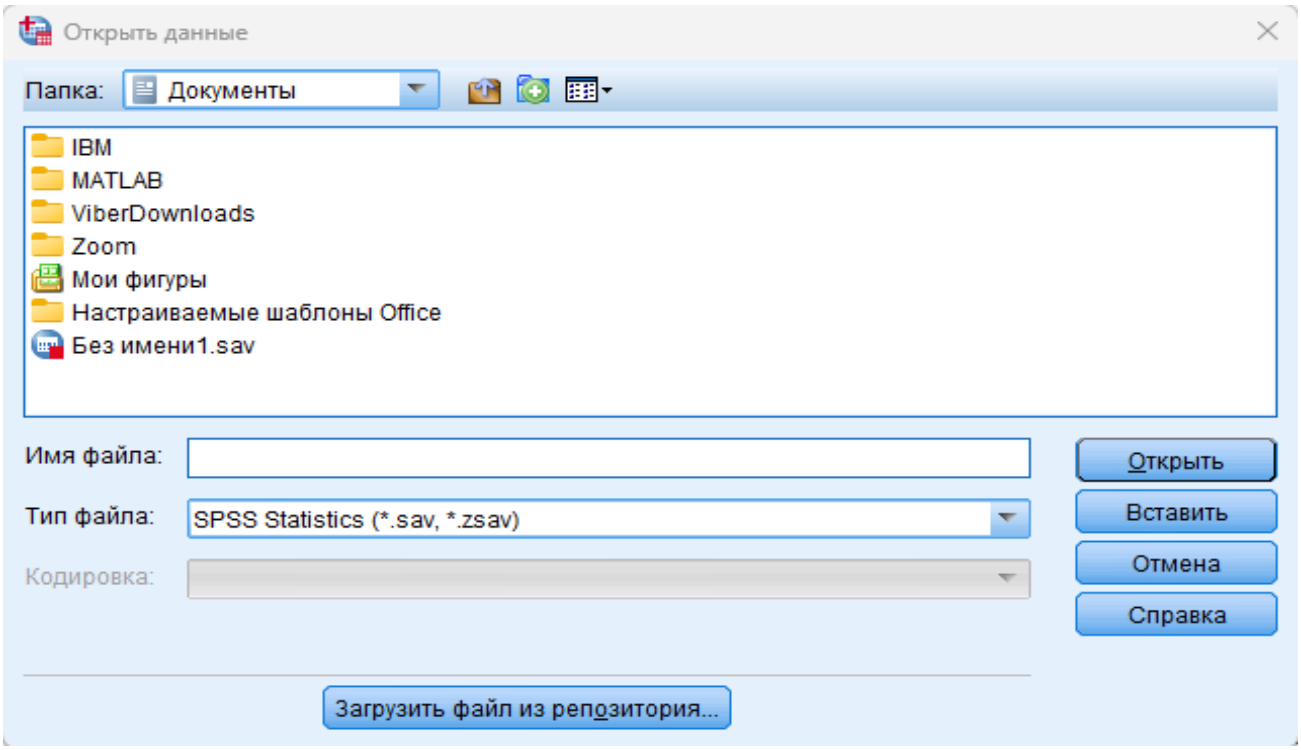


Рисунок 4.7 – Импорт данных в IBM SPSS Statistics

ШАГ 1.2. В новом диалоговом окне задайте диапазон ячеек. Если в первой строке Excel-файла находятся имена переменных, следует установить флажок *Читать имена переменных из первой строки данных*. Щелкнем на кнопке *ОК*.

На экране появятся импортированные данные в формате SPSS (рис. 4.8).

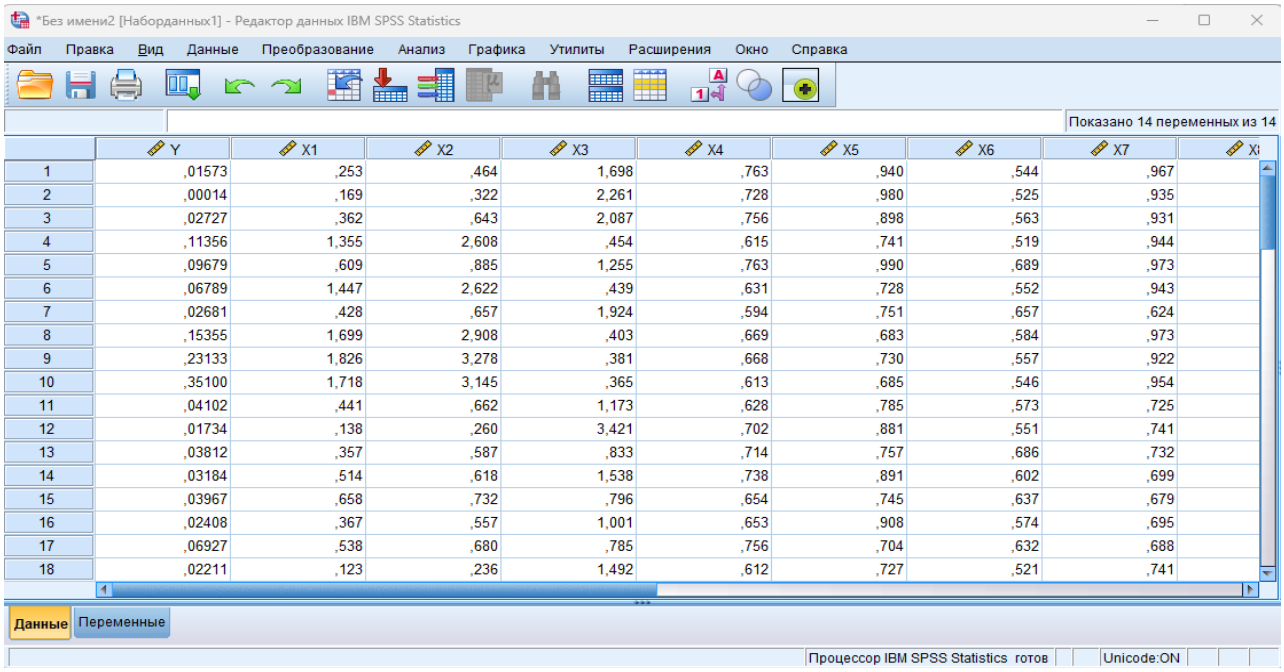


Рисунок 4.8 – Фрагмент данных задания представлены в формате SPSS

ШАГ 1.3. Для редактирования данных с использованием пакета SPSS нажмем в нижней строке меню кнопку *Переменные*. Это позволит просмотреть всю информацию о переменных, импортированных в SPSS-файл, и внести нужные изменения и дополнения. В столбце «Метка» (рис. 4.9) введем расширенное имя переменной (до 256 символов). Эта информация будет использована при создании отчетов.

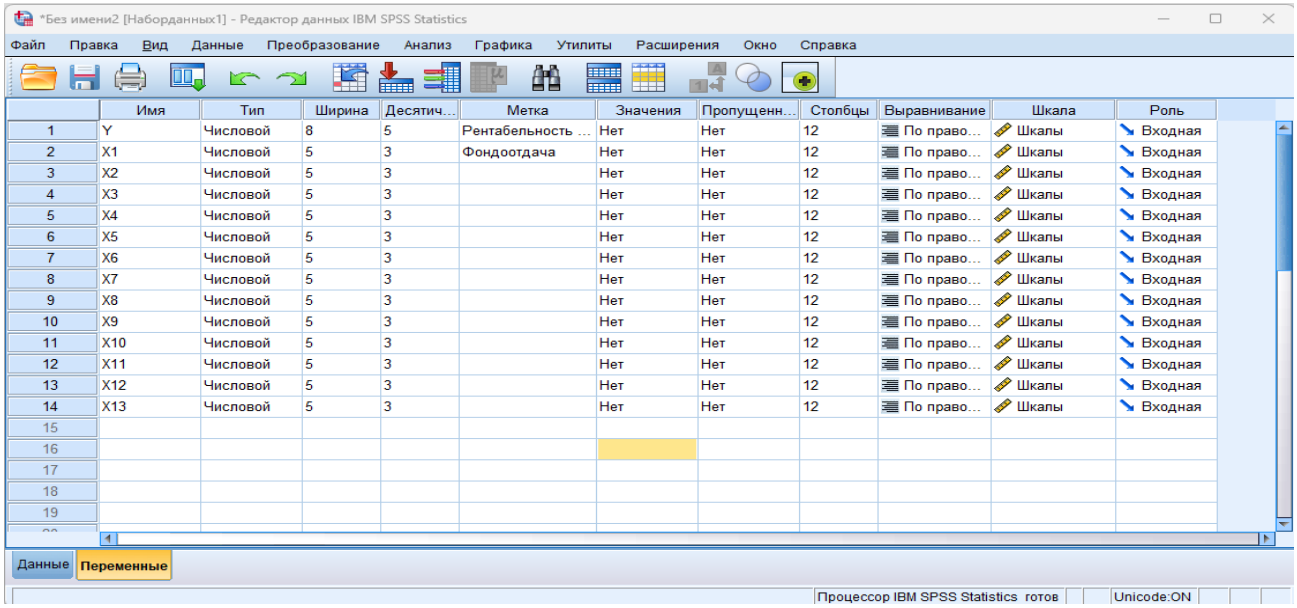


Рисунок 4.9 – Редактирование данных IBM SPSS Statistics

ШАГ 1.4. Каждой переменной в файле данных SPSS будет соответствовать отдельная колонка. Используя меню *Файл – Сохранить как*, сохраним данный файл с расширением <name>.sav (рис. 4.10).

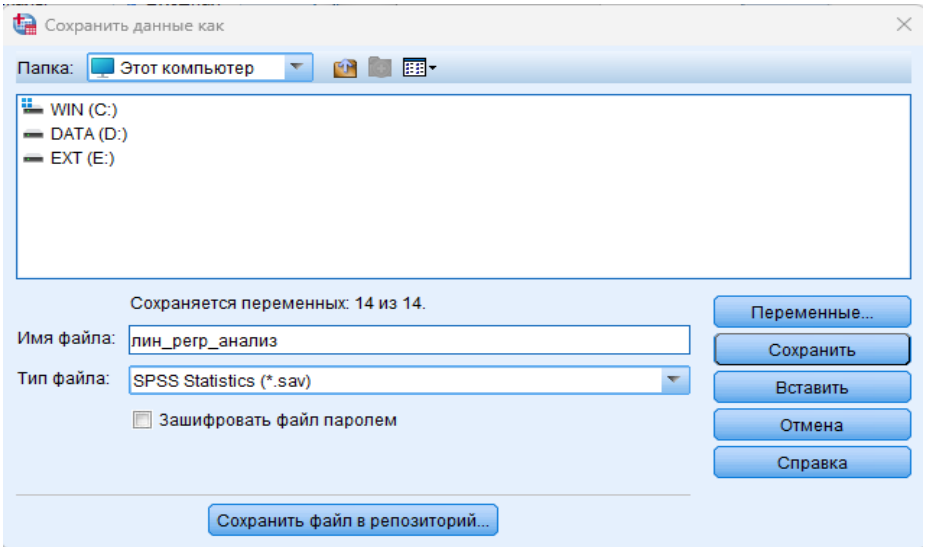


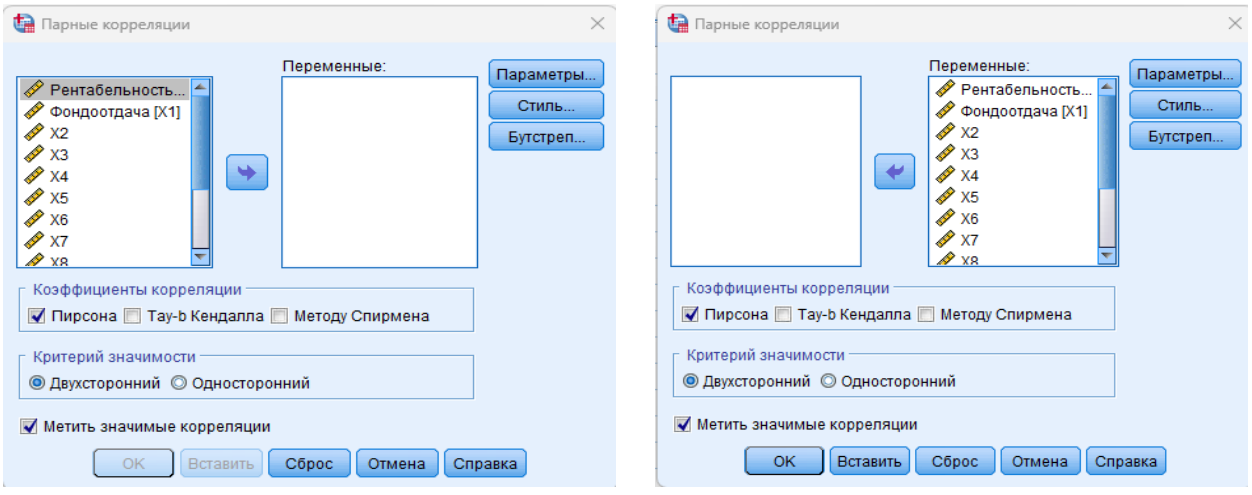
Рисунок 4.10 – Сохранение данных в IBM SPSS Statistics

ШАГ 1.5. Решение задания в IBM SPSS Statistics. Выбор факторных признаков для построения регрессионной модели на основе анализа матрицы коэффициентов корреляции. Для построения матрицы парной корреляции всех переменных с помощью пакета IBM SPSS Statistics выполним следующие действия.

Выберите в верхней строке меню *Анализ – Корреляция – Парные*.

Переменные, относительно которых проверяется степень корреляционной связи, поочередно переместим в поле тестируемых переменных справа (рис. 4.11).

Для начала расчета нажмите кнопку ОК.



а *б*

Рисунок 4.11 – Построение матрицы «Парные корреляции» в IBM SPSS Statistics

В результате в выходной области появится матрица парной корреляции всех переменных (рис. 4.12). Полученные результаты содержат коэффициент

корреляции Пирсона, вероятность ошибки, соответствующую предположению о ненулевой корреляции, и количество использованных пар значений (в нашем случае оно равно 54).

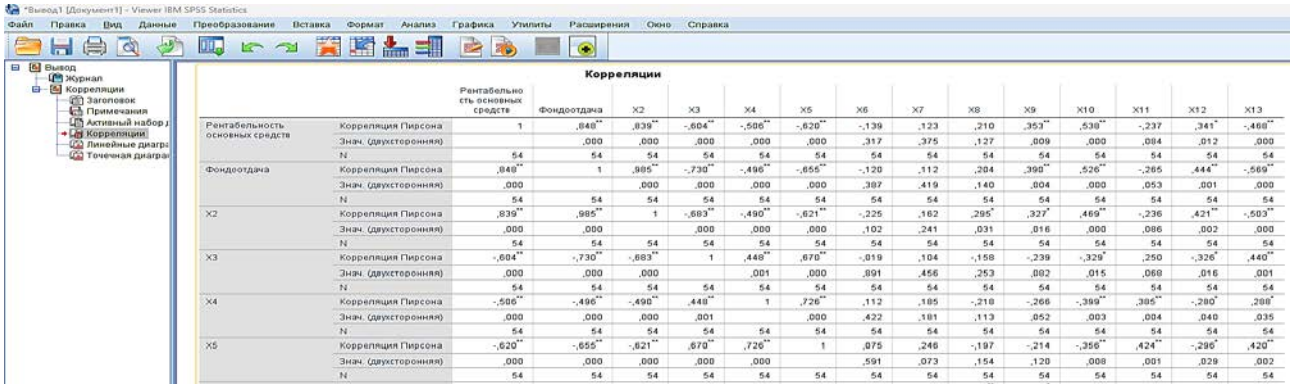


Рисунок 4.12 – Фрагмент корреляции в IBM SPSS Statistics

Коэффициент корреляции R между двумя переменными указывает силу связи между ними и принимает значения от -1 до +1: если значение находится ближе к единице, это означает наличие сильной связи, а если ближе к нулю, то слабой.

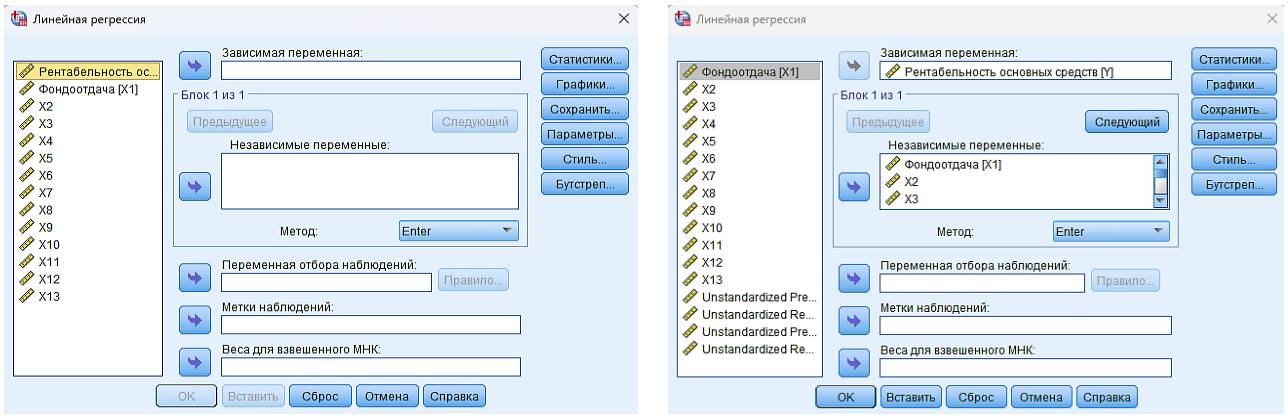
На основе рисунка 4.12 выполните анализ матрицы коэффициентов парной корреляции и сделайте вывод о целесообразности построения множественного регрессионного уравнения $Y = f(X_i \ldots X_n)$.

ШАГ 2. Построение линейного уравнения регрессии. По заданию необходимо построить регрессионную модель со значимыми факторами в линейной форме.

Оценка параметров регрессии осуществляется методом наименьших квадратов (МНК). Для проведения регрессионного анализа с помощью пакета SPSS выполните следующие действия.

Выберите в верхней строке меню *Анализ – Регрессия – Линейная*.

Поместите переменную *Y* в поле для зависимых переменных, объявив переменные $X_1 – X_{13}$ независимыми (рис. 4.13).



а *б*

Рисунок 4.13 – Диалоговое окно «Линейная регрессия» в IBM SPSS Statistics

ШАГ 2.1. Используйте метод исключения (или метод *«Назад»*), установленный в поле *Метод* (рис. 4.13).

В полях панели *Статистики* отметим флажками *Оценки. Согласие модели и Дарбин – Уотсон* (критерий Дарбина – Уотсона) (рис. 4.14), затем нажмите Продолжить.

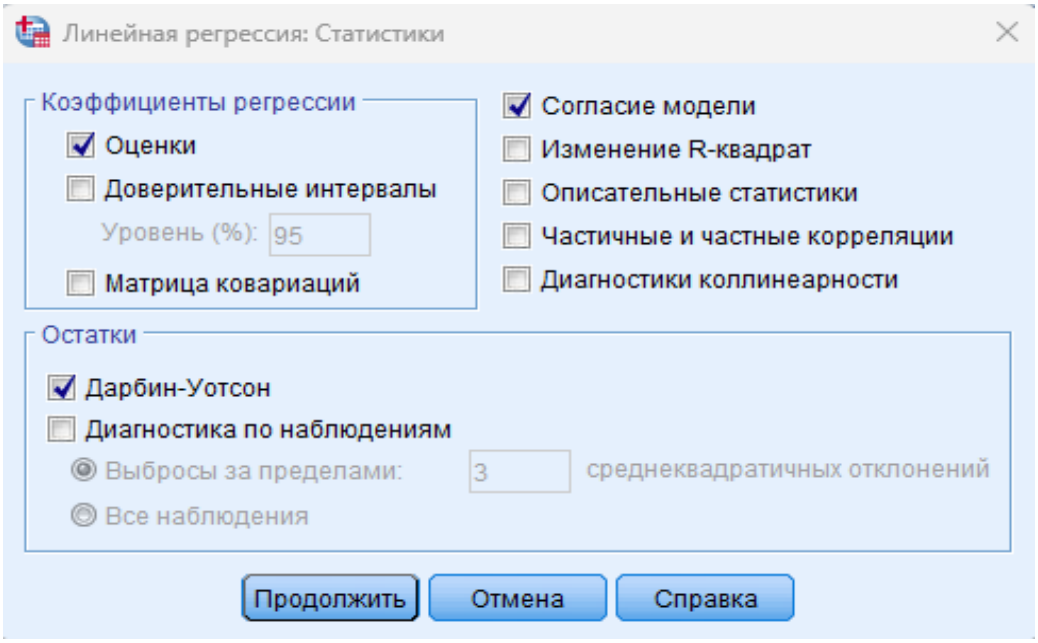


Рисунок 4.14 – Диалоговое окно «Линейная регрессия: Статистики» в SPSS

В полях панели *Сохранить* отметим флажками необходимые поля (рис. 4.15) и нажмите *Продолжить*.

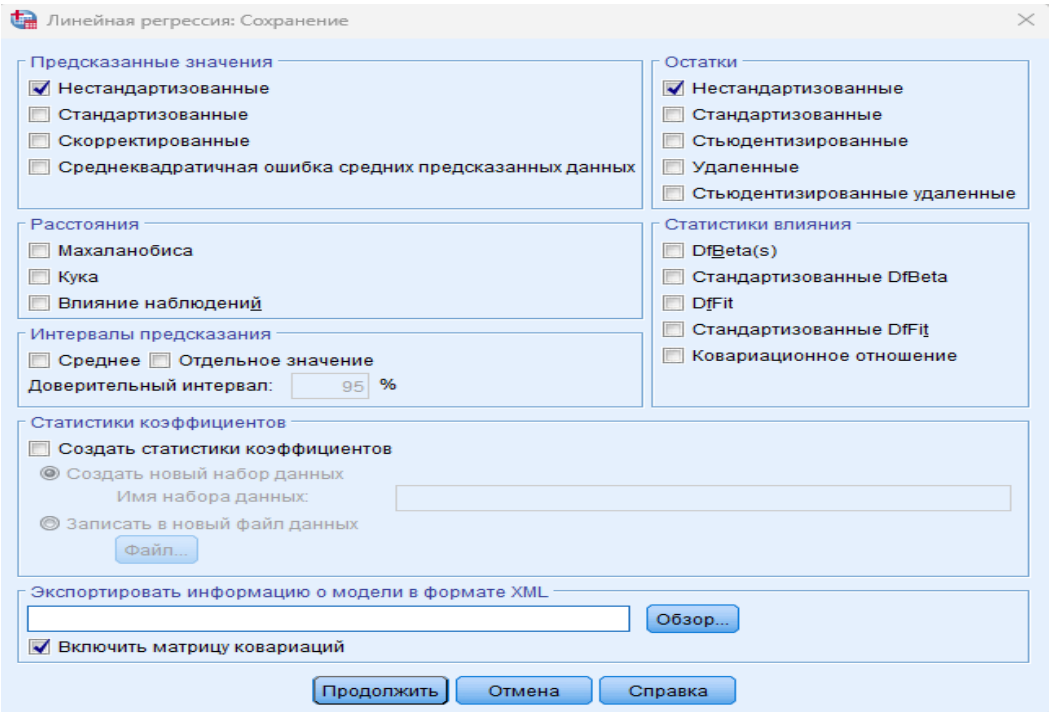


Рисунок 4.15 – Диалоговое окно «Линейная регрессия: Сохранение» в SPSS

Значения, установленные по умолчанию в полях панели *Параметры*, не меняем; они будут использованы для определения критерия Фишера с доверительной вероятностью 0,95. Начните вычисления нажатием *ОК*.

Результаты регрессионного анализа выводятся в табличном виде.

При последовательном подборе переменных в SPSS предусмотрена автоматизация, основанная на значимости включения и исключения переменных.

ШАГ 2.2. В таблице 4.4 «Включенные/исключенные переменные» перечислены переменные, которые были последовательно исключены на каждом шаге.

Таблица 4.4 – Введенные/удаленные переменные в IBM SPSS Statistics

Модель	Введенные переменные	Удаленные переменные	Метод
1	X13, X7, X8, X12, X11, X9, X3, X4, X2, X5, X6, X10, X1	.	Enter
2	.	X3	Обратный (критерий: Вероятность F для исключения >= ,100).
3	.	X8	Обратный (критерий: Вероятность F для исключения >= ,100).
4	.	X6	Обратный (критерий: Вероятность F для исключения >= ,100).
5	.	X4	Обратный (критерий: Вероятность F для исключения >= ,100).
6	.	X2	Обратный (критерий: Вероятность F для исключения >= ,100).
7	.	X7	Обратный (критерий: Вероятность F для исключения >= ,100).
8	.	X11	Обратный (критерий: Вероятность F для исключения >= ,100).
9	.	X5	Обратный (критерий: Вероятность F для исключения >= ,100).
10	.	X13	Обратный (критерий: Вероятность F для исключения >= ,100).
11	.	X12	Обратный (критерий: Вероятность F для исключения >= ,100).
a. Зависимая переменная: Y – рентабельность основных средств b. Все требуемые переменные введены.			

ШАГ 2.3. В таблице «Сводка для модели» (табл. 4.5) приведены значения коэффициента детерминации, коэффициента множественной корреляции, стандартной ошибки, коэффициента Дарбина – Уотсона последовательно для всех моделей. В последней строке приводятся данные для окончательной модели.

Таблица 4.5 – Сводка для модели в IBM SPSS Statistics

Модель	R	R-квадрат	Скорректированный R-квадрат	Стандартная ошибка оценки	Дарбин – Уотсон
1	,893[a]	,798	,732	,04245371	
2	,893[b]	,798	,738	,04193284	
3	,893[c]	,798	,745	,04143079	
4	,893[d]	,798	,751	,04094632	
5	,893[e]	,798	,756	,04048708	
6	,893[f]	,797	,761	,04005081	
7	,893[g]	,797	,766	,03965186	
8	,893[h]	,797	,771	,03925880	
9	,892[i]	,795	,774	,03897051	
10	,890[j]	,792	,775	,03890842	
11	**,885[k]**	**,784**	**,771**	**,03925552**	**2,046**

a. Предикторы: (константа), X13, X7, X8, X12, X11, X9, X3, X4, X2, X5, X6, X10, X1

b. Предикторы: (константа), X13, X7, X8, X12, X11, X9, X4, X2, X5, X6, X10, X1
c. Предикторы: (константа), X13, X7, X12, X11, X9, X4, X2, X5, X6, X10, X1
d. Предикторы: (константа), X13, X7, X12, X11, X9, X4, X2, X5, X10, X1
e. Предикторы: (константа), X13, X7, X12, X11, X9, X2, X5, X10, X1
f. Предикторы: (константа), X13, X7, X12, X11, X9, X5, X10, X1
g. Предикторы: (константа), X13, X12, X11, X9, X5, X10, X1
h. Предикторы: (константа), X13, X12, X9, X5, X10, X1
i. Предикторы: (константа), X13, X12, X9, X10, X1
j. Предикторы: (константа), X12, X9, X10, X1
k. Предикторы: (константа), X9, X10, X1
l. Зависимая переменная: *Y* – рентабельность основных средств.

ШАГ 2.4. В таблице 4.6 приведены результаты дисперсионного анализа и значения *F*-критерия, полученные на каждом шаге (здесь ст. св. – число степеней свободы, Знч. – значимость).

Таблица 4.6 – Дисперсионный анализ (ANOVA) в IBM SPSS Statistics

Модель		Сумма квадратов	ст.св.	Средний квадрат	F	Значимость
1	Регрессия	0,284	13	0,022	12,128	0,000[b]
	Остаток	0,072	40	0,002		
	Всего	0,356	53			
…	Регрессия					
	Остаток					
	Всего					
11	Регрессия	0,279	3	0,093	**60,396**	**0,000**[l]
	Остаток	0,077	50	0,002		
	Всего	0,356	53			

a. Зависимая переменная: Y – рентабельность основных средств
b. Предикторы: (константа), X13, X7, X8, X12, X11, X9, X3, X4, X2, X5, X6, X10, X1
…
l. Предикторы: (константа), X9, X10, X1

ШАГ 2.5. Выполните регрессионный анализ данных на основе полученных результатов в таблице 4.7 и запишите линейное уравнение регрессии зависимости рентабельности основных средств от всех факторов, полученное на последнем шаге. Сделайте развернутый вывод.

Отметим, таблица 4.7 содержит следующие данные:

первый столбец – номер модели;

второй – используемые в модели независимые переменные;

третий – коэффициенты уравнения регрессии;

четвертый – стандартные ошибки коэффициентов уравнения регрессии;

пятый – стандартизованные коэффициенты;

шестой – *t*-статистика, используемая для проверки значимости коэффициентов уравнения регрессии.

Таблица 4.7 – Коэффициенты регрессионного анализа в IBM SPSS Statistics

Модель		Нестандартизованные коэффициенты		Стандартизованные коэффициенты	t	Значимость
		B	Стандартная ошибка	Бета		
1	(Константа)	0,012	0,248		0,048	0,962
	X1	0,088	0,103	0,576	0,856	0,397
	X2	0,009	0,051	0,113	0,181	0,857
	X3	0,000	0,012	0,001	0,010	0,992
	X4	0,020	0,152	0,015	0,132	0,896
	X5	-0,051	0,116	-0,062	-0,443	0,660
	X6	0,005	0,219	0,004	0,024	0,981
	X7	0,011	0,054	0,019	0,198	0,844
	X8	0,005	0,241	0,003	0,019	0,985
	X9	-0,869	0,307	-0,716	-2,827	0,007
	X10	2,252	0,686	0,926	3,283	0,002
	X11	-0,101	0,312	-0,030	-0,322	0,749
	X12	-0,029	0,024	-0,105	-1,198	0,238
	X13	0,035	0,047	0,074	0,737	0,466
…	(Константа)	0,012	0,239		0,052	0,959
	X1	0,088	0,095	0,573	0,923	0,361
	X2	0,009	0,049	0,114	0,192	0,849
	X4	0,020	0,150	0,015	0,133	0,895
	X5	-0,051	0,111	-0,062	-0,459	0,649
	X6	0,005	0,212	0,003	0,023	0,982
	X7	0,011	0,054	0,019	0,202	0,841
	X8	0,004	0,231	0,003	0,018	0,986
	X9	-0,869	0,303	-0,717	-2,870	0,006
	X10	2,253	0,672	0,926	3,354	0,002
	X11	-0,101	0,307	-0,030	-0,328	0,745
	X12	-0,029	0,024	-0,105	-1,217	0,231
	X13	0,035	0,047	0,074	0,747	0,459
11	(Константа)	**-0,016**	**0,009**		**-1,831**	**0,073**
	X1	**0,103**	**0,013**	**0,675**	**8,120**	**0,000**
	X9	**-0,890**	**0,255**	**-0,733**	**-3,485**	**0,001**
	X10	**2,124**	**0,554**	**0,873**	**3,834**	**0,000**

а. Зависимая переменная: Y – рентабельность основных средств

ШАГ 3. Выполните оценку качества модели и сделайте выводы. Оценка качества всего уравнения регрессии производится с помощью коэффициентов детерминации R^2 и множественной корреляции R. Их значения можно найти в таблице 4.5 «Сводка для модели».

Выполните проверку значимости уравнения регрессии на основе F-критерия Фишера и сделайте вывод. Значение критерия Фишера можно найти в таблице 4.6 «Дисперсионный анализ».

ШАГ 4. Постройте прогноз по модели регрессии и сделайте выводы. Для построения прогноза вначале введите прогнозные значения выбранных факторов $X_1 = 1{,}549$; $X_9 = 0{,}406$; $X_{10} = 0{,}226$. (рис. 4.16).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13
41	0,10166	1,051	2,05	0,59	0,62	0,748	0,512	0,513	0,456	0,067	0,044	0,022	1,251	0,627
42	0,00124	0,658	0,935	1,407	0,751	0,98	0,708	0,704	0,273	0,328	0,104	0,023	1,386	0,798
43	0,10233	1,415	2,566	0,455	0,624	0,715	0,551	0,552	0,417	0,074	0,038	0,008	1,647	0,574
44	0,02371	0,384	0,621	2,041	0,618	0,775	0,646	0,617	0,337	0,031	0,017	0,011	1,403	0,753
45	0,1243	1,468	2,496	0,495	0,703	0,725	0,588	0,588	0,396	0,071	0,053	0,024	1,486	0,147
46	0,16733	1,447	2,562	0,441	0,632	0,712	0,565	0,565	0,391	0,028	0,013	0,008	1,275	0,704
47	0,28663	1,502	2,746	0,398	0,591	0,674	0,547	0,547	0,429	0,025	0,021	0,008	1,081	0,538
48	0,03417	0,421	0,657	1,162	0,637	0,792	0,661	0,714	0,301	0,042	0,025	0,022	1,471	0,654
49	0,01482	0,136	0,253	3,402	0,716	0,895	0,614	0,721	0,351	0,011	0,007	0,016	1,294	0,781
50	0,0254	0,347	0,583	0,826	0,721	0,874	0,629	0,675	0,334	0,019	0,011	0,013	1,179	0,714
51	0,03307	0,521	0,623	1,541	0,732	0,886	0,637	0,704	0,323	0,023	0,014	0,017	1,031	0,642
52	0,0438	0,673	0,748	0,804	0,649	0,736	0,653	0,692	0,304	0,037	0,021	0,015	1,282	0,617
53	0,02741	0,385	0,561	1,022	0,763	0,894	0,581	0,713	0,382	0,017	0,009	0,012	1,108	0,721
54	0,06504	0,532	0,674	0,781	0,604	0,728	0,625	0,671	0,336	0,053	0,031	0,028	1,438	0,677
55	0,02718	0,127	0,241	1,506	0,647	0,714	0,538	0,765	0,423	0,014	0,007	0,009	1,171	0,736
56	-	1,549	-	-	-	-	-	-	-	0,406	0,226	-	-	-

Рисунок 4.16 – Введены прогнозные значения объясняющих факторов

Затем запускаем процедуру *Анализ – Регрессия – Линейная*, в качестве независимых переменных указывайте все факторы, на основе которых получено линейное уравнение регрессии на последнем шаге (ШАГ 2.5) (рис. 4.17).

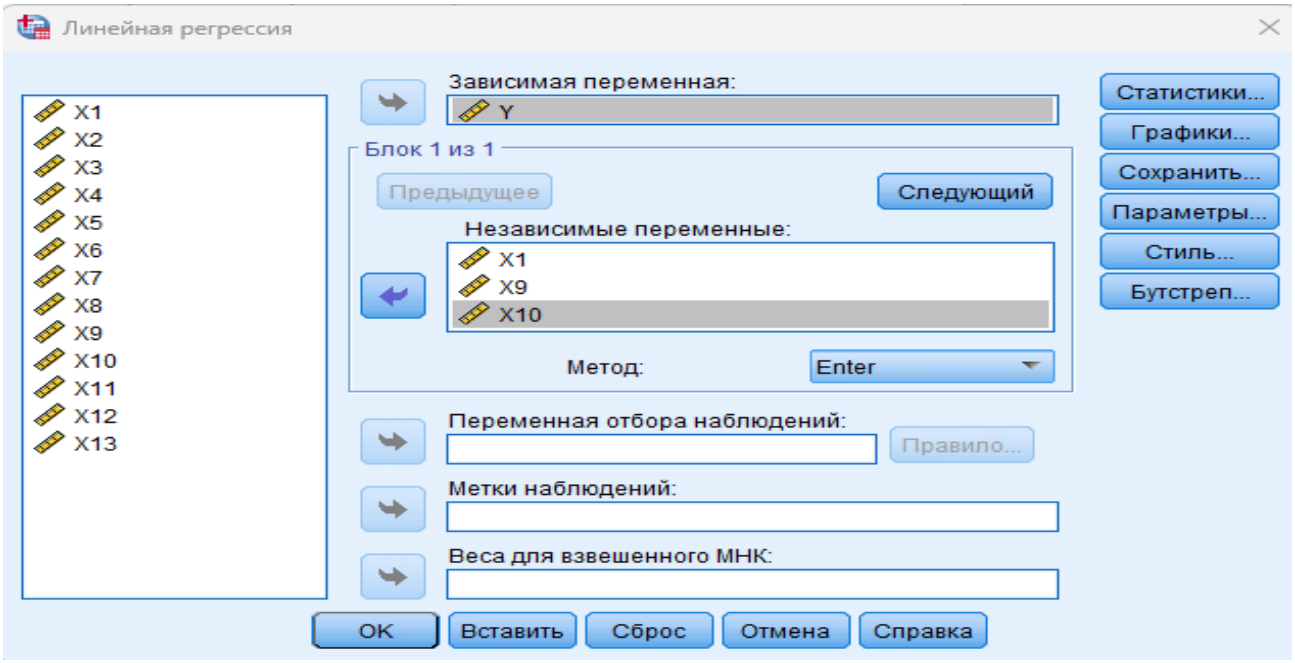


Рисунок 4.17 – Заполнение диалогового окна «Линейная регрессия» в SPSS

В полях панели *Сохранить* следует обязательно отметить необходимость сохранения интервалов предсказания для отдельных значений с вероятностью 90 % (рис. 4.18) и нажать *Продолжить*.

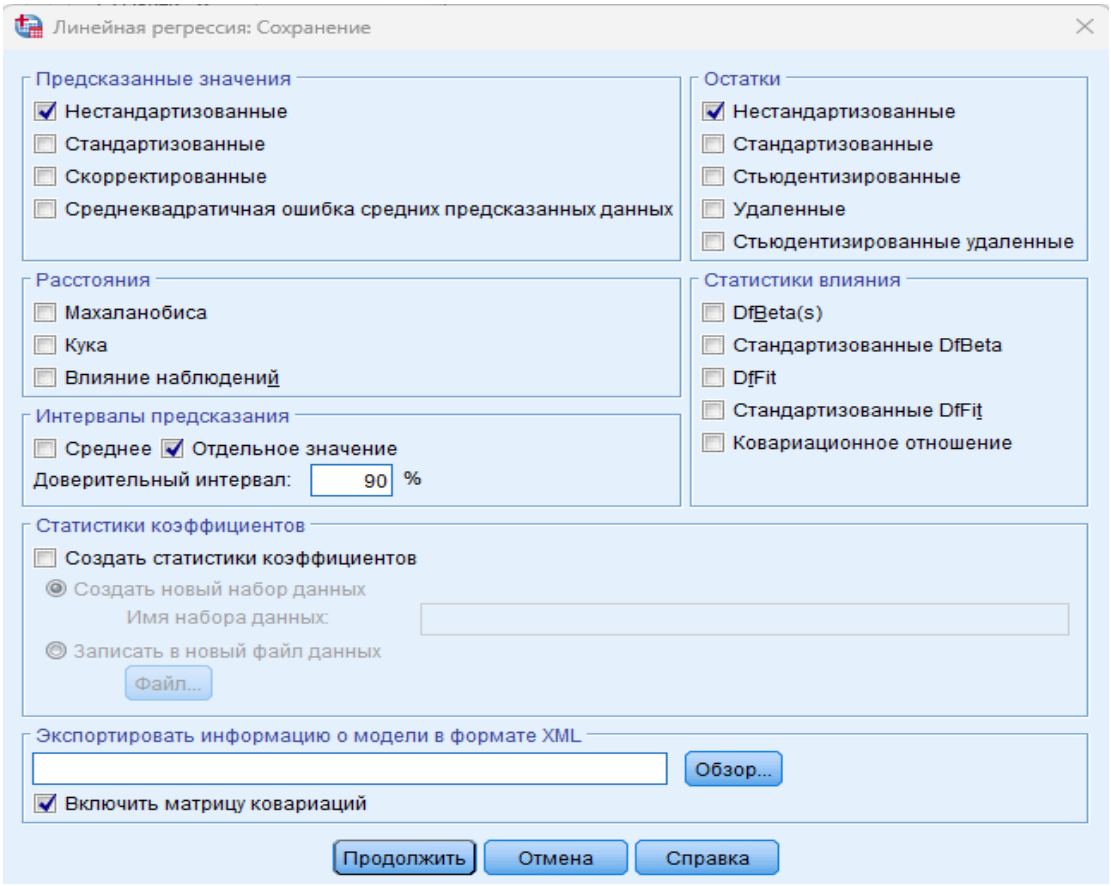


Рисунок 4.18 – Заполнение поля «Линейная регрессия: Сохранение» в SPSS

В таблице 4.8 приведены результаты прогнозирования по модели регрессии: точечный прогноз, верхняя и нижняя границы.

Таблица 4.8 – Статистика остатков и результаты прогнозирования в IBM SPSS Statistics

	Минимум	Максимум	Среднее	Стандартные отклонения	N
Предсказанное значение	-0,0193167	0,2946694	0,0676096	0,07258206	54
Стандартная «Предсказанное значение»	-1,198	3,128	0,000	1,000	54
Стандартная ошибка предсказанного значения	0,005	0,035	0,009	0,006	54
Скорректированное предсказываемое значение	-0,0941318	0,3008714	0,0661909	0,07494465	54
Остаток	-0,08752222	0,12565912	0,00000000	0,03812833	54
Стандартная «Остаток»	-2,230	3,201	0,000	0,971	54
Стьюдентизированный остаток	-2,318	3,353	0,012	1,026	54
Остаток для исключаемого наблюдения	-0,09458204	0,13959719	0,00141871	0,04373642	54
Стьюдентизированный остаток для исключаемого наблюдения	-2,429	3,770	0,028	1,094	54
Расстояние Махалонобиса	0,041	40,595	2,944	7,584	54
Расстояние Кука	0,000	1,158	0,046	0,170	54
Значение центрированной балансировки	0,001	0,766	0,056	0,143	54

a. Зависимая переменная: Y – Рентабельность основных средств.

Задания для самостоятельного решения по темам 2–4

Задание 2.3. Приведите все возможные способы вычисления статистической формулы множественного коэффициента (индекса) корреляции в MS Word

Наименование способа или величины для вычисления	Формула вычисления множественного коэффициента (индекса) корреляции и его составляющие
Матричный	
Дисперсионный	
Парные коэффициенты корреляции	
Регрессионные остатки	
Множественный коэффициент детерминации	
Уравнение в стандартизированном масштабе	

Задание 3.3. Выполните линеаризацию нелинейных регрессий в MS Word.

Регрессия	Исходное уравнение	Преобразованное уравнение

Задание 3.4. На основе приложения Б запишите в MS Word наименование и формулы расчета всех показателей раздела «Регрессионная статистика» программы MS Excel.

Наименование в отчете Excel	Принятые наименования в статистике	Формула и ее составляющие

Задание 3.5. Дайте характеристику всем вариантам на рисунке 4.19 (а, б, в) и сделайте вывод о взаимодействии трех переменных в уравнении регрессии.

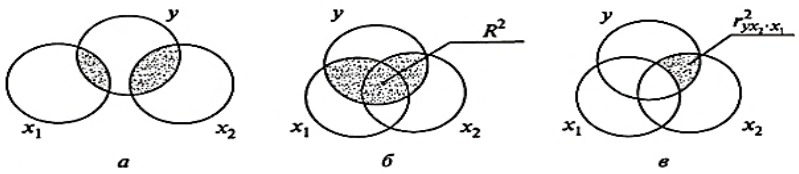


Рисунок 4.19 – Взаимодействие трех переменных в уравнении регрессии

Охарактеризуйте геометрический смысл коэффициента детерминации при использовании одной и двух объясняющих переменных (рис. 4.20 а, б).

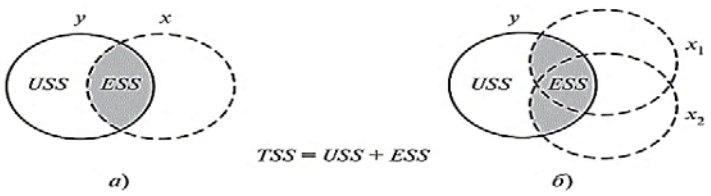


Рисунок 4.20 – Геометрический смысл коэффициента детерминации

Задание 3.6. На основе приложения Б запишите в MS Word наименование и формулы расчета всех показателей раздела «Дисперсионный анализ» программы MS Excel.

	df	SS	MS	*F*	*Значимость F*

Задание 4.6. На основании приведенных данных в задании 1.1 постройте множественную линейную регрессионную модель производительности труда одного рабочего по группе предприятий.

1. Рассчитайте уравнение множественной линейной регрессии.
2. Оцените тесноту связи между анализируемыми признаками с помощью множественного коэффициента корреляции.
3. Оцените качество уравнения регрессии с помощью множественного коэффициента детерминации и статистическую значимость уравнения множественной регрессии с помощью F-критерия Фишера (α = 0,05).
4. Оцените статистическую значимость коэффициентов множественной регрессии с помощью t-критерия Стьюдента (α = 0,05).
5. Рассчитайте доверительные интервалы для каждого коэффициента a_j линейного уравнения регрессии.
6. Выполните все этапы задания с помощью команды *Анализ данных* в MS Excel. Результаты задания представьте на рабочих листах Excel в соответствии с требованиями методических указаний к заданию.

Методические указания к выполнению задания 4.6

Выполните задание в MS Excel с помощью вкладки *Данные* – команды *Анализ данных* – инструментов *Корреляция* и *Регрессия.* Загрузите (или скопируйте) в MS Excel данные задания 1.1 для выполнения линейного множественного корреляционно-регрессионного анализа.

Пошаговый множественный корреляционно-регрессионный анализ в линейной форме с полным набором факторов выполните с помощью методических указаний к заданию 4.1 в Excel.

Поясните экономический смысл полученных результатов линейного множественного корреляционно-регрессионного анализа в Excel.

Задание 4.7. Имеются следующие статистические данные об одном из товаров коммерческой организации (USB Flash накопители), представленные в приложении К. Выполните линейный множественный корреляционно-регрессионный анализ данных в MS Excel и SPSS по следующим направлениям.

1. Осуществите выбор факторных признаков для построения регрессионной модели двумя способами:

а) на основе анализа матрицы коэффициентов парной корреляции;

б) с помощью шагового отбора методом исключения.

2. Выполните тест Фаррара – Глоубера на мультиколлинеарность.

3. Оцените параметры модели, дайте экономическую интерпретацию коэффициентов регрессии.

4. Для оценки качества модели:

а) определите коэффициент множественной корреляции и детерминации;

б) вычислите среднюю относительную ошибку аппроксимации;

в) осуществите оценку значимости уравнения регрессии.

5. Оцените с помощью *t*-критерия Стьюдента статистическую значимость коэффициентов уравнения множественной регрессии.

6. Оцените влияние факторов на зависимую переменную по модели.

7. Постройте точечный и интервальный прогноз результирующего показателя на два года вперед (α = 0,1).

Задание 4.8. Для анализа услуг связи по Витебской области отберите из статистических сборников четыре фактора, которые могут оказывать влияние на объем услуг связи, оказанных населению: наличие квартирных телефонных аппаратов сети общего пользования на 1000 человек населения (X1); число подключенных абонентских устройств подвижной радиотелефонной связи на 1000 человек населения (X2); число активных абонентов фиксированного широкополосного доступа к сети Интернет на 100 человек населения (X3); число активных абонентов подвижной радиотелефонной связи, использующих широкополосный доступ к сети Интернет на 100 человек населения (X4) (табл. 4.9).

Таблица 4.9 – Объем услуг связи и факторы, оказывающие на него влияние

№ п/п	Район	Y	X1	X2	X3	X4
1						
…						
n						

Требуется установить зависимость объема услуг связи (Y) от перечисленных факторов. Уровень вероятности суждения принять 0,95.

Выполните технологию решения линейного множественного корреляционно-регрессионного анализа данных с помощью шагового отбора методом исключения в MS Excel или Statistica.

Задание 4.9. Для анализа денежных доходов по Витебской области отберите из статистических сборников четыре фактора, которые могут оказывать влияние на среднедушевые денежные доходы населения (Y): среднемесячная номинальная начисленная заработная плата работников организаций (X1); средний размер назначенных пенсий (X2); численность населения с денежными доходами ниже величины прожиточного минимума (X3); численность занятых, приходящихся на одного пенсионера (X4). Требуется установить зависимость среднедушевых денежных доходов от перечисленных факторов в программе IBM SPSS. Уровень вероятности суждения принять 0,95.

ПРАКТИЧЕСКИЕ ЗАНЯТИЯ 5–7
«Снижение размерности признакового пространства с использованием статистических программных пакетов»

Цель занятия: получить навыки выполнения многомерного компонентного и факторного анализа данных с применением статистических программных пакетов IBM SPSS Statistics и STATISTICA».

Задание 5.1. Имеется информация по использованию основных средств коммерческими организациями за год (приложение Ж).

Выполните пошаговый алгоритм многомерного факторного анализа в IBM SPSS Statistics.

1. Выполните подготовительные шаги в программе IBM SPSS Statistics для факторного анализа экономических данным.
2. Выполните оценку применимости факторного анализа к выбранным данным (выборки) на основе следующих критериев: теста Кайзера-Майера-Олкина (КМО) и критерия сферичности Бартлетта.
3. Оцените важность каждой переменной в факторной модели с помощью анализа общности (таблица «Общности») и анализа коэффициентов корреляции Пирсона между рассматриваемыми переменными.
4. Выделите факторы для факторного анализа с помощью метода главных компонент и определите число значимых факторов (общих факторов) для последующего анализа на основе 1-го из 3-х критериев: критерий Кайзера, критерий Кеттелла или дисперсии.
5. Выполните факторный анализ без методов вращения и с помощью методов ортогонального вращения (метод «Варимакс»). По результатам сравнительного факторного анализа сделать выводы об изменении значений переменных по факторным нагрузкам.
6. Присвойте названия (имена) общим факторам из смыслового содержания входящих в них переменных, и дайте их экономическую интерпретацию.
7. Представьте в табличной форме в MS Word пошаговые результаты многомерного факторного анализа использования основных средств организаций и сделайте развернутые выводы.

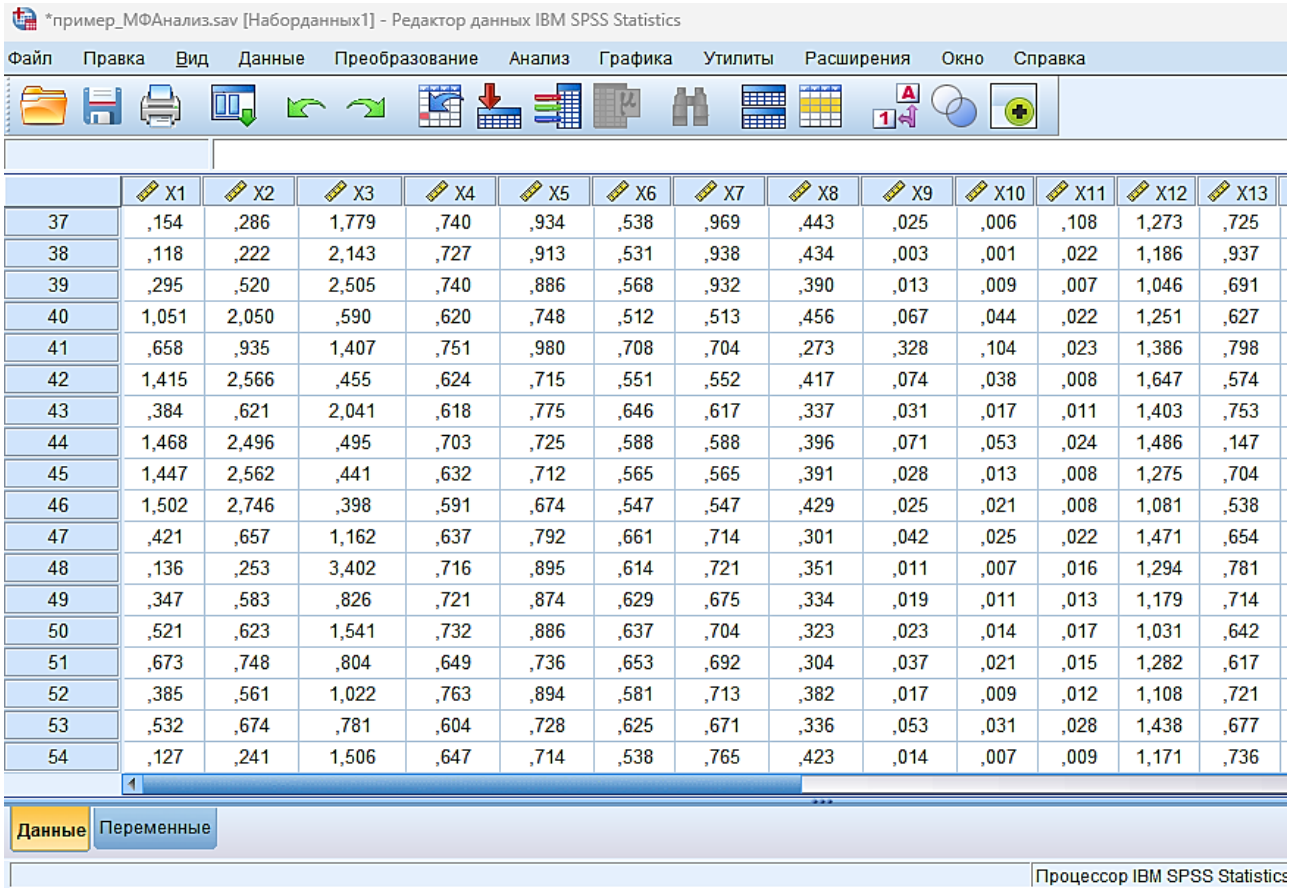
Методические указания к выполнению задания 5.1

При проведении многомерного факторного анализа сначала выполните ряд подготовительных шагов. Эти шаги (шаги 1.1 – 1.2) позволяют подготовить рабочий файл данных, запустить программу IBM SPSS Statistics 26 (версия 26 программы или др.) и открыть файл с исходными данными для анализа.

ШАГ 1. Группируя данные по показателям (по колонкам), используйте матрицу наблюдений, описывающую данные измерений 54 объектов по 13 соответствующим переменным x_1, ..., x_{13} (приложение Ж):

$$X = \begin{pmatrix} x_{1,1} & \cdots & x_{1,13} \\ \vdots & \ddots & \vdots \\ x_{54,1} & \cdots & x_{54,13} \end{pmatrix}.$$

ШАГ 1.1. Для описания данных используйте редактор данных в SPSS, т.е. рабочий лист, каждый столбец которого соответствует отдельной переменной, а каждая строка – отдельному наблюдению, в каждой ячейке хранится одно наблюдаемое значение переменной (рис. 5.1).



	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13
37	,154	,286	1,779	,740	,934	,538	,969	,443	,025	,006	,108	1,273	,725
38	,118	,222	2,143	,727	,913	,531	,938	,434	,003	,001	,022	1,186	,937
39	,295	,520	2,505	,740	,886	,568	,932	,390	,013	,009	,007	1,046	,691
40	1,051	2,050	,590	,620	,748	,512	,513	,456	,067	,044	,022	1,251	,627
41	,658	,935	1,407	,751	,980	,708	,704	,273	,328	,104	,023	1,386	,798
42	1,415	2,566	,455	,624	,715	,551	,552	,417	,074	,038	,008	1,647	,574
43	,384	,621	2,041	,618	,775	,646	,617	,337	,031	,017	,011	1,403	,753
44	1,468	2,496	,495	,703	,725	,588	,588	,396	,071	,053	,024	1,486	,147
45	1,447	2,562	,441	,632	,712	,565	,565	,391	,028	,013	,008	1,275	,704
46	1,502	2,746	,398	,591	,674	,547	,547	,429	,025	,021	,008	1,081	,538
47	,421	,657	1,162	,637	,792	,661	,714	,301	,042	,025	,022	1,471	,654
48	,136	,253	3,402	,716	,895	,614	,721	,351	,011	,007	,016	1,294	,781
49	,347	,583	,826	,721	,874	,629	,675	,334	,019	,011	,013	1,179	,714
50	,521	,623	1,541	,732	,886	,637	,704	,323	,023	,014	,017	1,031	,642
51	,673	,748	,804	,649	,736	,653	,692	,304	,037	,021	,015	1,282	,617
52	,385	,561	1,022	,763	,894	,581	,713	,382	,017	,009	,012	1,108	,721
53	,532	,674	,781	,604	,728	,625	,671	,336	,053	,031	,028	1,438	,677
54	,127	,241	1,506	,647	,714	,538	,765	,423	,014	,007	,009	1,171	,736

Рисунок 5.1 – Фрагмент исходных данных для факторного анализа в IBM SPSS

Переменные в SPSS можно определить следующим образом: щелкнуть по ярлычку *Переменные (Variable view)* в левой нижней части таблицы (рис. 5.1). Так осуществляется переход в режим просмотра переменных, где последовательно, строка за строкой описываются необходимые переменные. При выборе имени переменной руководствуйтесь следующими правилами:

1. Имена переменных могут содержать буквы латинского алфавита и цифры, также допускаются некоторые специальные символы (без пробелов).
2. Имя переменной должно начинаться с буквы.
3. Последний символ имени не может быть точкой или знаком подчеркивания.
4. Длина имени не должна превышать 8 символов.

Затем задайте тип переменной (численная, строковая и др.) и форматы ввода, вывода переменной (рис. 5.2).

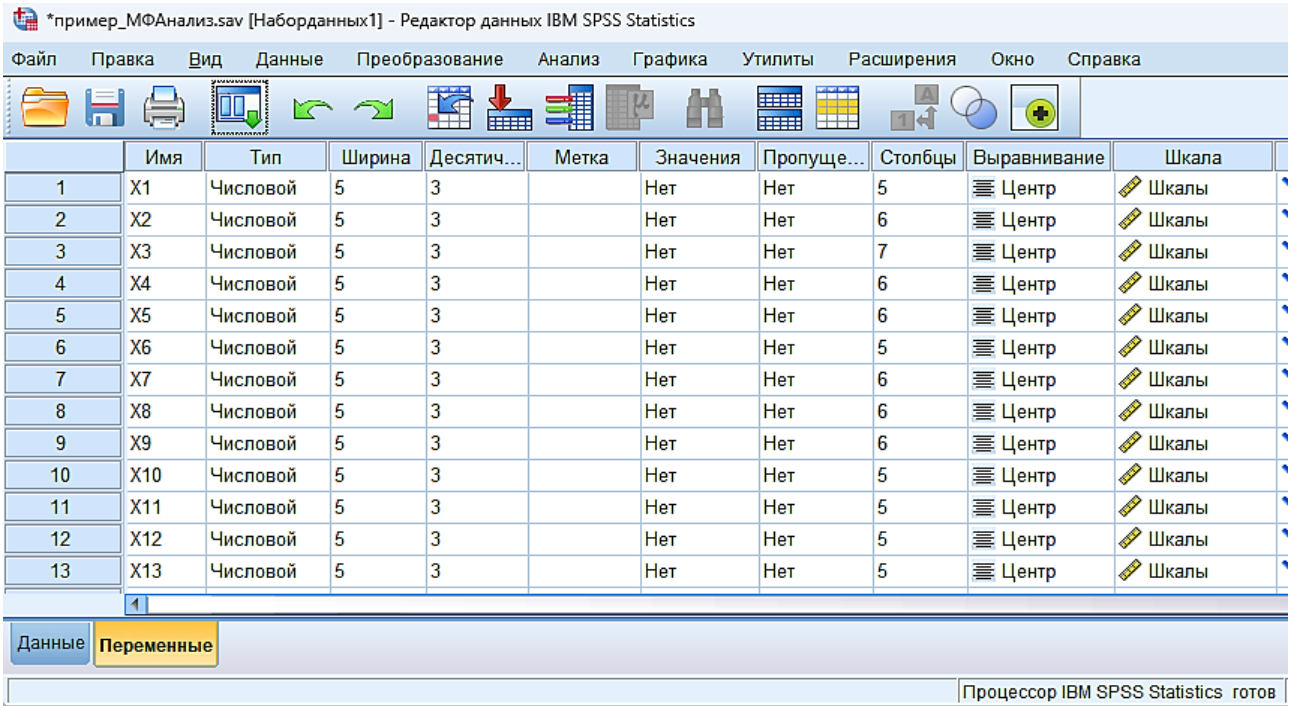


Рисунок 5.2 – Описание переменных в IBM SPSS для факторного анализа

Созданный файл сохраняется с помощью стандартного диалогового окна.

ШАГ 1.2. Для выполнения вычислений выберите в меню *Анализ – Снижение размерности (Сокращение данных) – Факторный анализ (Analyze – Data Reduction – Factor)* (рис. 5.3).

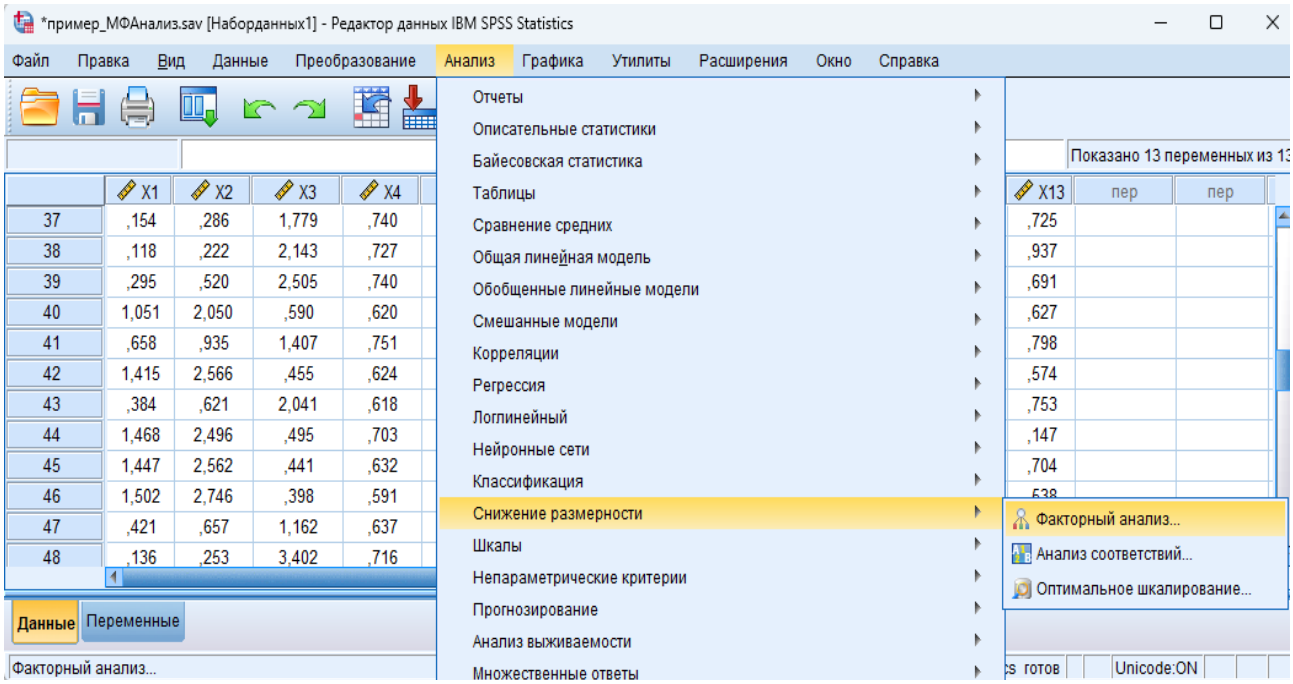
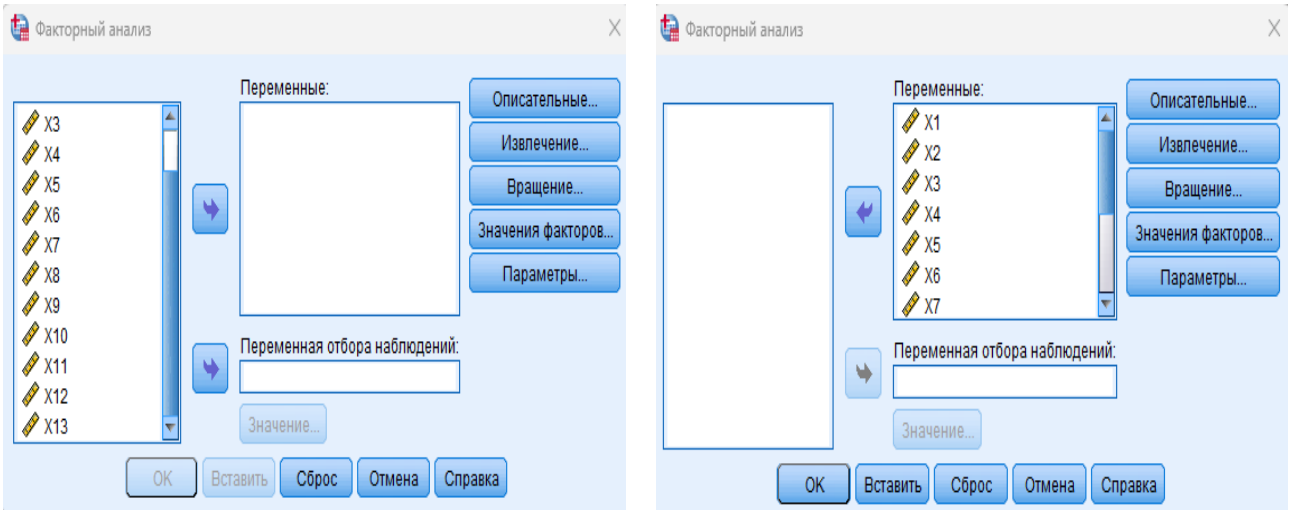


Рисунок 5.3 – Выбор факторного анализа в процедуре «Снижение размерности» в меню IBM SPSS Statistics 26

Откроется диалоговое окно *Факторный анализ* (рис. 5.4 *а*).



а *б*

Рисунок 5.4 – Диалоговое окно «Факторный анализ» в IBM SPSS Statistics

Выделите все переменные слева и перенесите их в окно справа *Переменные (Variables)* (рис. 5.4 *б*).

ШАГ 2. Щелкните по кнопке *Описательные... (Descriptive statistics...)* и оставьте по умолчанию вывод первичных результатов, включающий: первичные относительные дисперсии простых факторов, собственные значения и процентные доли объясненной дисперсии. Дополнительно можно вывести одномерные статистики *(Начальное решение (Initial solution))* и корреляционную матрицу (рис. 5.5).

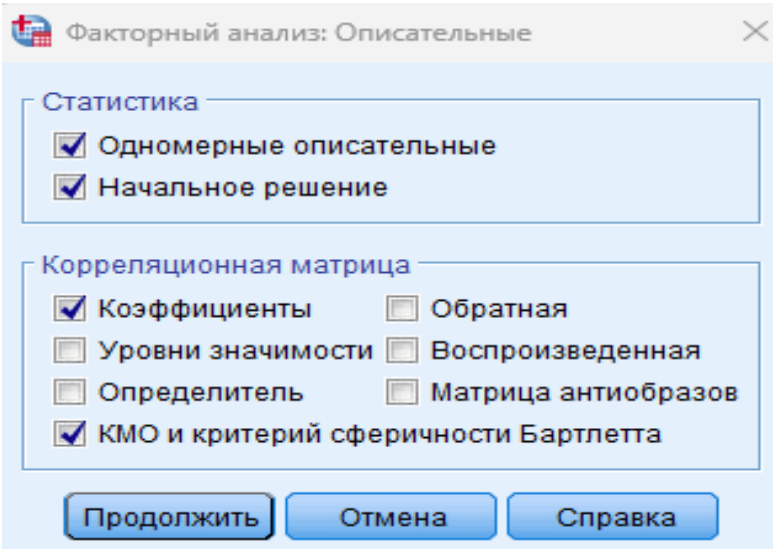


Рисунок 5.5 – Выбор вида выводимых результатов в окне «Факторный анализ: Описательные»

ШАГ 2.2. Оцените пригодность данных для проведения факторного анализа на основе теста Кайзера-Майера-Олкина (КМО) и критерия сферичности Бартлетта.

Мера адекватности выборки Кайзера-Майера-Олкина (КМО) и критерий

сферичности Бартлетта используются для проверки пригодности данных для проведения факторного анализа. Результат теста КМО, превышающий 0,5 показывает пригодность имеющихся данных для проведения факторного анализа и, соответственно, адекватность построенной факторной модели.

Критерий сферичности Бартлетта используется для подтверждения корреляционных связей между изучаемыми переменными (что является неотъемлемым условием факторного анализа, поскольку на основании корреляции переменные объединяются в факторы). Этот тест проверяет нулевую гипотезу об отсутствии корреляционных связей между переменными. Если уровень значимости меньше 0,05, то нулевую гипотезу можно отклонить. Это означает, что существует корреляция между переменными, позволяющая проводить факторный анализ.

ШАГ 3. Оценка важности каждой переменной в факторном анализе с помощью анализа общности (таблица «Общности») и анализа коэффициентов корреляции Пирсона между рассматриваемыми переменными.

ШАГ 3.1. «Начальное решение» выводит начальные общности и показывает процентные доли извлеченной (объясненной) дисперсии каждой переменной. Выводимая таблица «Общности» полезна для оценки важности той или иной переменной в факторной модели. Переменные с низкой объясненной дисперсией можно исключить из модели, тем самым повысив ее качество.

«Коэффициенты» необходимы для построения корреляционной матрицы, содержащей коэффициенты корреляции между изучаемыми переменными.

ШАГ 3.2. Исходным элементом для дальнейших расчетов является корреляционная матрица. Для построенной матрицы вычисляются собственные значения и соответствующие им собственные векторы, для определения которых используются оценочные значения диагональных элементов матрицы (относительные дисперсии простых факторов).

Собственные значения сортируются в порядке убывания, для чего чаще всего отбирается столько факторов, сколько имеется собственных значений, превосходящих по величине единицу. Собственные векторы, соответствующие этим собственным значениям, образуют факторы, а их собственные значения получают название факторных нагрузок, т. е. коэффициентов корреляции между соответствующими переменными и факторами.

ШАГ 4. Извлечение факторов для факторного анализа с помощью метода главных компонент и определение значимых факторов для анализа на основе 1-го из 3-х критериев: критерий Кайзера, критерий Кеттелла или дисперсии.

С помощью кнопки *Извлечение... (Extraction...)* (см. рис. 5.4) можно выбрать различные методы отбора факторов. Наиболее распространенным является метод главных компонент, предлагаемый по умолчанию (рис.5.6).

Количество отобранных факторов в этом случае приравнивается к числу собственных значений, превосходящих единицу. Дополнительно можно указать нужное количество факторов, вывести график собственных значений. График «каменистой осыпи» (собственных значений) используется для определения оптимального числа факторов в модели (рис. 5.7).

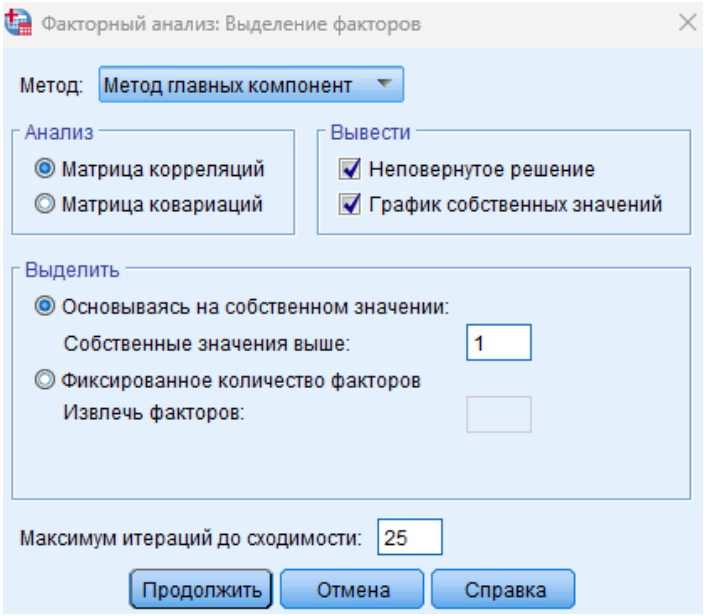


Рисунок 5.6 – Диалоговое окно «Факторный анализ: Выделение факторов»

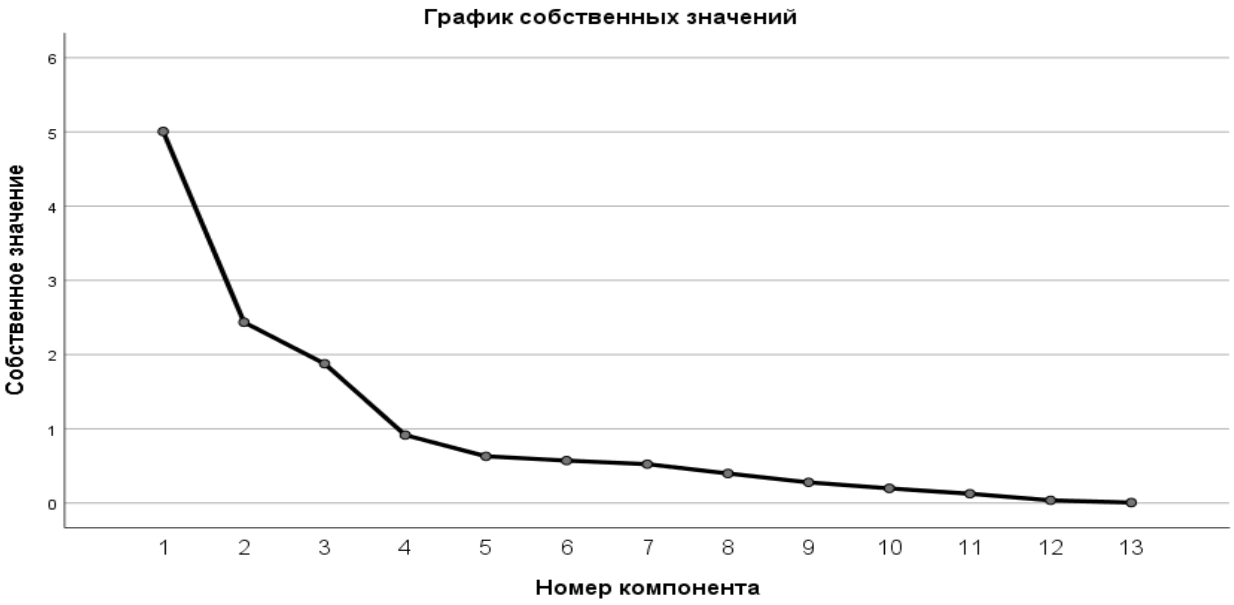


Рисунок 5.7 – График собственных значений

Окончательное решение об оптимальном числе факторов обычно принимается после смысловой интерпретации факторов, в ходе которой осуществляется неоднократное выделение различного числа факторов и сравнивается качество факторных моделей и возможности аналитической интерпретации полученных факторов.

Так как неповернутое факторное решение предоставляет малозначимую информацию, можно предотвратить его появление щелчком на соответствующей опции *Неповернутое решение (Not turned decision).*

ШАГ 5. Выполнение факторного анализа без методов вращения и с помощью методов ортогонального вращения (метод Варимакс (varimax)).

ШАГ 5.1. Кнопка *Вращение... (Rotation...)* (см. рис. 5.4) позволяет выбрать метод вращения. При вращении факторов желательно, чтобы каждый фактор имел ненулевые или значимые нагрузки (коэффициенты) только для небольшого количества переменных. Кроме того, желательно, чтобы каждая переменная имела ненулевые или значимые нагрузки с небольшим числом факторов (если можно, то с одним). Если несколько факторов имеют высокие значения факторных нагрузок с одной и той же переменной, то их трудно интерпретировать. Вращение не влияет на общности и процент объясненной полной дисперсии. Однако процент дисперсии, обусловленный влиянием отдельного фактора, меняется. Следовательно, различные методы вращения позволяют интерпретировать различные факторы.

Метод варимакс (varimax) – вращение, максимизирующее дисперсию; минимизирует число переменных с высокими значениями нагрузок, увеличивая тем самым интерпретируемость факторов; это самый распространенный метод.

Активизируйте метод варимакс (рис. 5.8) и оставьте активизированным вывод повернутой матрицы факторов.

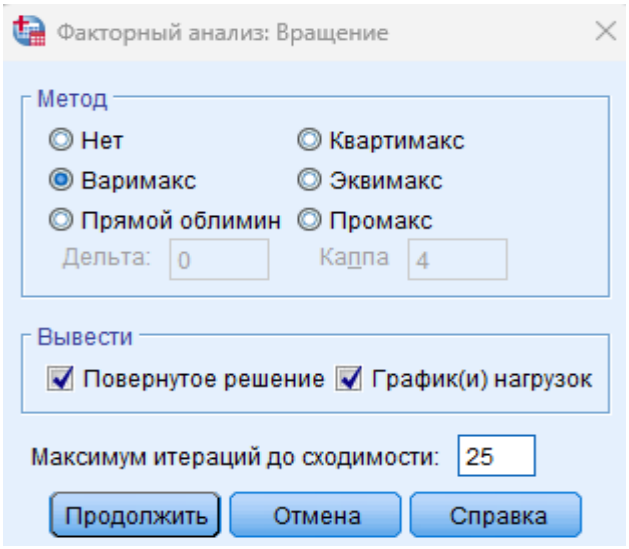


Рисунок 5.8 – Диалоговое окно «Факторный анализ: Выделение факторов»

Дополнительно можно организовать вывод факторных нагрузок в графическом виде, в котором первые три фактора будут представлены в трехмерном пространстве; при наличии двух факторов приводится изображение в одном слое.

Максимальное число шагов, которое может осуществить программа для выполнения вращения, отмечено 25, но если их будет недостаточно, то эту цифру стоит увеличить.

ШАГ 5.2. Также можно сохранить найденные значения факторов в виде дополнительных переменных. Для этого следует задействовать выключатель *Значения факторов... (Scores offactors...)* (см. рис. 5.4) и отметить *Сохранить как переменные (Save as variables)* (рис. 5.9). По умолчанию установлен регрессионный метод.

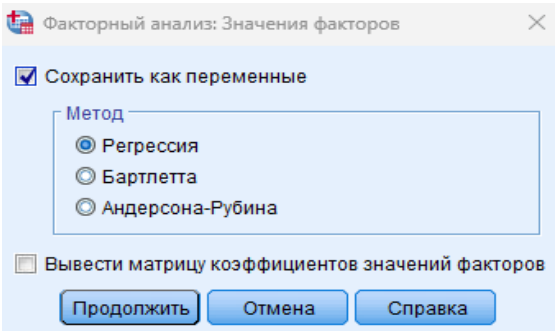


Рисунок 5.9 – Диалоговое окно «Факторный анализ: Значения факторов»

ШАГ 5.3. Выключатель *Параметры... (Options...)* (см. рис. 5.4) предназначен для обработки пропущенных значений. Здесь обеспечивается возможность заменить пропущенные значения средними значениями соответствующих переменных (рис. 5.10).

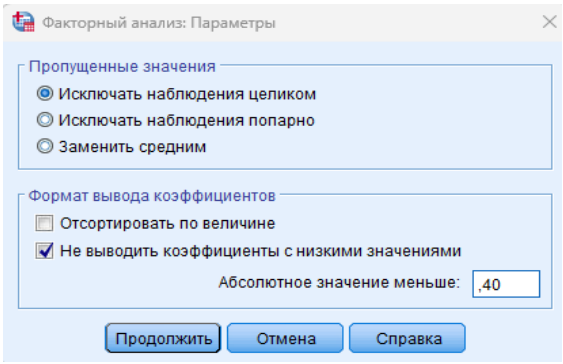


Рисунок 5.10 – Диалоговое окно «Факторный анализ: Параметры»

Для лучшей визуализации при анализе данных коэффициенты можно отсортировать по величине и не выводить те, которые по модулю меньше заданного значения (по умолчанию стоит 0,1, что приводит к выводу практически всех наблюдений, поэтому, активируя эту опцию, стоит повысить это значение, отсекая вывод тех коэффициентов, значения которых меньше обозначенного).

Для проведения расчетов щелкните *ОК* (см. рис. 5.4), в окне обзора появятся результаты факторного анализа (рис. 5.11).

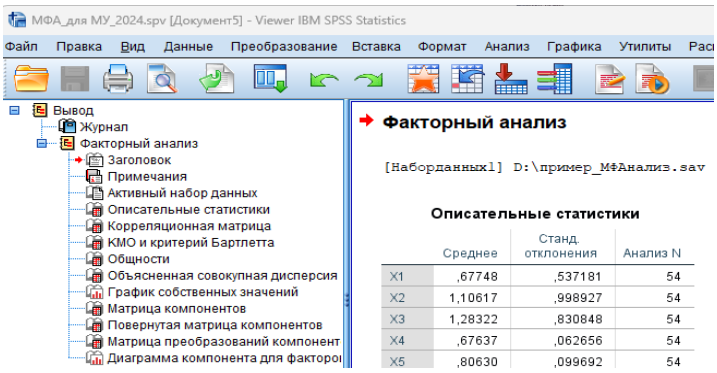
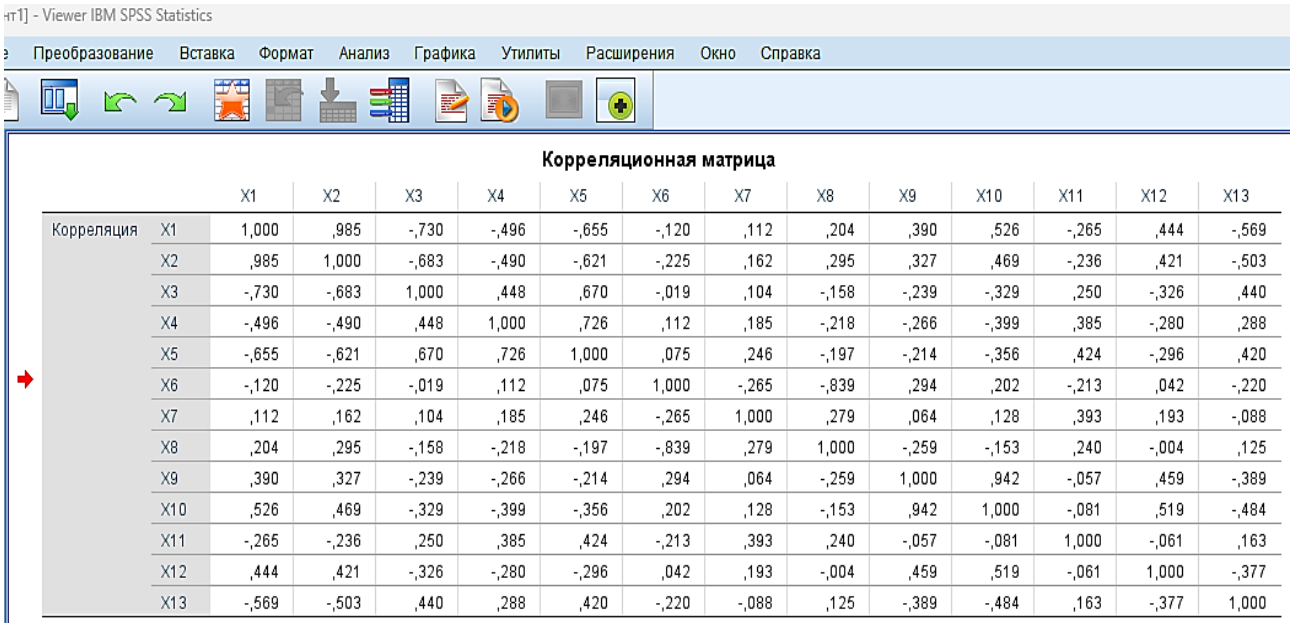


Рисунок 5.11 – Вывод результатов факторного анализа

Методические рекомендации к интерпретации результатов многомерного факторного анализа в IBM SPSS Statistics для задания 5.1

После запуска процедуры факторного анализа первыми выводятся таблицы «Описательные статистики» и «Корреляционная матрица» (рис. 5.12). В таблице «Описательные статистики» представлены среднее значение, стандартное отклонение и количество наблюдений для каждой переменной.

В таблице «Корреляционная матрица» (рис. 5.12) представлены коэффициенты корреляции при попарном сравнения переменных (максимальная корреляция 1,000 означает, что переменная сравнивается сама с собой).



Корреляционная матрица

		X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13
Корреляция	X1	1,000	,985	-,730	-,496	-,655	-,120	,112	,204	,390	,526	-,265	,444	-,569
	X2	,985	1,000	-,683	-,490	-,621	-,225	,162	,295	,327	,469	-,236	,421	-,503
	X3	-,730	-,683	1,000	,448	,670	-,019	,104	-,158	-,239	-,329	,250	-,326	,440
	X4	-,496	-,490	,448	1,000	,726	,112	,185	-,218	-,266	-,399	,385	-,280	,288
	X5	-,655	-,621	,670	,726	1,000	,075	,246	-,197	-,214	-,356	,424	-,296	,420
	X6	-,120	-,225	-,019	,112	,075	1,000	-,265	-,839	,294	,202	-,213	,042	-,220
	X7	,112	,162	,104	,185	,246	-,265	1,000	,279	,064	,128	,393	,193	-,088
	X8	,204	,295	-,158	-,218	-,197	-,839	,279	1,000	-,259	-,153	,240	-,004	,125
	X9	,390	,327	-,239	-,266	-,214	,294	,064	-,259	1,000	,942	-,057	,459	-,389
	X10	,526	,469	-,329	-,399	-,356	,202	,128	-,153	,942	1,000	-,081	,519	-,484
	X11	-,265	-,236	,250	,385	,424	-,213	,393	,240	-,057	-,081	1,000	-,061	,163
	X12	,444	,421	-,326	-,280	-,296	,042	,193	-,004	,459	,519	-,061	1,000	-,377
	X13	-,569	-,503	,440	,288	,420	-,220	-,088	,125	-,389	-,484	,163	-,377	1,000

Рисунок 5.12 – Вывод результатов корреляционного анализа в IBM SPSS

Чем выше значение корреляции между переменными, тем выше степень связи между ними, на основании которой переменные объединяются в фактор. Напомним, что корреляция может быть и обратной, о чем свидетельствует отрицательное значение коэффициента корреляции.

Особое внимание нужно обратить на таблицу с коэффициентами тестов КМО и Бартлетта (табл. 5.1).

Таблица 5.1 – КМО и критерий Бартлетта

Мера адекватности выборки Кайзера-Майера-Олкина (КМО)		0,743
Критерий сферичности Бартлетта	Примерная Хи-квадрат	619,733
	Степень свободы (ст. св.)	78
	Значимость (Знч.)	0,000

Если полученные значения этих тестов не отвечают соответствующим требованиям (КМО > 0,5 и значимость Бартлетта < 0,05), то следует попробовать исключить переменные, имеющие низкую дисперсию. Это можно оценить по таблице «Общности», в которой показываются доли

(максимальная – 1) объясненной дисперсии каждой переменной построенной факторной моделью (табл. 5.2).

Таблица 5.2 – Общности

	Начальная	Извлечение
X1	1,000	0,855
X2	1,000	0,837
X3	1,000	0,639
X4	1,000	0,599
X5	1,000	0,811
X6	1,000	0,819
X7	1,000	0,677
X8	1,000	0,865
X9	1,000	0,779
X10	1,000	0,829
X11	1,000	0,614
X12	1,000	0,491
X13	1,000	0,503

Метод выделения факторов: метод главных компонент.

Для определения оптимального количества факторов анализируют таблицу «Объясненная совокупная дисперсия» (табл. 5.3).

Таблица 5.3 – Объясненная совокупная дисперсия

Компонента	Начальные собственные значения			Ротация суммы квадратов нагрузок		
	Всего	% дисперсии	Суммарный %	Всего	% дисперсии	Суммарный %
1	5,006	38,504	38,504	3,689	28,379	28,379
2	2,436	18,740	57,244	3,261	25,083	53,462
3	1,877	14,438	71,682	2,369	18,220	71,682
4	0,915	7,040	78,722			
5	0,629	4,842	83,564			
6	0,571	4,395	87,959			
7	0,524	4,030	91,990			
8	0,398	3,061	95,050			
9	0,278	2,141	97,191			
10	0,198	1,520	98,711			
11	0,126	0,965	99,677			
12	0,036	0,277	99,953			
13	0,006	0,047	100,000			

Метод выделения факторов: метод главных компонент.

Таблица 5.3 демонстрирует долю обшей дисперсии, объясняемую каждым фактором в отдельности и всей факторной моделью в целом. Из таблицы 5.3, видно, что только три фактора имеют собственные значения (столбец «Всего») больше единицы и три отобранных фактора (компоненты 1, 2 и 3) объясняют 71,7 % совокупной дисперсии. При этом треть информации оказалась потерянной, но число переменных сократилось в 4,3 раза (с 13 до 3).

Помочь определить оптимальное количество факторов также может и «График собственных значений» (см. рис. 5.7). На графике надо найти точку резкого спада, слева (сверху) от нее расположены факторы.

Если в результате факторного анализа принято решение о сокращении количества выделенных факторов, то необходимо вернуться в процедуру выполнения факторного анализа и в диалоговом окне «Выделение факторов» (см. рис. 5.6). отметить «Фиксированное количество факторов» и указать требуемое число факторов, которое следует извлечь (например, 2, а не 3).

Для упрощения интерпретации выделенных факторов осуществляют вращение факторов, в основном, используя метод «Варимакс» (см. рис. 5.8). Сопоставляя матрицу компонентов и повернутую матрицу компонентов, можно увидеть, как изменились значения переменных.

Покажем в таблицах изменения компонент, которые происходят с переменными до вращения факторов (табл. 5.4) и после вращения факторов (табл. 5.5), выделив три главных фактора. Сравнивая эти таблицы, видно, что значения факторных нагрузок изменилось.

Таблица 5.4 – Матрица факторных нагрузок (до вращения факторов)

	Компонента		
	1	2	3
X1	0,903		
X2	0,862		
X3	-0,760		
X4	-0,683		
X5	-0,783		0,421
X6		0,902	
X7			0,751
X8		-0,922	
X9	0,591	0,477	0,450
X10	0,719		0,431
X11			0,647
X12	0,587		
X13	-0,663		

Метод выделения факторов: метод главных компонент.
а. Извлечено компонентов - 3.

Таблица 5.5 – Матрица повернутых нагрузок (после вращения факторов)

	Компонента		
	1	2	3
X1	**-0,706**	0,543	
X2	**-0,688**	0,492	
X3	**0,733**		
X4	**0,754**		
X5	**0,884**		
X6			**-0,864**
X7		0,415	**0,590**
X8			**0,884**
X9		**0,846**	
X10		**0,878**	
X11	0,622		0,459
X12		0,668	
X13		-0,596	

Метод выделения факторов: метод главных компонент.
Метод вращения: варимакс с нормализацией Кайзера.
а. Вращение сошлось за 11 итераций.

Чем больше факторная нагрузка, тем сильнее корреляция между переменной и фактором, и, соответственно, тем значительнее вклад данной переменной в фактор. В этом и состоит цель вращения – упростить факторную структуру, в которой бы определенные переменные оказывали большую нагрузку на один фактор и малую на все остальные.

После вращения появляется повернутая матрица компонентов (табл. 5.5), в которой отражаются коэффициенты корреляции между переменными и компонентами (факторами), на основании которых выявляются переменные, входящие в тот или иной фактор. В одну группу объединяются переменные, имеющие наиболее сильную связь с компонентом. Именно эта таблица необходима для определения количества переменных, входящих в конкретный

фактор, и смысловой интерпретации образованных факторов.

Для упрощения анализа можно не выводить коэффициенты с низкими значениями (например, 0,3 или 0,4) и отсортировать переменные по величине коэффициентов. Это делается в диалоговом окне «Параметры» (см. рис. 5.10).

Замечание. Если переменная даже в результате вращения обнаруживает значительные связи с несколькими факторами, то вопрос ее отнесения к какому-либо фактору остается в компетенции исследователя, который принимает решение на основании собственных представлений о модели связи между изучаемыми переменными (признаками). В некоторых ситуациях приходится отказаться от проведения факторного анализа и попытаться применить другие статистические процедуры.

ШАГ 6. Интерпретация факторов. Интерпретация фактора предполагает содержательное осмысление переменных, показывающих большие нагрузки с данным фактором. В нашем случае были выявлены три главных фактора на основе полученных результатов матрицы повернутых нагрузок (табл. 5.5): первый фактор – эффективность использования основных средств (или техническое состояние основных средств); второй фактор – движение и обновление основных средств; третий фактор – структура основных средств.

В результате выполнения интерпретации главных факторов можно столкнуться с трудностями смыслового содержания полученных факторов. Такая проблема возникала и в задании 5.1. Одним из ее решений является сокращение количества факторов. Однако это приводит к уменьшению доли объясненной дисперсии. Вторым решением является обращение к таблице «Общности» (табл. 5.2) и исключить из факторного анализа те переменные, которые извлекают небольшую долю дисперсии. Решение об этом принимается на основе содержательного факторного анализа и возможности смысловой интерпретации. Сократив количество переменных, надо вновь выполнить ряд шагов факторного анализа, обращая внимание на критерии КМО и Бартлетта.

На последнем этапе факторного анализа созданным главным факторам надо снова присвоить название, исходя из смыслового содержания входящих в них переменных, и интерпретировать их. В дальнейшем с этими факторами можно работать как с переменными, используя их в различных видах статистического анализа. Сохранив факторы в качестве новых переменных в исходной базе данных, им надо присвоить названия. Значения новых переменных (факторов) содержатся в окне «Представление данных» (рис. 5.13).

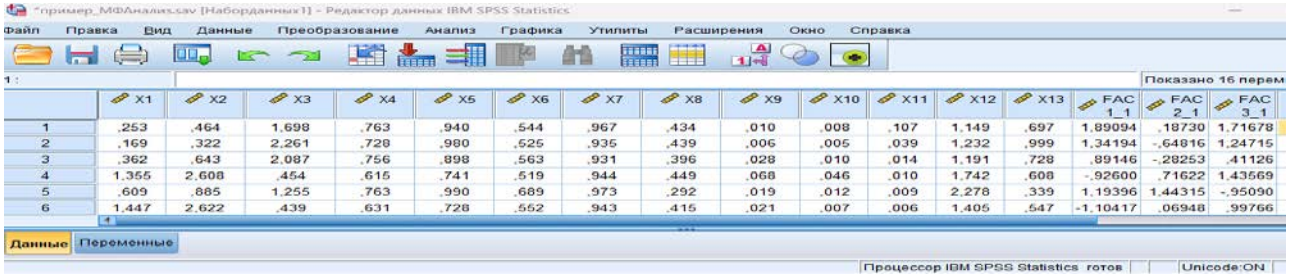


Рисунок 5.13 – Фрагмент диалогового окна «Представление данных» в SPSS

Задания для самостоятельного решения по темам 5–7

Задание 7.1. Имеются следующие данные по 30 коммерческим организациям: уровень средней месячной заработной платы рабочих (X1), стаж работы (X2) и средняя производительность труда (X3) представлены в табл. 7.1.

Таблица 7.1 – Исходные данные организаций для статистического анализа

№ организации	X1	X2	X3	№ организации	X1	X2	X3
1	2360	5	21	16	2400	8	37
2	3000	10	40	17	2800	10	33
3	2480	6	31	18	2650	15	28
4	2140	8	28	19	3500	4	45
5	2800	2	30	20	2950	12	29
6	1650	4	15	21	2700	10	38
7	1800	2	17	22	1800	5	24
8	3150	20	45	23	2500	7	28
9	2000	7	20	24	3100	12	34
10	2740	11	37	25	3450	15	38
11	1940	5	25	26	2200	6	26
12	2670	18	32	27	1800	3	18
13	2800	10	45	28	1750	3	20
14	3200	12	50	29	1900	6	21
15	3800	18	55	30	2150	3	29

Выполните факторный анализ на основании приведенных переменных (признаков) с использованием пакетов IBM SPSS или Statistica. Составьте отчет в MS Word и поясните полученные результаты.

Задание 7.2. Для анализа социально-экономического положения всех районов Витебской области отберите из статистических сборников основные показатели (индексы): промышленного производства (X1), строительство жилых домов (X2), продукция сельского хозяйства (X3), оборот розничной торговли (X4), объем платных услуг населению (X5), инвестиции в основной капитал (X6), иностранные инвестиции (X7) (табл. 7.2).

Таблица 7.2 – Основные показатели социально-экономического положения районов Витебской области в 20** г.

№ п/п	Район	X1	X2	X3	X4	X5	X6	X7
1								
…								
n								

Требуется: выделить и интерпретировать главные факторы по месторасположению и по экономическим показателям отдельных районов; выполнить ранжирование районов по выделенным переменным (признакам). Выполните технологию решения многомерного факторного анализа данных с помощью IBM SPSS Statistics.

Методические указания к выполнению задания 7.2

Выполните задание в IBM SPSS с помощью команды *Анализ данных*. Загрузите (или скопируйте) в IBM SPSS данные из таблицы 7.2 для выполнения факторного анализа.

ШАГ 1. Первая задача – выявить скрытую структуру по расположению районов (т. е. по географическому фактору). Для этого сначала необходимо перегруппировать исходные данные на рабочем листе. Воспользуемся инструментом *Транспонирование (Transposing)* (в меню *Данные*) для транспонирования исходных переменных. Затем запустите надстройку *Анализ – Снижение размерности – Факторный анализ (Analyze – Data Reduction – Factor),* совершая те же действия, что и в задании 5.1.

ШАГ 2. Вторая задача анализа – выявить факторные признаки среди экономических показателей. Возьмите данные из условия задания (без транспонирования). Запустите факторный анализ и проанализируйте полученные результаты.

Задание 7.3. Для анализа о среднедушевом потреблении товаров и услуг в некоторых странах отберите из статистических сборников следующие показатели: продукты (X1), квартплата (X2), обустройство дома (X3), транспорт (X4), лечение (X5), обучение (X6), одежда (X7) прочее (X8) (табл. 7.3).

Таблица 7.3 – Среднедушевое потребление товаров и услуг в 20** г.

Страна	X1	X2	X3	X4	X5	X6	X7	X8
Беларусь								
Эстония								
Украина								
Казахстан								
Таджикистан								
США								
Япония								
Германия								
Турция								
Россия								

Требуется: выделить и интерпретировать главные факторы. Выполните технологию решения многомерного факторного анализа данных с помощью IBM SPSS Statistics.

Методические указания к выполнению задания 7.3

Выполните задание в IBM SPSS с помощью команды *Анализ данных*. Загрузите (или скопируйте) в IBM SPSS данные из таблицы 7.3 для выполнения факторного анализа.

ШАГ 1. Сначала рассмотрите потребление по странам. Для этого необходимо перегруппировать исходные данные на рабочем листе.

Воспользуйтесь инструментом *Транспонирование (Transposing)* (в меню *Данные*) для транспонирования исходных переменных. Затем запустите надстройку *Анализ – Снижение размерности – Факторный анализ (Analyze – Data Reduction – Factor),* совершая те же действия, что и в задании 5.1.

ШАГ 2. Определите скрытые факторы в структуре расходов (данные до транспонирования). Запустите факторный анализ в пакете IBM SPSS и проанализируйте полученные результаты. Дайте экономическую интерпретацию главным факторам.

Задание 7.4. Требуется выявить скрытые факторы, влияющие на характеристики вкладчиков коммерческого банка (табл. 7.4).

Таблица 7.4 – Характеристика вкладчиков банка

№ п/п	Пол	Образование	Вклад, тыс. ед.	Возраст
1	Ж	Высшее	1350	35
2	М	Высшее	1300	30
3	Ж	Незаконченное высшее	2000	21
4	М	Среднее	1250	42
5	М	Среднее	1200	20
6	Ж	Высшее	1400	38
7	Ж	Незаконченное высшее	1350	34
8	М	Среднее	1800	19
9	М	Высшее	4000	54
10	Ж	Среднее	2800	48
11	М	Высшее	1250	63
12	Ж	Высшее	1350	69
13	Ж	Незаконченное высшее	2600	53
14	М	Среднее	1350	28
15	М	Среднее	1200	42
16	Ж	Высшее	1750	55
17	Ж	Незаконченное высшее	1350	51
18	М	Среднее	1900	33
19	М	Высшее	3500	53
20	Ж	Среднее	3000	69

Методические указания к выполнению задания 7.4

Выполните задание в IBM SPSS с помощью команды *Анализ данных*. Загрузите (или скопируйте) в IBM SPSS данные из таблицы 7.4 для выполнения факторного анализа.

ШАГ 1. Сначала с помощью встроенных инструментов SPSS перекодируйте нечисловую информацию в числовую. Для этого воспользуйтесь функцией меню *Преобразовать – Перекодировать – В те же переменные... (Recode into Same Variables)*. Исходные данные (пол: женский (1), мужской (-1); образование: среднее (1), незаконченное высшее (2) и т. д.).

ШАГ 2. Запустите факторный анализ в IBM SPSS и выполните интерпретацию выделенных главных факторов.

ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 8
«Кластерный анализ многомерных наблюдений с использованием статистических программных пакетов»

Цель занятия: получить навыки выполнения многомерного кластерного анализа данных с применением статистических пакетов SPSS и STATISTICA.

Задание 8.1. В таблице 8.1 приведены индексы цен и тарифов в секторах экономики по районам области за 2022 год: индекс потребительских цен (X1), индекс цен производителей промышленной продукции (X2), индекс цен производителей сельскохозяйственной продукции (X3), индекс цен на строительно-монтажные работы (X4), индекс тарифов на перевозку грузов (X5).

Таблица 8.1 – Индексы цен и тарифов в секторах экономики в %

Районы	X1	X2	X3	X4	X5
А	112,8	89,4	100,8	107,1	105,2
Б	113,8	111,6	105,6	109,2	105,0
В	113,2	104,3	117,4	107,8	114,2
Г	112,3	107,5	101,7	105,7	121,9
Д	112,7	110,3	102,9	111,9	112,7
Е	115,6	114,0	120,7	107,3	129,0
Ж	114,7	106,4	97,0	112,3	102,1
З	113,3	102,4	100,8	112,7	116,2
И	113,0	89,7	106,5	107,4	108,2
К	114,1	111,8	109,6	111,6	105,1
Л	112,9	111,8	105,9	109,0	108,1
М	112,1	105,8	109,9	104,6	107,6
Н	110,3	105,7	98,3	108,6	97,9
О	112,9	104,4	99,0	106,5	107,6
П	112,3	106,5	97,8	107,8	116,6
Р	109,7	106,2	107,7	108,6	105,3
С	111,6	107,9	104,7	103,6	118,3

Выполните пошаговый алгоритм многомерного кластерного анализа с помощью программы IBM SPSS Statistics.

1. Выполните подготовительные шаги в программе IBM SPSS Statistics для кластерного анализа экономических данных.
2. Определите наиболее подходящее количество кластеров для разбиения районов на группы по всем пяти признакам.
3. Выполните классификацию районов области иерархическим методом.
4. Дайте графическое представление процесса объединения районов в виде дендрограммы.
5. Для каждого из полученных кластеров дайте содержательную интерпретацию, определив кластерные профили.

Методические указания к выполнению задания 8.1 в IBM SPSS Statistics

При проведении многомерного кластерного анализа сначала выполните ряд подготовительных шагов. Эти шаги (шаги 1.1 – 1.2) позволяют подготовить рабочий файл данных, запустить программу IBM SPSS Statistics 26 (версия 26 программы или др.) и открыть файл с исходными данными для анализа.

ШАГ 1. Программы кластерного анализа являются составной частью многих пакетов статистической обработки данных, в том числе и SPSS. В части кластеризации программа SPSS дает возможность проведения двухэтапного кластерного анализа, иерархической кластеризации и кластеризации *k*-средними. В строке меню Анализ выберите команды *Анализ – Классификация – Иерархическая кластеризация (Analyze – Classify – Hierarchical Cluster Analysis)* (рис. 8.1).

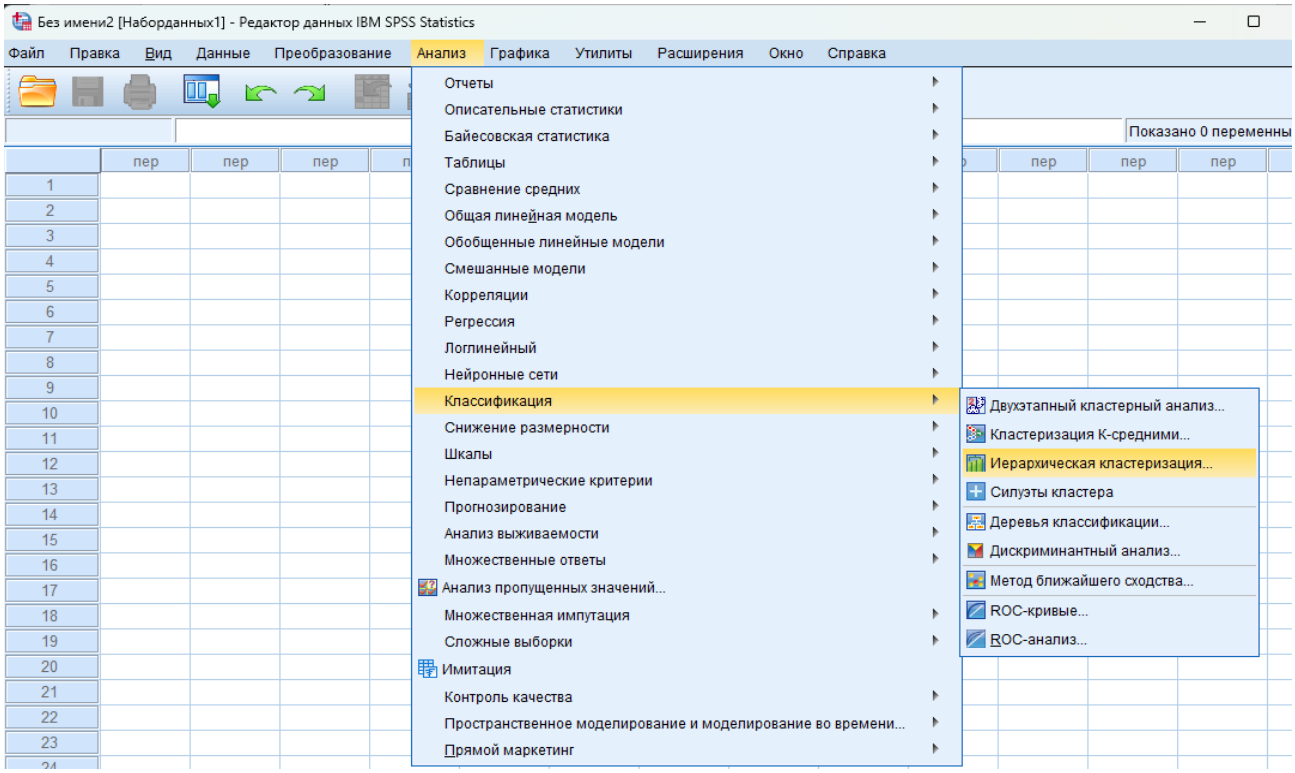


Рисунок 8.1 – Выбор иерархической кластеризации в процедуре «Классификация» в меню «Анализ» IBM SPSS Statistics 26

ШАГ 1.1. Подготовка данных для использования пакета IBM SPSS Statistics. Данные могут быть введены непосредственно в SPSS или импортированы из текстовых файлов, файлов формата Excel, ACCESS и т. п. В нашем случае имеется файл данных в формате Excel. Выполните импорт данных в файл IBM SPSS. Для импорта данных из файла формата Excel в IBM SPSS Statistics выполните подготовительные действия в соответствии с методическими указаниями к заданию 4.5.

ШАГ 1.2. В редакторе данных создайте таблицу, в которой одна может быть текстовой переменной и пять числовых (рис. 8.2).

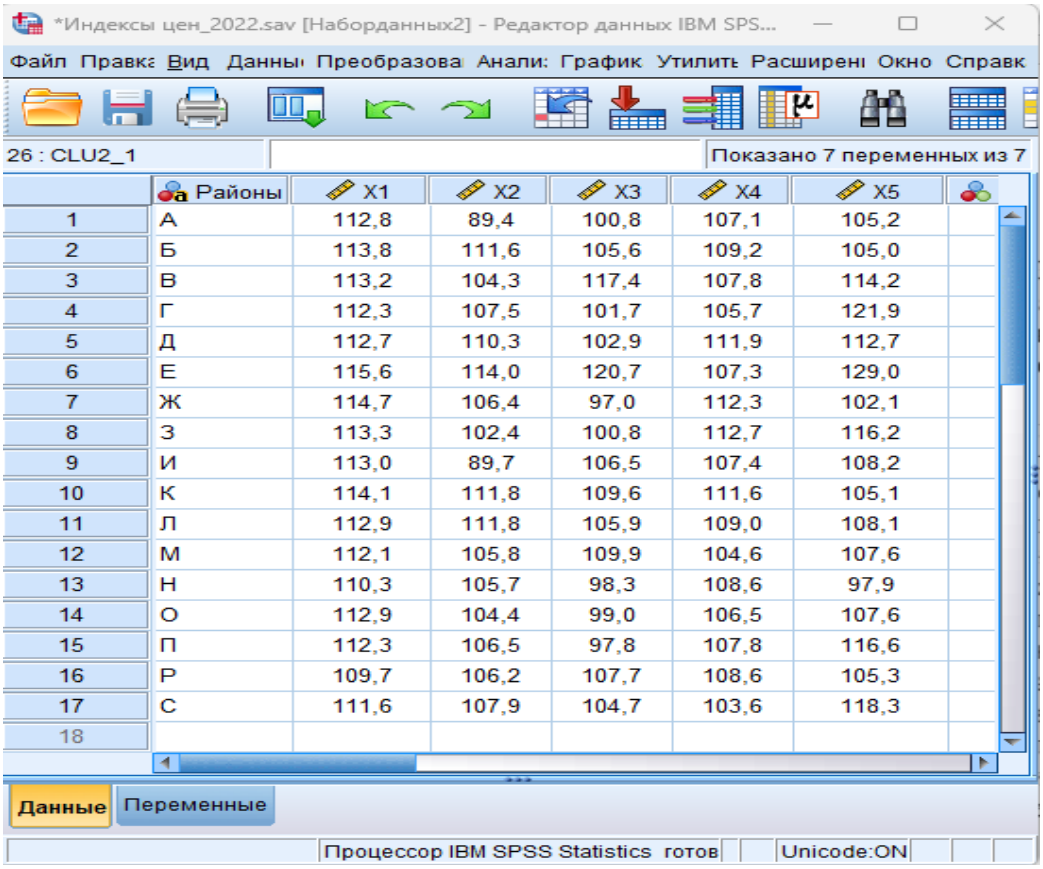
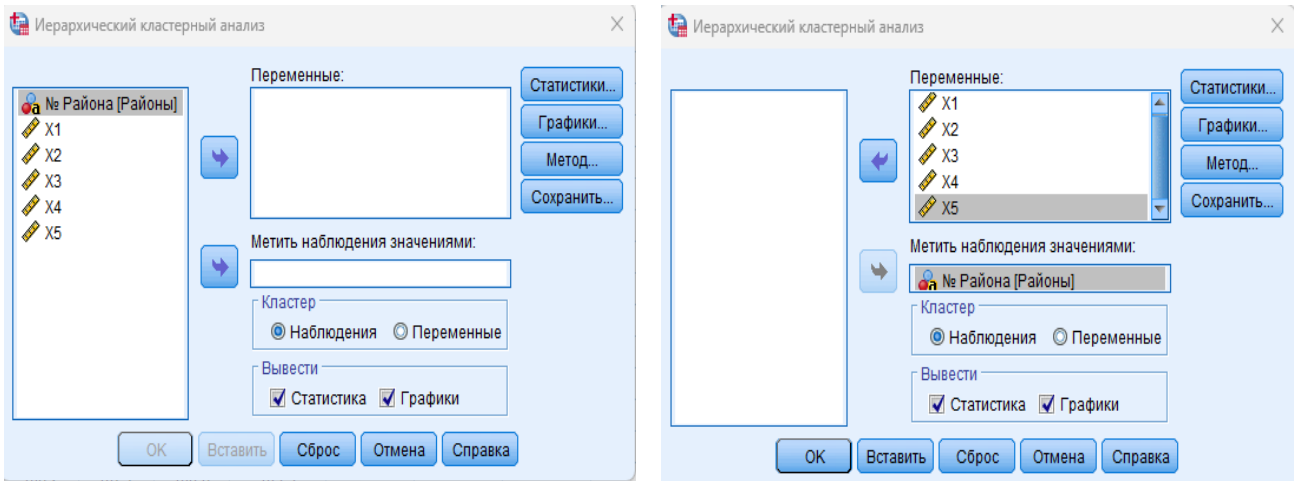


Рисунок 8.2 – Исходные данные для кластерного анализа в IBM SPSS

ШАГ 2 Определим наиболее подходящее количество кластеров для разбиения районов на группы по всем пяти признакам.

Для начала выполните предварительную классификацию, которая позволит определить наиболее целесообразное количество кластеров. Для этого выберите в меню *Анализ – Классификация – Иерархическая кластеризация... (Analyze – Classify – Hierarchical Cluster Analysis...).*

Появится диалоговое окно *Иерархический кластерный анализ* (рис. 8.3 а).



а *б*

Рисунок 8.3 – Диалоговое окно «Иерархический кластерный анализ» в SPSS

Поместите числовые переменные, соответствующие индексам цен, в поле тестируемых переменных, а текстовую переменную «Районы» будем использовать для обозначения (маркировки) наблюдений (рис. 8.3 б).

Для начала должно быть достаточно только вывода обзорной таблицы порядка агломерации. Для этого в окне *Иерархический кластерный анализ (Hierarchical Cluster Analysis)* активизируем кнопку *Статистики... (Statistics...)*. Откроется диалоговое окно *Иерархический кластерный анализ: Статистики (Hierarchical Cluster Analysis: Statistics)*, в котором активизируйте вывод последовательности слияния, опция *Порядок агломерации (Agglomeration schedule)* (рис. 8.4).

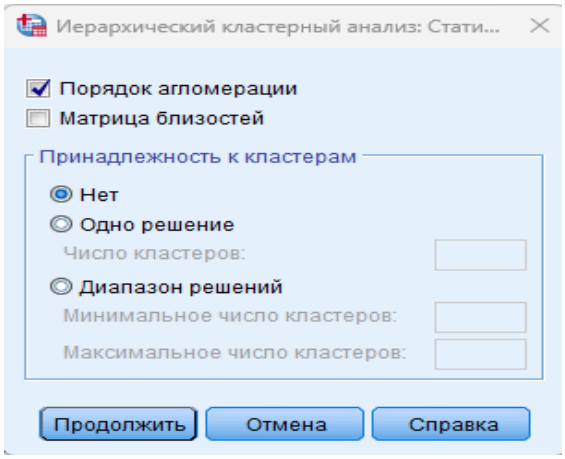


Рисунок 8.4 – Диалоговое окно «Иерархический кластерный анализ: Статистики» в IBM SPSS

В результате в окне просмотра (табл. 8.2) будет приведен обзор принадлежности, из которого можно определить очередность построения кластеров, а также их оптимальное количество.

Таблица 8.2 – Порядок агломерации (шаги кластеров)

Этап	Объединенный кластер		Коэффициенты	Этап первого появления кластера		Следующий этап
	Кластер 1	Кластер 2		Кластер 1	Кластер 2	
1	2	11	10,590	0	0	2
2	2	10	26,395	1	0	8
3	4	17	27,020	0	0	11
4	12	16	32,050	0	0	8
5	1	9	41,710	0	0	15
6	8	15	50,980	0	0	9
7	7	13	52,870	0	0	10
8	2	12	69,595	2	4	12
9	5	8	76,350	0	6	11
10	7	14	91,285	7	0	12
11	4	5	101,193	3	9	13
12	2	7	169,022	8	10	13
13	2	4	243,965	12	11	14
14	2	3	318,532	13	0	15
15	1	2	430,794	5	14	16
16	1	6	**862,758**	15	0	0

Число строк в таблице 8.2 всегда на единицу меньше, чем участвующих в кластерном анализе объектов (наблюдений). Здесь действует правило $n - 1$. Поэтому в рассматриваемом задании 17 районов, но 16 строк.

Для определения того, какое количество кластеров следовало бы считать оптимальным, решающее значение имеет показатель, выводимый под заголовком *«Коэффициент» (Coefficient)* (табл. 8.2). Под этим коэффициентом подразумевается расстояние между двумя кластерами, определенное на основании выбранной дистанционной меры (в нашем случае это квадрат евклидова расстояния). На том этапе, когда мера расстояния между двумя кластерами увеличивается скачкообразно, процесс объединения в новые кластеры необходимо остановить, так как в противном случае были бы объединены уже кластеры, находящиеся на относительно большом расстоянии друг от друга. В нашем задании оптимальным считается число кластеров, равное разности количества наблюдений (17) и номера шага (15), после которого коэффициент увеличивается скачкообразно. Таким образом, в нашем задании оптимальное количество кластеров равно 2 (17 – 15). Вторым способом является сложение числа строчек снизу таблицы, начиная с той, на которой произошел скачок, то есть тоже два кластера, так как две строчки (15 и 16) (см. табл. 8.2).

Теперь организуйте для каждого наблюдения вывод информации о принадлежности к кластеру. Для этого вновь откройте диалоговое окно *Иерархический кластерный анализ (Hierarchical Cluster Analysis)* и щелкните по выключателю *Статистики... (Statistics...)* (см. рис. 8.4). В разделе *Принадлежность к кластерам (Clusters Membership)* активизируем опцию *Одно решение (Single solution)* и укажем желаемое количество кластеров 2 (рис. 8.5).

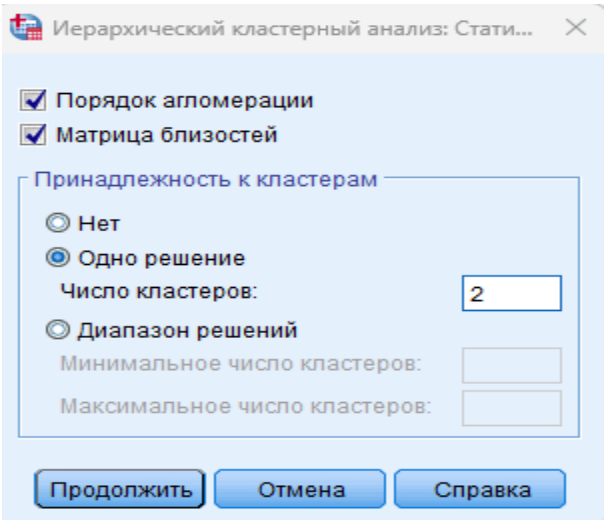


Рисунок 8.5 – Заполнение опций «Матрица близостей» и «Принадлежность к кластерам» в диалоговом окне «Иерархический кластерный анализ: Статистики»

В этом же окне (рис. 8.5) можно заказать вывод матрицы расстояний между объектами исследования, опция *Матрица близостей (Proximity matrix).*

Замечание: программа дает возможность получения диапазона кластеров, который задается с помощью опции *Диапазон решений (Range of solutions).* Так, если ввести в поле «Минимальное число» кластеров 2, а в поле «Максимальное число» кластеров число 4, то в выводимые результаты будут включены все решения с 2, 3 и 4 кластерами.

Информацию о принадлежности каждого наблюдения к определенному кластеру можно сохранить в новой переменной. Используя выключатель главного диалогового окна *Сохранить... (Save...)*, активизируйте опцию *Одно решение (Single solution)* и для указания желаемого количества кластеров введите 2 (рис. 8.6).

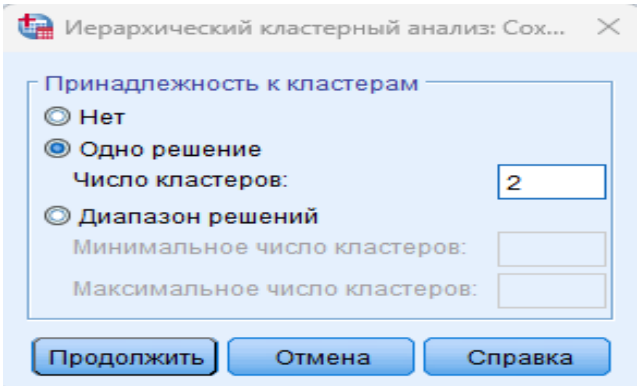


Рисунок 8.6 – Диалоговое окно «Иерархический кластерный анализ: Сохранить» в IBM SPSS

Теперь помимо таблицы порядка агломерации для каждого наблюдения будет выводиться также информация о принадлежности к кластеру.

ШАГ 3 Выполним классификацию районов области иерархическим методом. Вернувшись в главное диалоговое окно, щелкнем по выключателю *Графики... (Plots...)* (см. рис. 8.3 б). Активизируем оппию вывода древовидной диаграммы *Дендрограмма (Dendrogram)* и посредством опции *Не надо* или *нет (None)* отмените вывод накопительной диаграммы (рис. 8.7).

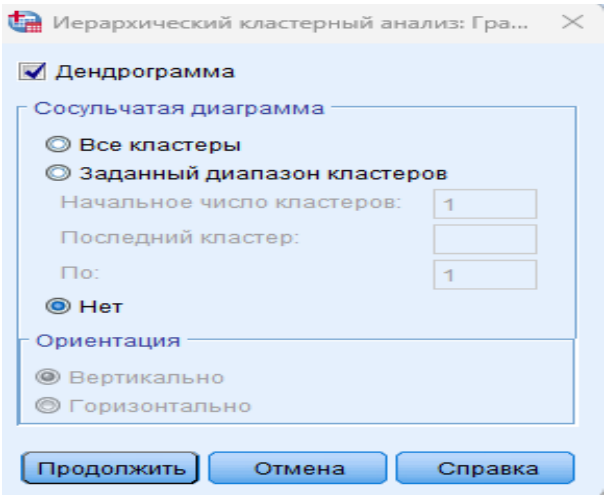
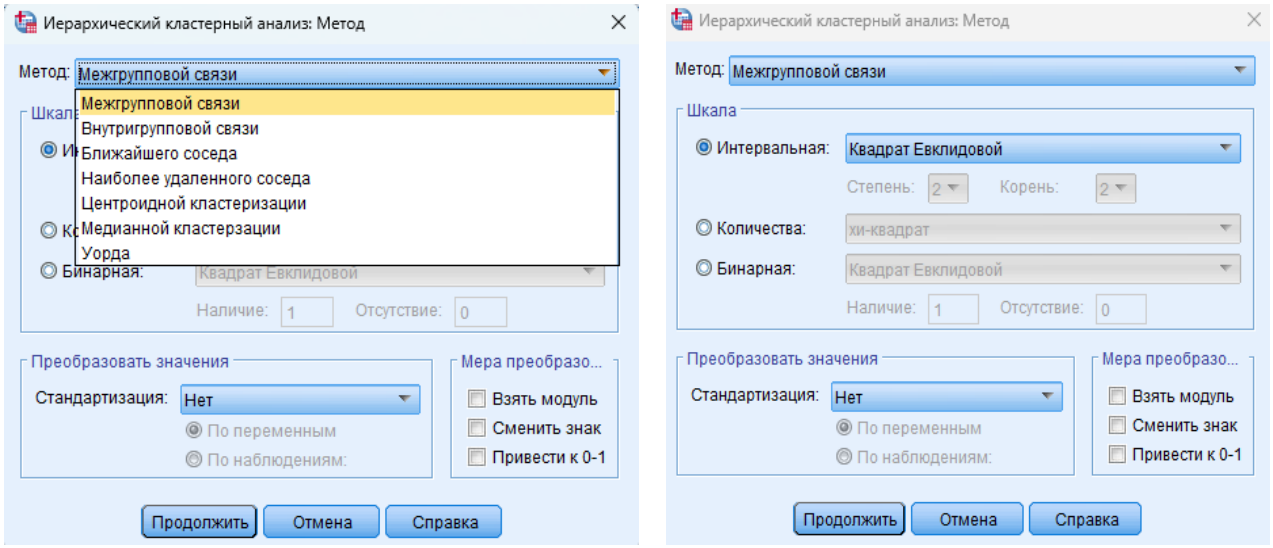


Рисунок 8.7 – Диалоговое окно «Иерархический кластерный анализ: Графики»

С помощью кнопки *Метод... (Method...)* (см. рис. 8.3 б) получим возможность выбрать метод образования кластеров, а также метод расчета дистанционной меры и меры подобия. Программа SPSS предлагает семь различных методов объединения (рис. 8.8 а). Метод *Межгрупповые связи (Between-groups linkage)* устанавливается по умолчанию (рис. 8.8 б).



а *б*

Рисунок 8.8 – Диалоговое окно «Иерархический кластерный анализ: Метод»

Дистанционные меры (шкалы) и меры подобия зависят от вида переменных, участвующих в анализе (рис. 8.9).

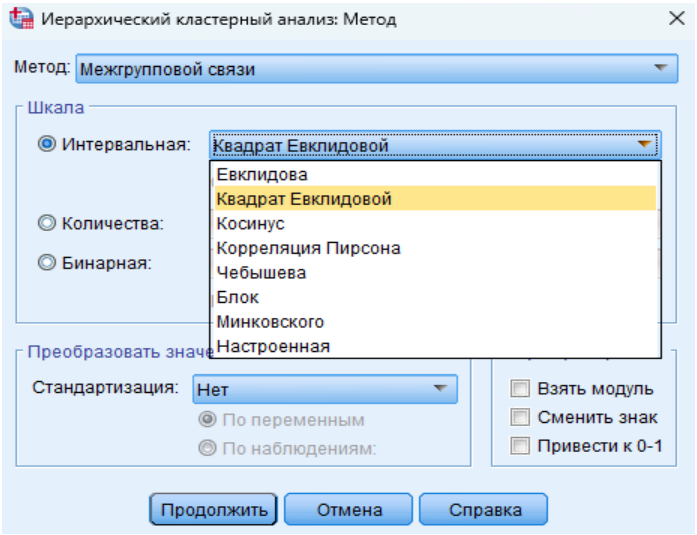


Рисунок 8.9 – «Шкала» в диалоговом окне «Иерархический кластерный анализ: Метод»

По умолчанию в качестве дистанционной меры (шкалы) для интервальных переменных устанавливается *Квадрат расстояния Евклида (Squared Euclidean distance)* (рис. 8.8 б).

Программа IBM SPSS дает возможность стандартизировать значения,

если это необходимо. Соответствующие установки можно сделать в поле Стандартизация (Transform Values) (рис. 8.8). На практике чаще всего применяют z-стандартизацию.

ШАГ 4 Выполним графическое представление процесса объединения районов в виде дендрограммы.

Дендрограмма (рис. 8.10) визуализирует процесс слияния, приведенный в обзорной таблице порядка агломерации (см. табл. 8.2). Она идентифицирует объединенные кластеры и значения коэффициентов на каждом шаге. При этом отображаются не исходные значения коэффициентов, а значения, приведенные к шкале от 0 до 25 (числа от 0 до 25 являются условной шкалой этих расстояний; 0 соответствует наименьшему расстоянию на первом этапе, а 25 – наибольшему расстоянию на последнем этапе). Кластеры, полученные в результате слияния, отображаются горизонтальными линиями.

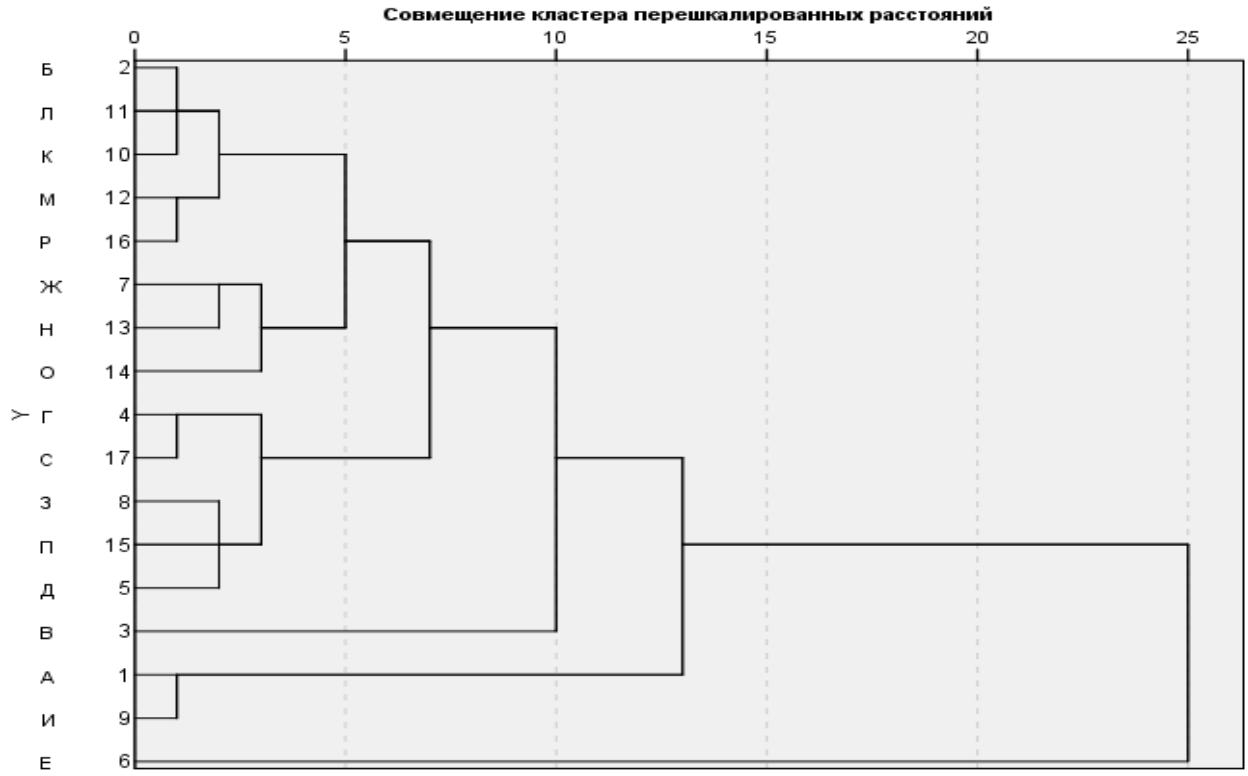


Рисунок 8.10 – Дендрограмма кластеризации, полученная с использованием метода межгрупповых связей

На дендрограмме (рис. 8.10) любое решение характеризуется вертикальной линией, число точек пересечения которой с деревом соответствует количеству кластеров на текущем этапе. Так, для нашего задания эту линию следует расположить на уровне 10–15: между шагами 15 и 16 кластеризации. В этом случае получается два кластера. Для того, чтобы установить состав каждого кластера, необходимо вернуться к корням дерева и выяснить соответствующие номера объектов (наблюдений). Напомним, что в результате выполнения шага 2 (рис. 8.6) в файле «Индексы цен_2022.sav»

появилась новая переменная, определяющая принадлежность каждого объекта (наблюдения) к одному из двух кластеров (рис. 8.11).

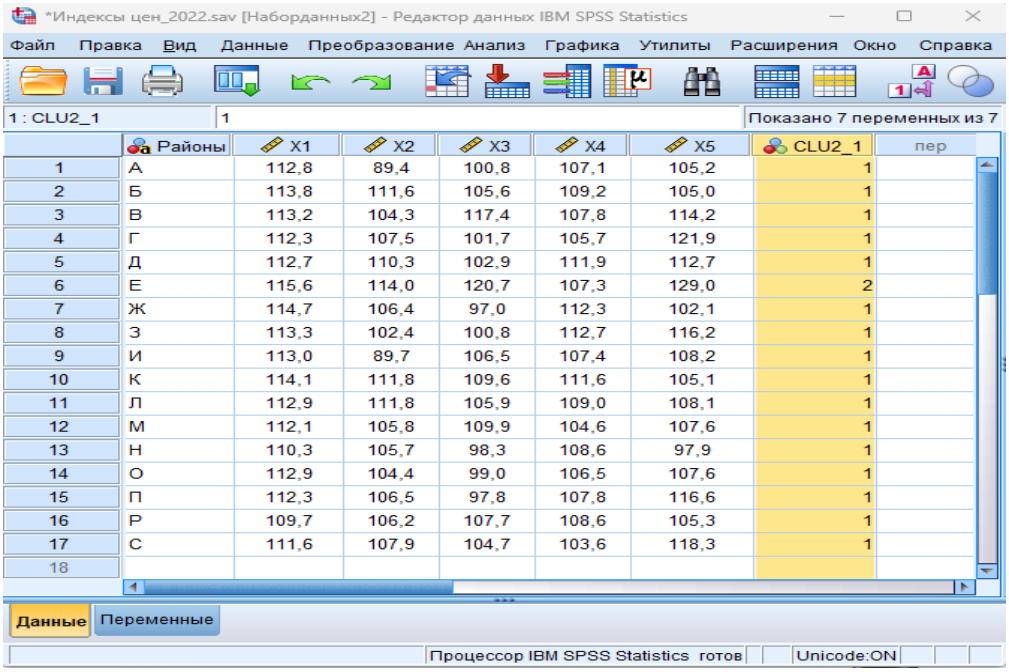


Рисунок 8.11 – Исходные данные вместе с переменной кластерной принадлежности clu2_1, полученной в результате выполнения иерархического кластерного анализа в IBM SPSS

Из рис. 8.11 видно, что в первый кластер попали 16 районов анализируемой области (этот кластер получился самым большим), во второй – 1 район (район Е). Аналогичные результаты анализа представлены в таблице «Принадлежность к кластерам» (табл. 8.3).

Таблица 8.3 – Принадлежность к кластерам

Наблюдение	Кластеры 2
1:А	1
2:Б	1
3:В	1
4:Г	1
5:Д	1
6:Е	2
7:Ж	1
8:З	1
9:И	1
10:К	1
11:Л	1
12:М	1
13:Н	1
14:О	1
15:П	1
16:Р	1
17:С	1

Матрица близостей (рис. 8.12) позволяет увидеть, например, что наиболее близкими характеристиками (индексами цен) обладают в первом кластере следующие районы: Б и К, Б и Л, Г и С, а также М и Р.

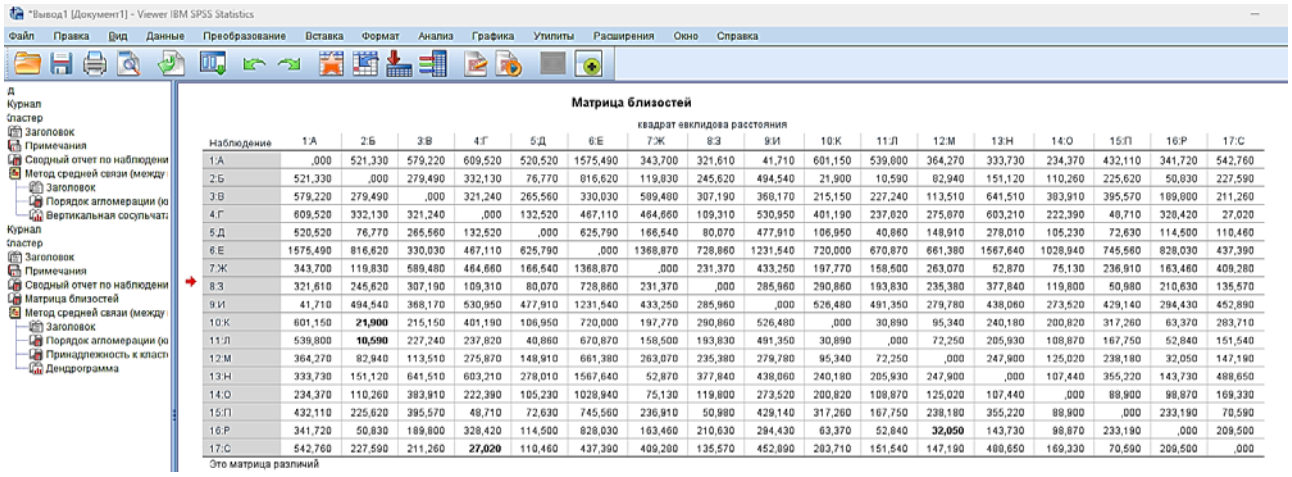


Рисунок 8.12 – «Матрица близостей», полученная в результате выполнения иерархического кластерного анализа в IBM SPSS Statistics 26

Однако пока неясно, что означают полученные три кластера. Разобраться в этом нам помогут кластерные профили.

ШАГ 5 Для каждого из полученных кластеров сделаем содержательную интерпретацию, определив кластерные профили.

Кластерные профили представляют собой распределенные по кластерной принадлежности средние значения переменных, которые включены в анализ.

Если рассмотреть содержимое таблицы в редакторе данных, то заметно, что там добавилась переменная clu2_1 (см. рис. 8.11). Эта переменная указывает на кластерную принадлежность каждого наблюдения и может быть использована для расчета кластерного профиля.

Выберите в меню *Анализ – Сравнение средних – Средние (Analyze – Compare Means – Means...)* (рис. 8.13).

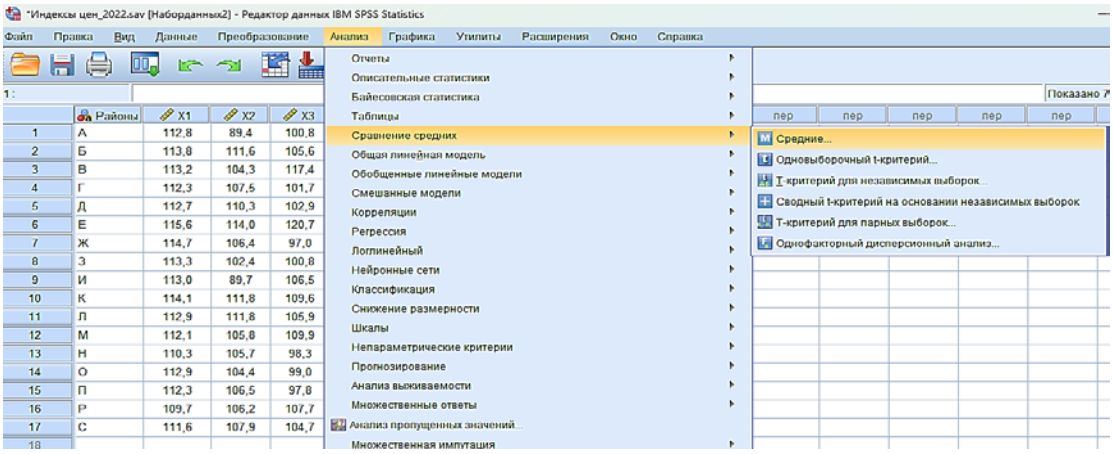
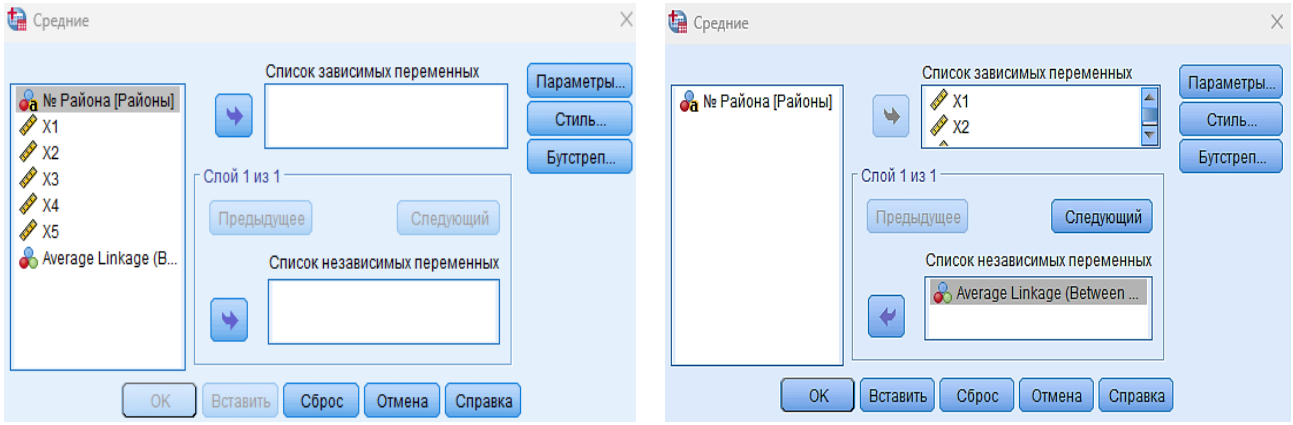


Рисунок 8.13 – Выбор «Средние…» в процедуре «Сравнение средних» в меню «Анализ» IBM SPSS Statistics 26

Переменным заданиям присвойте статус зависимых, переменной clu2_1 – статус независимой (рис. 8.14) и выполните расчет.



а *б*

Рисунок 8.14 – Диалоговое окно «Средние» в IBM SPSS Statistics 26

В качестве результатов выводятся средние значения (или стандартные отклонения) переменных (табл. 8.4).

Таблица 8.4 – Средние значения переменных

Average Linkage (Between Groups)		Потребители (X1)	Промышленность (X2)	Сельское хозяйство (X3)	Строительство (X4)	Перевозки (X5)
1	Среднее	112,606	105,106	104,100	108,400	109,500
	N	16	16	16	16	16
	Стандартное отклонение	1,2772	6,6856	5,4411	2,6964	6,5052
2	Среднее	115,600	114,000	120,700	107,300	129,000
	N	1	1	1	1	1
	Стандартное отклонение	.	.	.	.	.
Всего	Среднее	112,782	105,629	105,076	108,335	110,647
	N	17	17	17	17	17
	Стандартное отклонение	1,4341	6,8232	6,6306	2,6244	7,8766

Из таблицы 8.4 видим, что в первый кластер попали районы, имеющие примерно одинаковый уровень цен по всем пяти секторам (при этом, немного выше индекс потребительских цен), район из второго кластера отличается высокими средними значениями следующих переменных: индексом тарифов на грузовые перевозки – 129 %, индексом цен производителей на реализованную сельскохозяйственную продукцию – 120,7 % и индексом потребительских цен.

Для принятия решения об оценке кластерных профилей необходимо измерить колеблемость показателей, т. е. определить степень отклонения значения от средней величины. На практике применяют два критерия: дисперсию и среднее квадратичное отклонение действительных результатов от среднего значения.

Для сравнительной оценки важное значение составляют показатели относительной колеблемости, т. е. коэффициент вариации. *Коэффициент вариации (variation coefficient)* – это относительная величина, с помощью

которой можно сравнивать колеблемость (изменчивость) признаков (переменных), выраженных в разных величинах измерения и определяется как отношение среднего квадратического отклонения к среднему значению.

Коэффициент вариации может изменяться от 0 до 100 (до 10 % – слабая колеблемость, 10–25 % – умеренная, свыше 25 % – высокая).

Сравнительный анализ коэффициентов вариации для оценки кластерных профилей представлен в таблице 8.5.

Таблица 8.5 – Коэффициенты вариации для оценки кластерных профилей

Average Linkage (Between Groups)		Потребители (X1)	Промышленность (X2)	Сельское хозяйство (X3)	Строительство (X4)	Перевозки (X5)
1	Коэффициент вариации, %	1,13	6,36	5,23	2,49	5,94
2	Коэффициент вариации, %	–	–	–	–	–
Всего	Коэффициент вариации, %	1,27	6,46	6,31	2,42	7,12

Таким образом, коэффициенты вариации первого кластера характеризуются слабой колеблемостью, не превышающей 6,5 %. В целом по двум кластерам изменчивость всех коэффициентов вариации также сохранялась слабой, не превышающей 7,15 %. Вместе с тем, наименьшее значение коэффициента вариации сложилось в секторе «Потребители» – 1,27 % (индекс потребительских цен), а наибольшее в секторе «Перевозки» до 7,12 %. Это свидетельствует о том, что на кластерные профили наибольшее влияние оказывало изменение в районах индекса тарифов на грузовые перевозки, а наименьший разброс результатов относительно среднего обусловлен незначительным изменением индекса потребительских цен в районах.

Замечание. Иерархическим кластерный анализ не накладывает каких-либо серьезных ограничений на данные, необходимые для его проведения. Переменные могут быть как количественными, так и качественными. Соответственно формы связи между признаками и между признаками и факторами могут быть любыми. Нормальное распределение данных также не требуется. Однако у этого метода есть и свои недостатки. Один из главных заключается в отсутствии статистических критериев оценки качества построенной модели. Также остается открытым вопрос о возможности перенесения результатов кластерного анализа на всю генеральную совокупность.

Задание 8.2. В таблице 8.1 приведены индексы цен и тарифов в секторах экономики по районам области за 2022 год. Выполните пошаговый алгоритм многомерной кластеризации данных с помощью метода *k*-средних в программе IBM SPSS Statistics.

1. Выполните подготовительные шаги в программе IBM SPSS Statistics для кластерного анализа экономических данных.

2. Определите наиболее подходящее количество кластеров для разбиения районов на группы по всем пяти признакам.

3. Выполните классификацию районов области методом *k*-средних.

4. Дайте содержательную интерпретацию результатов кластерного анализа методом *k*-средних.

Методические указания к выполнению задания 8.2 в IBM SPSS Statistics

При проведении многомерного кластерного анализа сначала выполните ряд подготовительных шагов. Подготовьте рабочий файл данных, запустите программу IBM SPSS Statistics 26 и откройте файл с исходными данными для анализа.

ШАГ 1. В строке меню *Анализ* выберите команды *Анализ – Классификация – Кластеризация k-средними (Analyze – Classify – K-Means Cluster...)* (рис. 8.15).

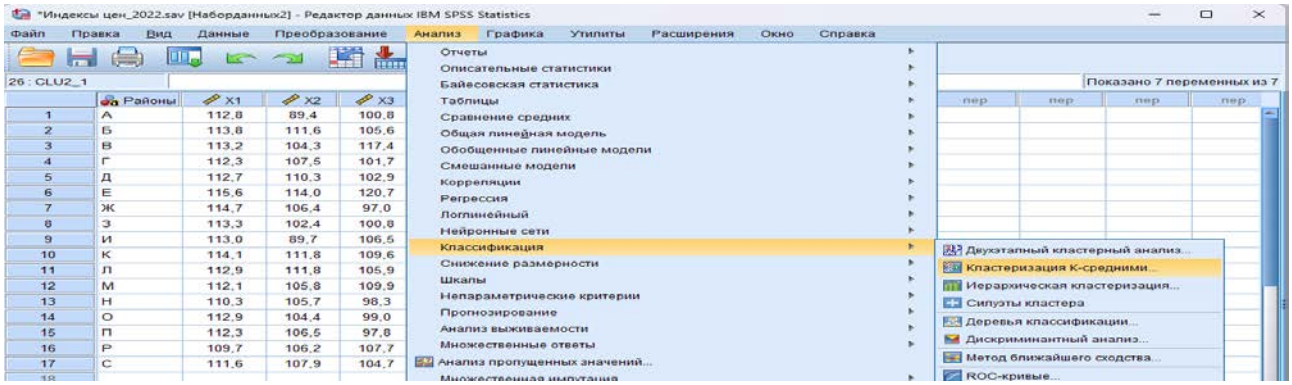
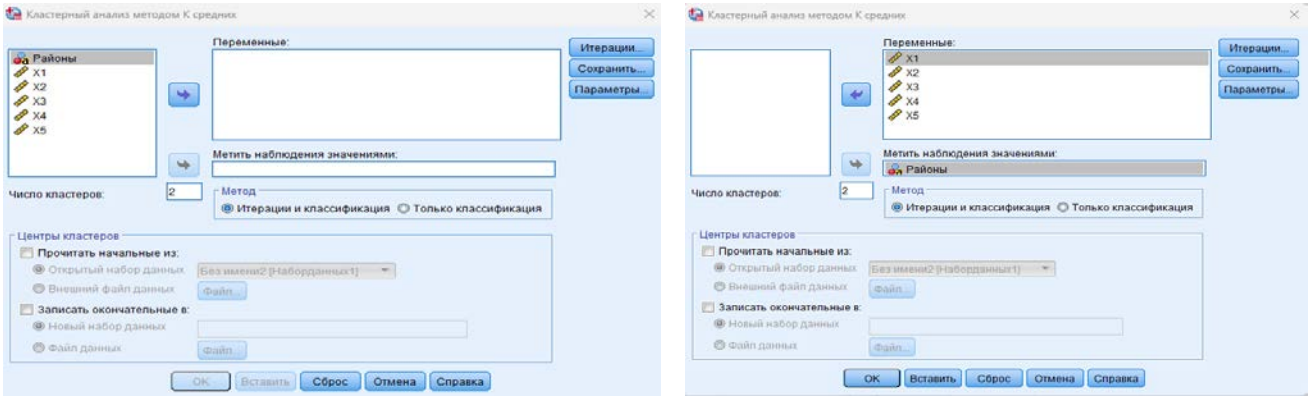


Рисунок 8.15 – Выбор кластеризации *k*-средними в процедуре «Классификация» в меню «Анализ» IBM SPSS Statistics 26

Метод *k*-средних относят к итеративным методам классификации. Недостаток их заключается в том, что необходимо заранее задать количество кластеров, а не получить его в качестве результата, как в иерархическом анализе.

Откройте диалоговое окно *Кластерный анализ методом k-средних (K-Means Cluster Analysis)* (рис. 8.16 а).

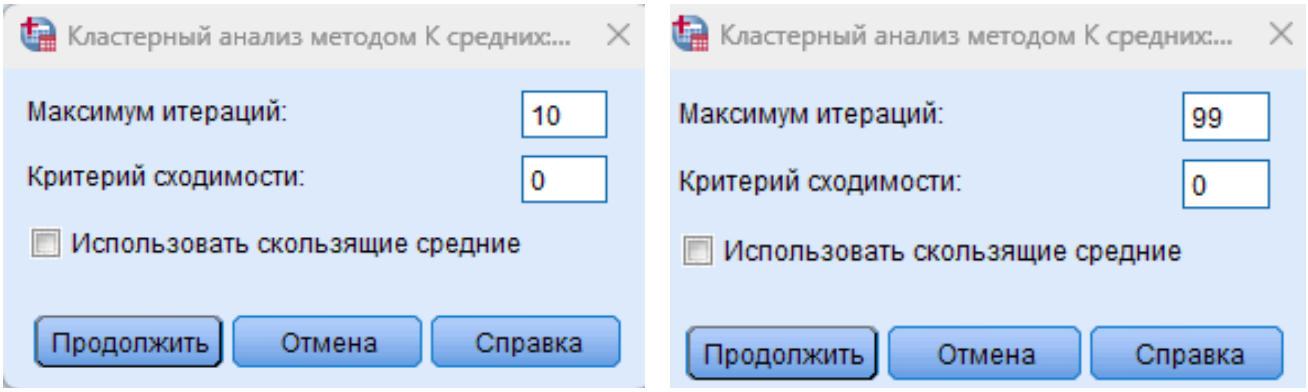


а *б*

Рисунок 8.16 – Диалоговое окно «Кластерный анализ методом *k*-средних»

ШАГ 2. Переменные поместите в поле тестируемых переменных (рис. 8.16 б). Теперь нужно указать количество кластеров. Предварительно проведенный для этих данных иерархический кластерный анализ (см. задание 8.1 ШАГ 2) позволил определить оптимальное количество кластеров, равное 2. Это значение и вводим в поле *Число кластеров (Number of Clusters)*.

ШАГ 3. Через выключатель *Итерации... (Iterate...)* укажите число итераций, равное 99 (рис. 8.17 б); установленного по умолчанию количества итераций 10 будет недостаточно (рис. 8.17 а). Как и ранее, воспользуйтесь выключателем *Сохранить... (Save...)*, чтобы с помощью дополнительных переменных зафиксировать принадлежность наблюдений к кластеру. Для начала расчетов нажмите ОК.



а *б*

Рисунок 8.17 – Диалоговое окно «Кластерный анализ методом *k*-средних: Итерации…» в IBM SPSS Statistics 26

Во вкладке *Параметры* можно задать вывод статистки модели: начальные и конечные центры кластеров, конечный кластер для каждого наблюдения, а также таблицу дисперсионного анализа (ANOVA) (рис. 8.18).

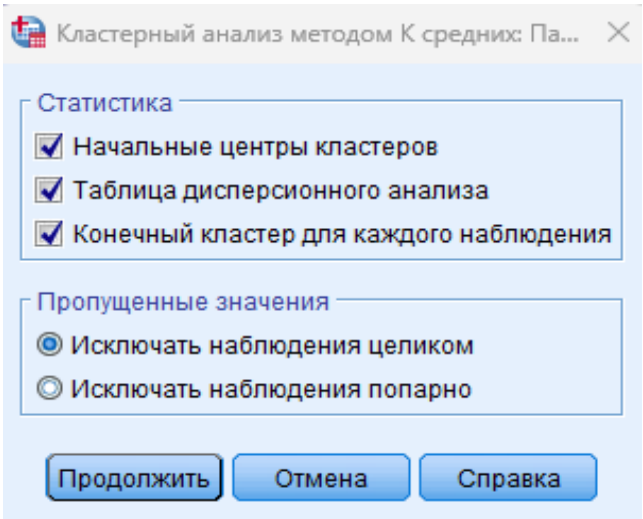


Рисунок 8.18 – Диалоговое окно «Кластерный анализ методом *k*-средних: Параметры…» в IBM SPSS Statistics 26

Замечание. В кластерном анализе методом К-средних нет возможности стандартизировать переменные, значения которых представлены в разном масштабе. Поэтому такие переменные следует преобразовывать заранее, сохраняя их как новые стандартизированные переменные. Это можно сделать в окне *«Описательные»* вкладки *«Описательные статистики»* меню *«Анализ»* (рис. 8.19).

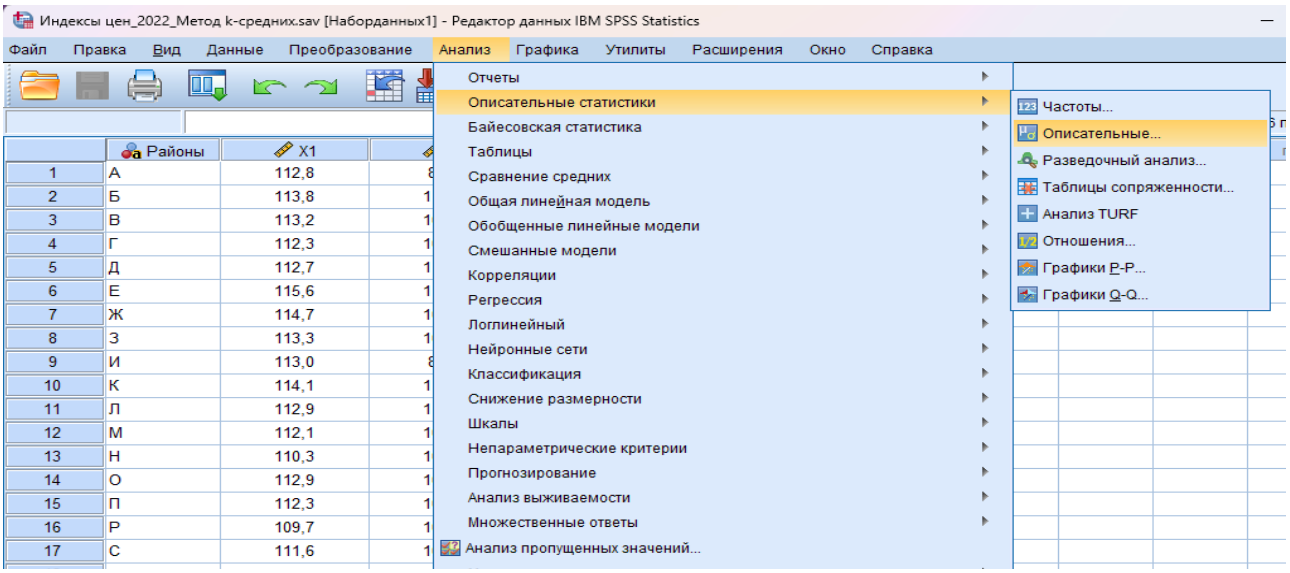
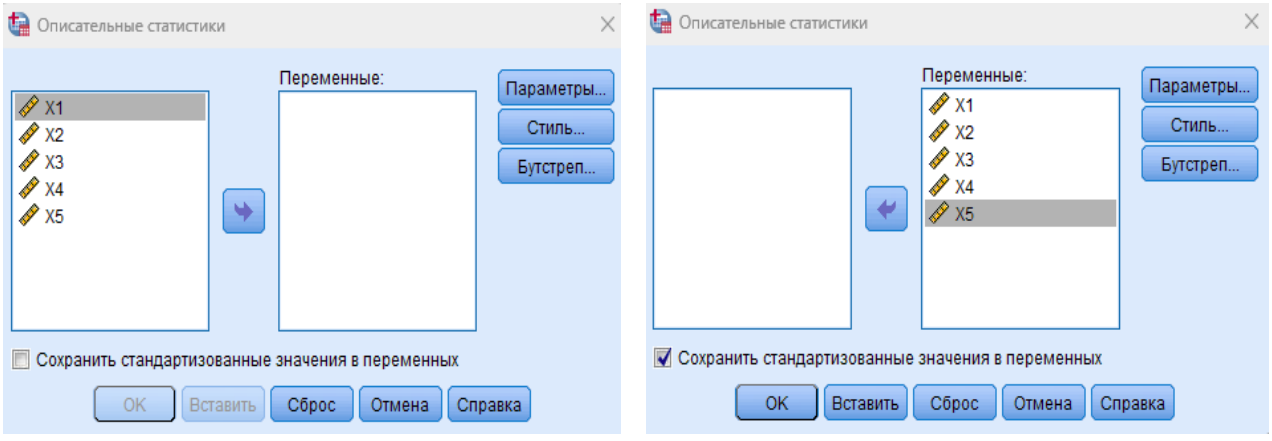


Рисунок 8.19 – Выбор «Описательные» во вкладке «Описательные статистики» в меню «Анализ» IBM SPSS Statistics 26 для стандартизации переменных

Выберите переменные, которые требуется стандартизировать (рис. 8.20 а), и, перенеся их в поле *«Переменные»* справа, поставить галочку *«Сохранить стандартизированные значения в переменных»* (рис. 8.20 б).



а *б*

Рисунок 8.20 – Диалоговое окно «Описательные статистики» в IBM SPSS Statistics 26 для стандартизации переменных

После этого в исходном файле данных появляются Z-переменные. В таком случае кластерному анализу подлежат именно эти переменные.

Методические рекомендации к интерпретации результатов кластерного анализа методом *k*-средних в IBM SPSS Statistics для задания 8.2

ШАГ 4. В отчете (табл. 8.6–8.12) могут быть представлены первичные кластерные центры, история итераций (т. е. как изменялись координаты кластерных центров с каждым шагом и на какой итерации это изменение прекратилось), окончательные центры кластеров, расстояние между конечными центрами, информация о принадлежности к кластерам, расстоянии объектов до кластерных центров, числе наблюдений в каждом кластере и т.п.

В таблице «Начальные центры кластеров» показаны начальные оценки положения средних для каждого кластера (табл. 8.6).

Таблица 8.6 – Начальные центры кластеров

	Кластер	
	1	2
Потребители (X1)	112,8	115,6
Промышленность (X2)	89,4	114,0
Сельское хозяйство (X3)	100,8	120,7
Строительство (X4)	107,1	107,3
Перевозки (X5)	105,2	129,0

В таблице «Хронология итераций» демонстрируется, сколько потребовалось осуществить итераций до момента стабилизации кластерных центров и как эти центры изменялись (табл. 8.7).

Таблица 8.7 – Хронология итераций[a]

Итерация	Изменение центров кластеров	
	1	2
1	15,846	14,018
2	0,000	0,000

a. Сходимость достигнута благодаря небольшому изменению центров кластеров или без него. Максимальное изменение абсолютной координаты для любого центра: 0,000. Текущая итерация – 2. Минимальное расстояние между начальными центрами: 39,692.

В таблице «Конечные центры кластеров» показаны центры каждого кластера по двум группирующим переменным (табл. 8.8). Сопоставляя конечные и начальные центры кластеров можно проследить динамику их изменения.

Таблица 8.8 – Конечные центры кластеров

	Кластер	
	1	2
Потребители (X1)	112,7	113,2
Промышленность (X2)	104,8	108,4
Сельское хозяйство (X3)	103,2	111,1
Строительство (X4)	109,0	106,1
Перевозки (X5)	107,5	120,9

В таблице «Принадлежность к кластерам» для каждого наблюдения указывается финальный кластер, к которому оно отнесено, и евклидово

расстояние до центра этого кластера (табл. 8.9).

Таблица 8.9 – Принадлежность к кластерам

Номер наблюдения	Районы	Кластер	Расстояние
1	А	1	15,846
2	Б	1	7,743
3	В	2	10,174
4	Г	2	9,577
5	Д	1	8,120
6	Е	2	14,018
7	Ж	1	9,243
8	З	1	10,046
9	И	1	15,527
10	К	1	10,232
11	Л	1	7,553
12	М	1	8,101
13	Н	1	11,095
14	О	1	4,933
15	П	1	10,799
16	Р	1	5,999
17	С	2	7,536

В таблице «Расстояния между конечными центрами кластеров» выводится евклидово расстояние между центрами финальных кластеров (табл. 8.10).

Таблица 8.10 – Расстояния между конечными центрами кластеров

Кластер	1	2
1		16,210
2	16,210	

Все кластеры соотносятся друг с другом, поэтому значения верхней правой части таблицы аналогичны значениям нижней левой. Эта таблица помогает определить оптимальное число кластеров. Чем больше усредненное расстояние между центрами кластеров, тем лучше модель. Минимальное пороговое значение – 2. В хорошей кластерной модели все расстояния должны быть больше 2 (иначе считается, что кластеры расположены слишком близко друг от друга и их сложно определить).

Таблица «ANOVA» (однофакторный дисперсионный анализ) служит только для целей описания, но не для проверки гипотезы о равенстве средних кластеров (табл. 8.11). Строгая статистическая интерпретация данных таблицы дисперсионного анализа невозможна, поскольку в ходе кластерного анализа объекты (наблюдения) специально разделяются по кластерам так, чтобы максимально усилить различия между ними, а в основе дисперсионного анализа лежит предположение о равенстве дисперсий. Проще говоря, в кластерном анализе группы должны быть разнородными, а в дисперсионном анализе – однородными. Поэтому использовать таблицу ANOVA можно лишь в информационных целях: из нее видно, по каким исходным переменным средние значения различаются существенно, а по каким – нет.

Таблица 8.11 – Дисперсионный анализ (ANOVA)

	Кластер		Ошибка		F	Значимость
	Средний квадрат	ст.св.	Средний квадрат	ст.св.		
Потребители (X1)	0,806	1	2,140	15	0,377	0,548
Промышленность (X2)	40,880	1	46,934	15	0,871	0,365
Сельское хозяйство (X3)	191,366	1	34,138	15	5,606	0,032
Строительство (X4)	26,136	1	5,604	15	4,664	0,047
Перевозки (X5)	544,523	1	29,875	15	18,227	0,001

F-критерий следует использовать только для целей описания, так как кластеры выбраны так, чтобы разница между наблюдениями в разных кластерах была максимальной. Наблюдаемые уровни значимости не скорректированы для этого, и поэтому их нельзя использовать для проверки гипотезы о равенстве средних кластеров.

F-критерий и его статистическая значимость по каждой группирующей переменной (потребители, промышленность, сельское хозяйство, строительство и перевозки) помогают установить оптимальное число кластеров. Каждая переменная вносит определенный вклад в кластеризацию, по одной из переменных кластеры могут различаться, а подругой нет. Чем больше значение F-критерия, тем больший вклад вносит переменная в кластеризацию. Если статистическая значимость переменной (последний столбец таблицы 8.11) больше или равна 0,05, то это свидетельствует о том, что данная переменная не существенна для классификации.

В таблице «Число наблюдении в каждом кластере» показано, сколько объектов вошло в каждый кластер (табл. 8.12).

Таблица 8.12 – Число наблюдений в каждом кластере

Кластер	1	13,000
	2	4,000
Валидные		17,000
Пропущенные		0,000

Наполненность кластеров также является одним из показателей качества модели. Хорошая модель должна быть достаточно наполненной, с относительно равномерным распределением кластеров. Исследовательские задачи могут быть разными, тем не менее, одно из основных свойств любой модели – ее информативность.

Еще одним показателем для определения оптимального числа кластеров является компактность кластеров, показывающая дисперсию расстояния объектов от центров кластеров. Лучшей считается та модель, где дисперсия расстояния объектов вокруг центров кластера меньше.

В заключение отметим, что для содержательной интерпретации групп объектов (наблюдений), полученных при использовании метода *k*-средних, нет необходимости определять дополнительно кластерные профили, так как ими фактически являются конечные центры кластеров.

Задания для самостоятельного решения

Задание 8.3. В таблице 8.13 приведены индексы цен и тарифов в секторах экономики Республики Беларусь за 2007–2023 годы: индекс потребительских цен (X1), индекс цен производителей промышленной продукции (X2), индекс цен производителей сельскохозяйственной продукции (X3), индекс цен в строительстве (X4), индекс тарифов на перевозку грузов (X5), индекс тарифов на услуги связи для юридических лиц и индивидуальных предпринимателей (X6), индекс цен оптовых продаж товаров (продукции) (X7).

Таблица 8.13 – Индексы цен и тарифов в секторах экономики Беларуси в %

Год	X1	X2	X3	X4	X5	X6	X7
2007	108,4	121,2	113	110,8	111,8	102,4	111,1
2008	114,8	115,2	135,5	118,3	118,0	103,3	117,1
2009	113,0	113,6	103,8	111,4	133,2	104,0	112,8
2010	107,8	113,6	117,6	111,2	139,3	99,9	114,3
2011	153,2	171,4	185,4	148,1	154,7	110,3	100,0
2012	159,2	176,0	191,2	191,2	177,2	193,9	168,3
2013	118,3	113,6	112,9	124,2	121,9	120,8	97,0
2014	118,1	112,8	120,7	113,4	109,5	113,9	99,6
2015	113,5	117,2	105,2	113,2	115,3	118,8	118,9
2016	111,8	112,0	107,7	109,5	110,3	114,4	115,1
2017	106,0	109,8	113,6	106,2	104,1	107,4	112,9
2018	104,9	106,8	109,2	110,9	107,8	104,1	114,5
2019	105,6	106,3	107,7	109,0	106,2	104,6	103,8
2020	105,5	105,6	108,0	109,8	112,8	103,2	99,2
2021	109,5	112,2	115,3	111,7	110,7	102,7	123,9
2022	115,2	115,2	124,3	114,2	120,1	104,0	127,0
2023	105,1	105,8	104,6	111,4	112,8	100,8	98,7

Выполните пошаговый алгоритм многомерного кластерного анализа в IBM SPSS Statistics.

1. Выполните подготовительные шаги в программе IBM SPSS Statistics для кластерного анализа экономических данных.
2. Определите наиболее подходящее количество кластеров для разбиения наблюдений на группы по всем семи признакам.
3. Выполните классификацию наблюдений иерархическим методом.
4. Дайте графическое представление процесса объединения наблюдений в виде дендрограммы.
5. Для каждого из полученных кластеров дайте содержательную интерпретацию, определив кластерные профили.

Задание 8.4. В таблице 8.14 представлены значения шести показателей, характеризующих условия жизни населения в 27 странах: потребление мяса и мясопродуктов на душу населения, кг. (X1), потребление масла животного на душу населения, кг. (X2), оценка валового внутреннего продукта по паритету

покупательной способности в некотором году на душу населения, % к США (X3), расходы на здравоохранение, % от ВВП (X4), потребление фруктов и ягод на душу населения, кг. (X5), потребление хлебных продуктов на душу населения, кг. (X6).

Таблица 8.14 – Макроэкономические показатели уровня жизни населения по странам

Страна	X1	X2	X3	X4	X5	X6
Австралия	100	2,6	71,4	8,5	121	87
Австрия	93	5,3	78,7	9,2	146	74
Азербайджан	20	4,1	12,1	3,3	52	141
Армения	20	3,7	10,9	3,2	72	134
Беларусь	72	3,6	20,4	5,4	38	120
Бельгия	85	6,9	79,7	8,3	83	72
Болгария	65	3,0	17,3	5,4	92	156
Великобритания	67	3,5	69,7	7,1	91	91
Венгрия	73	1,7	24,5	6,0	73	106
Германия	88	6,8	76,2	8,6	138	73
Греция	83	1,0	44,4	5,7	99	108
Грузия	21	3,8	11,3	3,5	55	140
Дания	98	5,0	79,2	6,7	89	77
Ирландия	99	3,3	57,0	6,7	87	102
Испания	89	0,4	54,8	7,3	103	72
Италия	84	2,2	72,1	8,5	169	118
Казахстан	61	4,2	13,4	3,3	10	191
Канада	98	3,1	79,9	10,2	123	77
Киргизия	46	4,1	11,2	3,4	20	134
Нидерланды	86	3,4	72,4	8,7	176	59
Португалия	73	3,2	48,6	7,3	150	83
Россия	55	3,9	20,4	3,2	28	124
США	115	1,9	100,0	14,1	99	103
Финляндия	62	5,8	63,9	8,8	82	94
Франция	91	8,8	77,5	9,8	84	85
Чехия	82	8,2	34,7	1,9	65	114
Япония	40	0,7	83,5	7,3	60	119

Выполните классификацию наблюдений иерархическим методом в программах IBM SPSS Statistics или Statistica.

Для каждого из полученных кластеров дайте содержательную интерпретацию результатов, определив кластерные профили.

Задание 8.5. В таблице 8.15 приведены индексы цен и тарифов в секторах экономики Республики Беларусь за 2011 и 2019 годы по месяцам: индекс потребительских цен (X1), индекс цен производителей промышленной продукции (X2), индекс цен производителей сельскохозяйственной продукции (X3), индекс цен на строительно-монтажные работы (в строительстве) (X4), индекс тарифов на перевозку грузов (X5), индекс тарифов на услуги связи для юридических лиц и индивидуальных предпринимателей (X6).

Таблица 8.15 – Индексы потребительских цен и цен производителей в 2011 г. в секторах экономики Беларуси в %

	X1	X2	X3	X4	X5	X6
Январь	101,4	102,6	107,0	103,1	103,9	99,3
Февраль	104,2	105,1	109,9	97,9	104,6	99,6
Март	106,1	107,7	112,6	98,1	106,3	99,6
Апрель	110,9	111,8	119,1	104,6	109,1	100,0
Май	125,4	122,4	126,9	106,4	145,5	101,4
Июнь	136,2	148,5	135,0	112,9	150,9	104,2
Июль	141,0	157,6	148,2	122,9	168,9	106,1
Август	153,6	167,1	163,3	132,1	170,9	106,1
Сентябрь	174,5	182,5	193,3	144,0	180,3	106,1
Октябрь	188,7	207,2	231,8	159,8	246,5	110,1
Ноябрь	204,0	240,0	261,8	177,4	253,6	134,0
Декабрь	208,7	249,6	274,4	197,6	250,3	138,6

Таблица 8.16 – Индексы потребительских цен и цен производителей в 2019 г. в секторах экономики Беларуси в %

	X1	X2	X3	X4	X5	X6
Январь	100,9	100,9	100,9	102,3	101,4	101,4
Февраль	102,2	101,5	104,1	96,9	102,1	102,1
Март	102,6	102,1	104,6	96,6	101,5	101,5
Апрель	102,7	102,8	104,4	100,9	101,3	101,3
Май	103,0	103,3	102,2	101,7	100,9	100,9
Июнь	102,9	103,9	100,6	102,4	100,7	100,7
Июль	103,1	104,0	102,6	103,3	100,5	100,5
Август	102,9	104,2	102,5	103,9	107,3	107,3
Сентябрь	103,4	104,3	104,3	105,5	107,3	107,3
Октябрь	103,9	104,1	106,5	103,8	107,1	107,1
Ноябрь	104,2	104,0	106,9	105,3	107,7	107,7
Декабрь	104,4	104,4	108,1	106,2	108,7	108,7

Выполните пошаговый алгоритм многомерного кластерного анализа методом *k*-средних в пакете IBM SPSS Statistics или Statistica.

Для каждого из полученных кластеров дайте содержательную интерпретацию результатов, определив кластерные профили.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Анализ данных : учебник / В. С. Мхитарян [и др.]; под ред. В. С. Мхитаряна. – Москва : Изд-во Юрайт, 2023. – 490 с.
2. Архипова, М. Ю. Анализ и моделирование патентной активности в России и развитых странах мира / М. Ю. Архипова, Е. С. Карпов // Ресурсы. Информация. Снабжение. Конкуренция (РИСК). – Москва : ОАО «ИТКОР». 2012. – № 4. – С. 286–293.
3. Бондарик, И. А. Исследование качества жизни населения районов Витебской области / И. А. Бондарик, Е. А. Дубровина, К. Р. Быков// Тезисы докладов 46 Республиканской науч.-технич. конф. преп. и студ. / УО ВГТУ. – Витебск, 2013. – С. 37.
4. Бородич, С. А. Эконометрика. Практикум : учебное пособие / С.А. Бородич. – Москва : ИНФРА-М, 2021. – 329 с. : [Электронный ресурс] – URL: https://znanium.com/catalog/product/1228789 – Дата доступа: 19.08.2024.
5. Быков, К. Р. Снижение размерности системы факторов в анализе эффективности использования основных средств / К. Р. Быков // Материалы докладов междунар. науч.-практ. конф. «Социально-экономическое развитие организаций и регионов Беларуси: цели, приоритеты, механизмы структурных преобразований» : в 2 ч. / УО «ВГТУ». – Витебск, 2011. – Ч. 1. – С. 78–82 : [Электронный ресурс] – URL: https://rep.vstu.by/handle/123456789/4220. – Дата обращения: 28.07.2024.
6. Быков, К. Р. Анализ системы факторов повышения эффективности использования основных средств с помощью методов прикладной статистики на предприятиях машиностроения Беларуси / К. Р. Быков, Ф. Б. Марголин // Эффективное управление предприятием и регионом: сб. науч. ст. в 2 ч. / ГрГУ им. Я. Купалы; редкол.: Ли Чон Ку (гл. ред.) [и др.]. – Гродно : ГрГУ, 2011. – Ч. 1. – С. 204–211. : [Электронный ресурс] – URL: https://elib.grsu.by/doc/3124. – Дата обращения: 03.08.2024.
7. Быков, К. Р. Прогнозирование эффективности использования основных средств организаций / К. Р. Быков // Вест. Витебск. гос. технолог. ун-та. – 2012. – Вып. 23. – С. 125–136. : [Электронный ресурс] – URL: https://rep.vstu.by/handle/123456789/380. – Дата обращения: 01.09.2024.
8. Быков, К. Р. Моделирование дифференцированной оценки производственной эффективности деятельности организаций. // Междунар. науч.-технич. конф., приуроченная к 50-летию МРТИ-БГУИР (Минск, 18–19 марта 2014 г.) : материалы конф. В 2 ч. Ч. 2. – Минск, 2014. – С. 229–231. : [Электронный ресурс] – URL: https://libeldoc.bsuir.by/handle/123456789/27346. – Дата обращения: 15.10.2024.
9. Быков, К. Р. Методические инструменты оценки качества жизни населения в эпоху цифровизации / К. Р. Быков // Материалы докладов 57-й Международной науч.-технич. Конф. преподавателей и студентов : в 2 т. / УО «ВГТУ». – Витебск, 2024. – Т. 1. – С. 158–160. : [Электронный ресурс] –

URL: http://www.rep.vstu.by/handle/123456789/19177. – Дата обращения: 06.11.2024.

10. Григорьев, А. А. Методы и алгоритмы обработки данных : учеб. пособие / А. А. Григорьев, Е. А. Исаев. – 2-е изд., перераб. и доп. – Москва : ИНФРА-М, 2022. – 383 с.

11. Дубров, А. М. Многомерные статистические методы : учебник / А. М. Дубров, В. С. Мхитарян, Л. И. Трошин. – Москва : Финансы и статистика, 2003. – 352 с.

12. Иванова, И. И. Факторный анализ эффективности использования основных средств ОАО «Красный Октябрь» / И. И. Иванова, О. Раева, К. Р. Быков // Тезисы докладов 45 Республиканской науч.-технич. конф. преп. и студ., посвященной году книги. – Витебск : УО ВГТУ, 2012. – С. 153–154.

13. Калинина, В. Н. Анализ данных. Компьютерный практикум : учебное пособие / В. Н. Калинина, В. И. Соловьев. – Москва : КноРус, 2022. – 166 с. – [Электронный ресурс] URL: https://book.ru/book/942681 – Дата обращения: 12.12.2024.

14. Козлов, А. Ю. Статистический анализ данных в MS Excel : учеб. пособие / А. Ю. Козлов, В. С. Мхитарян, В. Ф. Шишов. – Москва : ИНФРА-М, 2023. – 320 с.

15. Лексакова, Е. М. Многомерный факторный анализ эффективности использования основных средств организации / Е. М. Лексакова, К. Р. Быков // Тезисы докладов 46 Республиканской науч.-технич. конф. преп. и студ. / УО ВГТУ. – Витебск, 2013. – С. 34.

16. Логунова, О. С. Обработка экспериментальных данных на ЭВМ : учебник / О. С. Логунова, П. Ю. Романов, Е.А. Ильина. – 2-е изд., испр. и доп. – Москва : ИНФРА-М, 2021. – 377 с.

17. Ляликова, В. И Анализ факторов, влияющих на конкурентоспособность регионов Беларуси / В. И. Ляликова, А. В. Зенькова // Конкурентоспособность предприятий и регионов в глобальной экономике : сб. науч. статей. В 2 ч. Ч. 1 / УО «Гродненский гос. ун-т имени Янки Купалы» ; ред. кол.: Ли Чон Ку [и др.]. – Гродно : ГрГУ, 2009. – С. 271–277. : [Электронный ресурс] – URL: https://elib.grsu.by/doc/16131 – Дата доступа: 10.07.2024.

18. Ляликова, В. И Классификация районов Гродненской области по качеству жизни населения / В. И. Ляликова // Веснік Гродзенскага дзяржаўнага ўніверсітэта імя Янкі Купалы. Сер. 5, Эканоміка. – 2009. – № 2. – С. 63–71. : [Электронный ресурс]. – URL: https://elib.grsu.by/doc/17723 – Дата доступа: 02.10.2024.

19. Ляликова, В. И. Взаимосвязь индекса человеческого развития и качества жизни населения в регионах Беларуси и Польши / В. И. Ляликова, Е. Н. Калинина // Экономика, моделирование, прогнозирование : сб. науч. тр. Вып. 12 / ГНУ «НИЭИ Минэкономики Респ. Беларусь» ; ред. кол.: М. К. Кравцов [и др.]. – Минск : НИЭИ Минэкономики Респ. Беларусь, 2018. – С. 35–42. : [Электронный ресурс] – URL: https://elib.grsu.by/doc/48135 – Дата

доступа: 06.07.2024.

20. Максимов, С.И. Excel 2013 и SPSS 21 в решении задач прикладной статистики : учебно-методическое пособие / С. И. Максимов, Е. М. Зайцева; ГУО «Республиканский ин-т высш. шк.». – 3-е изд., испр. и доп. – Минск : РИВШ, 2023. – 134 с.

21. Максимов, С. И. Технологии обработки данных исследований в IBM SPSS Statistics: учебно-методическое пособие / С. И. Максимов, Е. М. Зайцева; ГУО «Республиканский ин-т высш. шк.». – Минск : РИВШ, 2016. – 99 с.

22. Малиновская, А. В. Факторный анализ финансового состояния ОАО «КИМ» / А. В. Малиновская, М. М. Макаревич, К. Р. Быков // Тезисы докладов 45 Республиканской науч.-технич. конф. преп. и студ., посвященной году книги. – Витебск : УО «ВГТУ» , 2012. – С. 156–157.

23. Методология статистических исследований социально-экономических процессов : монография / Н. А. Садовникова, Е. В. Зарова, М. В. Карманов [и др.]. – Москва : Русайнс, 2023. – 256 с. – [Электронный ресурс] – URL: https://book.ru/book/952655 – Дата обращения: 12.12.2024.

24. Многомерный статистический анализ в экономических задачах: компьютерное моделирование в SPSS : учебное пособие / под ред. И. В. Орловой. – Москва : Вузовский учебник : ИНФРА-М, 2024. – 310 с.: [Электронный ресурс] – URL: https://znanium.com/catalog/product/2073497 – Дата доступа: 09.11.2024.

25. Моосмюллер, Г. Маркетинговые исследования с SPSS : учеб. пособие / Г. Моосмюллер, Н. Н. Ребик. – 2-е изд. – Москва : ИНФРА-М, 2021. – 200 с.

26. Наследов, А. SPSS 19: профессиональный статистический анализ данных: учебник / А. Наследов. – СПб.: Питер, 2011. – 400 с.

27. Ниворожкина, Л. И. Многомерные статистические методы в экономике : учебник / Л. И. Ниворожкина, С. В. Арженовский. – Москва : РИОР: ИНФРА-М, 2023. – 203 с.

28. Орлова, И. В. Экономико-математические методы и модели: компьютерное моделирование : учебное пособие / И. В. Орлова, В.А. Половников. – 3-е изд., перераб. и доп. – Москва : Вузовский учебник : ИНФРА-М, 2024. – 389 с. : [Электронный ресурс] – URL: https://znanium.com/catalog/product/2056791 – Дата обращения: 21.10.2024.

29. Официальный сайт Национального статистического комитета Республики Беларусь // Режим доступа: https://www.belstat.gov.by/ofitsialnaya-statistika/publications/. – Дата обращения: 12.11.2024.

30. Салин, В. Н. Статистический анализ данных цифровой экономики в системе «STATISTICA» : учебно-практическое пособие / В. Н. Салин, Э. Ю. Чурилова. – Москва : КНОРУС, 2023. – 238 с.

31. Сошникова, Л. А. Многомерные статистические методы. Практикум : учеб. пособие / Л. А. Сошникова. – Минск : БГЭУ, 2024. – 230 с.

32. Теория статистики : учебник / под ред. Г.Л. Громыко. – 4-е изд., перераб. и доп. – Москва : ИНФРА-М, 2021. – 465 с. – [Электронный ресурс] – URL: https://znanium.com/catalog/product/1497872. – Дата обращения:

02.08.2024.

33. Тихомиров, Д. А. Анализ данных (с применением программы SPSS) : учебник / Д. А. Тихомиров. – Москва : КноРус, 2022. – 246 с. [Электронный ресурс]. – URL: https://book.ru/book/942678 – Дата обращения: 15.11.2024.

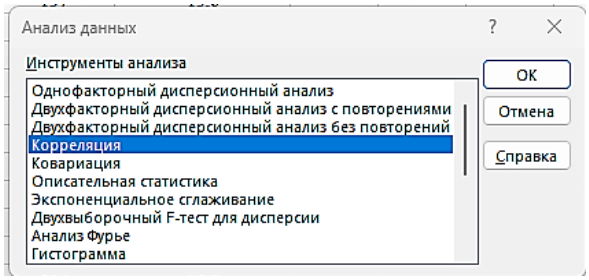
34. Толмачев, М. Н. Бизнес-аналитика: Статистическая обработка данных : учебное пособие / М. Н. Толмачев, Э. Ю. Чурилова. – Москва : КноРус, 2024. – 283 с. – [Электронный ресурс] – URL: https://book.ru/book/955151 – Дата обращения: 12.12.2024.

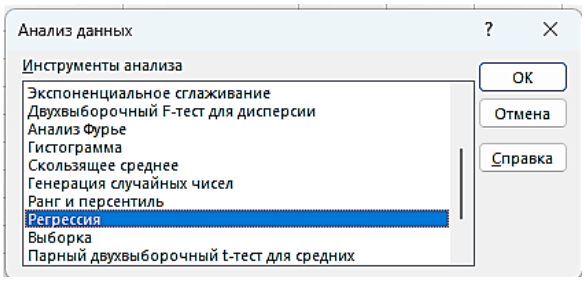
35. Фролов, Ю. В. Анализ результатов маркетинговых исследований в системе STATISTICA (на примерах) : монография / Ю. В. Фролов, О. М. Игрунова. – Москва : Русайнс, 2021. – 247 с. – [Электронный ресурс] – URL: https://book.ru/book/945051 – Дата обращения: 12.12.2024.

36. Черткова, Е. А. Статистика. Автоматизация обработки информации : учеб. пособие / Е. А. Черткова. – 2-е изд., испр. и доп. – Москва : Изд-во Юрайт, 2023. – 195 с.

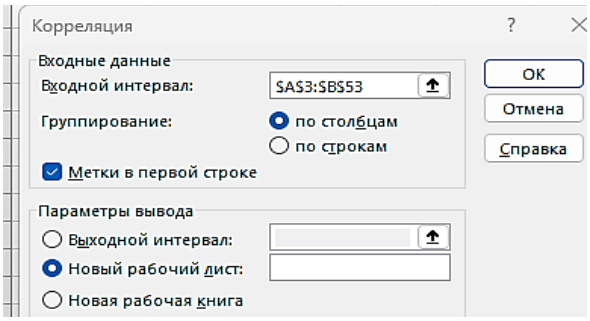
37. Яковлев, В. Б. Эконометрика в Excel и Statistica: учебное пособие / В. Б. Яковлев. – Москва : КНОРУС, 2022. – 382 с.

Приложение А
Пошаговый алгоритм выполнения корреляционно-регрессионного анализа в MS Excel

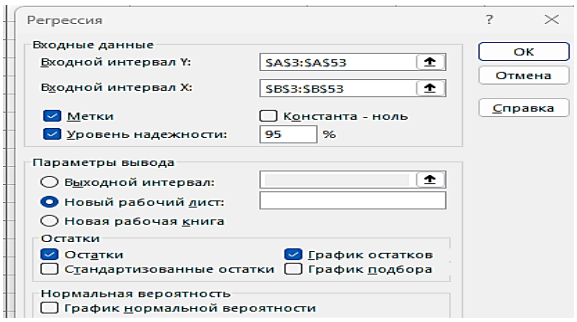




ШАГ 1.1 – Инструменты анализа данных в MS Excel



ШАГ 1.2 – Настройки инструмента «Корреляция»



ШАГ 1.3 – Настройки инструмента «Регрессия»

ВЫВОД ИТОГОВ								
Регрессионная статистика								
Множественный R	0,840310353							
R-квадрат	0,706121489							
Нормированный R-квадрат	0,69999902							
Стандартная ошибка	52,33184036							
Наблюдения	50							
Дисперсионный анализ								
	df	*SS*	*MS*	*F*	*Значимость F*			
Регрессия	1	315852,8873	315852,8873	115,3327999	2,32863E-14			
Остаток	48	131453,8327	2738,621516					
Итого	49	447306,72						
	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*	*Нижние 95,0%*	*Верхние 95,0%*
Y-пересечение	4059,710113	348,5990778	11,64578558	1,37181E-15	3358,804691	4760,616	3358,805	4760,616
X1-Цена	-245,4019219	22,85080699	-10,73931096	2,32863E-14	-291,3465487	-199,457	-291,347	-199,457

ШАГ 1.4 – Итоги регрессионного анализа в MS Excel

ВЫВОД ОСТАТКА		
Наблюдение	*Предсказанное Y*	*Остатки*
1	157,8195546	-31,81955461
2	182,3597468	-52,3597468
3	157,8195546	-22,81955461
4	231,4401312	-94,44013118
5	206,899939	-65,89993899
6	182,3597468	-36,3597468
7	206,899939	-49,89993899
8	231,4401312	-65,44013118
9	255,9803234	-71,98032337
10	206,899939	-15,89993899

ШАГ 1.5 – Вывод остатка

ШАГ 1.6 – График остатков

Рисунок А. 1– Пошаговый алгоритм выполнения линейного корреляционно-регрессионного анализа в программе MS Excel

Приложение Б
Показатели регрессионной статистики в программе MS Excel

Таблица Б. 1 – Название показателей линейного корреляционно-регрессионного анализа по разделам в программе MS Excel

Наименование показателя и величины в MS Excel	Принятые наименования показателей в статистике
***Раздел** «Регрессионная статистика»*	
Множественный R	Коэффициент (индекс) множественной корреляции
R-квадрат	Коэффициент (индекс) детерминации
Нормированный R-квадрат	Скорректированный коэффициент (индекс) детерминации (используется для множественной регрессии)
Стандартная ошибка	Стандартная ошибка уравнения регрессии (или среднеквадратическое отклонение от модели)
***Раздел** «Дисперсионный анализ»*	
Df (регрессия)	Число объясняющих переменных
Df (остаток)	Число степеней свободы
Df (итого)	Размер выборки минус 1
SS	Сумма квадратов отклонений
MS	Дисперсия на одну степень свободы вариации
F	Фактическое значение *F-критерия* для оценки связи между объясняемой и объясняющими переменными
Значимость F	Уровень значимости *F-критерия*
Коэффициенты	Параметры регрессии
Коэффициенты (Y-пересечение*)*	Оценка параметра a_0
Стандартная ошибка	Стандартные ошибки параметров уравнения регрессии
Стандартная ошибка (Y-пересечение*)*	Среднеквадратичная ошибка оцененного параметра a_0
t-статистика	Фактические значения *t*-критерия Стьюдента
t-статистика (Y-пересечение*)*	Отношение *Коэффициент* / *Стандартная ошибка* для параметра a_0
p-Значение	Уровень значимости *t*-критерия Стьюдента
p-Значение (Y-пересечение*)*	Уровень значимости α для значений *t-статистика* для параметра a_0
Нижние 95 % *Верхние 95 %*	Границы доверительно интервала для параметров уравнения регрессии, вычисляемые при различных уровнях значимости α

Приложение В
Основы прикладного статистического анализа экономических данных

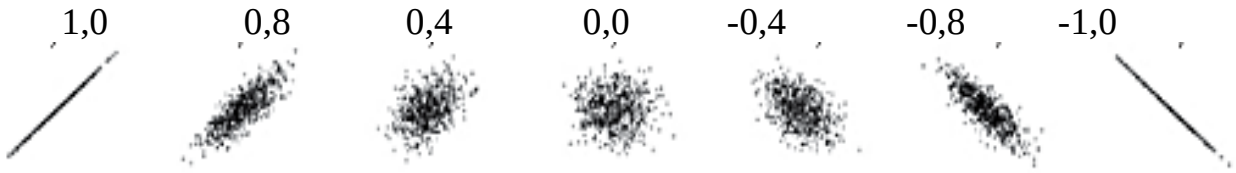


Рисунок В.1 – Поля корреляции между переменными *x* и *y*

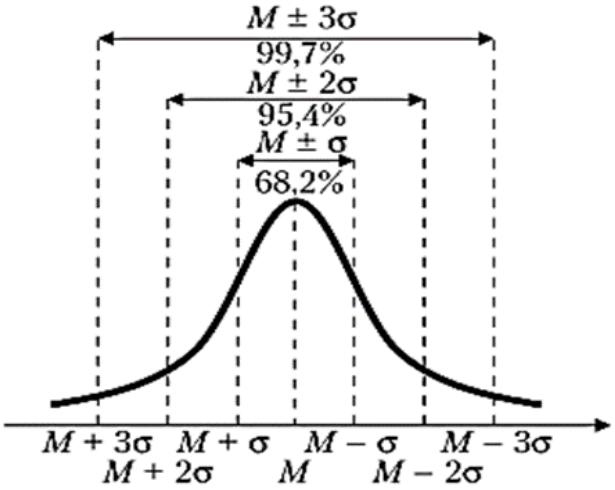


Рисунок В.2 – График нормального распределения случайной величины X

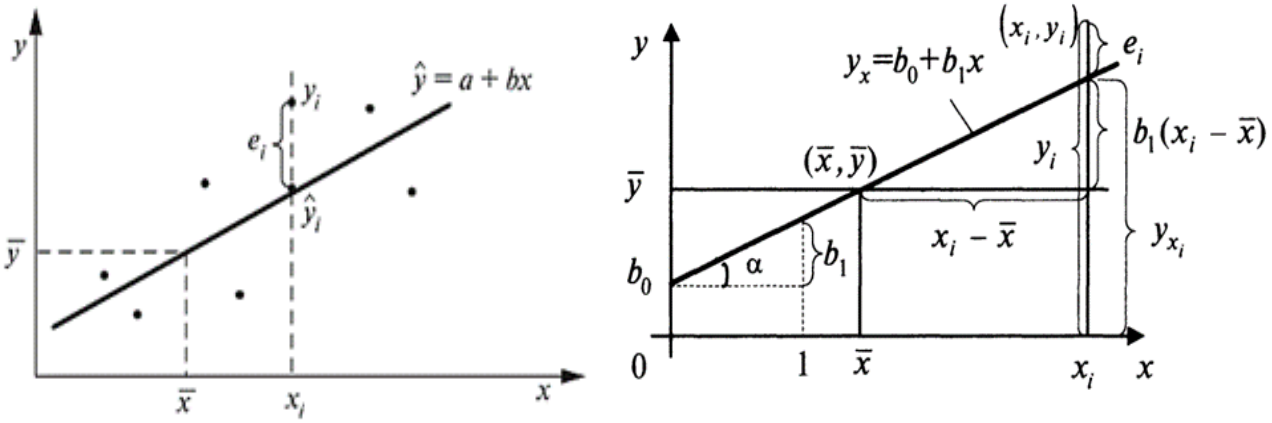


Рисунок В.3 – Диаграмма рассеяния наблюдений и линия регрессии

Приложение Г
Матричная алгебра

Матрицы – это массивы данных, имеющие прямоугольную форму. Данные располагаются в строках и столбцах. Размер матрицы выражается числом строк и столбцов, т. е. матрица с пятью строками и четырьмя столбцами будет матрицей размера 5×4. Например, следующая матрица имеет размер 3×3

$$Y = \begin{bmatrix} 1 & 4 & 6 \\ 3 & 7 & 5 \\ 8 & 6 & 3 \end{bmatrix}$$

Положение отдельных ячеек в матрице определяется сначала по позиции их строки, а затем по позиции их столбца. Таким образом, тройка слева в средней строке в приведенной выше матрице находится в ячейке 2, 1, в то время как тройка справа в нижней строке находится в ячейке 3, 3.

Векторы – это строки или столбцы данных, фактически это тоже матрицы, только с одной строчкой или одним столбцом. По правилу определения ячеек в матрицах вектор строки из n элементов будет матрицей размера $1 \times n$, а вектор столбца – матрицей размера $n \times 1$.

Часто для идентификации элемента в матрице используются двузначные подстрочные индексы. Например, элемент на пересечении i-ой строки и j-го столбца матрицы Y будет отображен как y_{ij}

Матрицы могут перемножаться, если их формы удовлетворяют особым условиям. Они могут складываться и вычитаться, только если это матрицы одной формы.

Вычисление определителя матрицы (МОПРЕД)

Определение. Любой квадратной матрице A по определенному правилу можно сопоставить некоторое число, которое называется *определителем (детерминантом)* матрицы A и обозначается $\det(A)$ или $|A|$. Если A – матрица n-го порядка ($n \geq 1$), то вместо слов «определитель матрицы A» часто говорят «определитель n-го порядка».

Пример Г.1. Вычислите определитель матрицы A: $\mathrm{A} = \begin{pmatrix} 8 & 6 \\ 7 & 4 \end{pmatrix}$.

Пусть дана матрица $\mathrm{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, тогда ее определитель второго порядка вычисляется по формуле $\Delta_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$.

Решение: определитель матрицы A составит $|\mathrm{A}| = 8 \times 4 - 6 \times 7 = -10$.

Пример Г.2. Вычислите определитель матрицы B: $\mathrm{B} = \begin{pmatrix} 5 & 5 & 1 \\ -1 & 3 & 2 \\ 4 & 0 & 2 \end{pmatrix}$.

Пусть дана матрица $\mathrm{B} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$, тогда ее определитель третьего порядка вычисляется по формуле $\Delta_3 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} +$

$+ a_{21}a_{32}a_{13} + a_{12}a_{23}a_{31} - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33}$.

Решение: определитель матрицы B составит |B| = 5 × 3 × 2 + (-1) × 0 × 1 + + 5 × 2 × 4 – 1 × 3 × 4 – 5 × 2 × 0 – 5 × (-1) × 2 = 68.

Транспонирование матрицы (ТРАНСП)

Определение. Пусть A – матрица $m \times n$. Матрица A' или A^T размера $n \times m$ называется *транспонированной* к A, если $a'_{ij} = a_{ji}$, $i = 1, 2, 3, \ldots, n$; $j = 1, 2, 3, \ldots, m$.

Очевидно, что матрица A' получается поворотом вокруг главной диагонали матрицы A. При этом повороте i-й столбец матрицы A преобразуется в i-ю строку матрицы A', $i = 1, 2, 3, n$, например:

$$\mathrm{A} = \begin{pmatrix} 2 & 3 & 8 \\ 4 & 11 & 5 \\ 7 & 20 & -5 \end{pmatrix}, \quad \mathrm{A} = \begin{pmatrix} 2 & 4 & 7 \\ 3 & 11 & 20 \\ 8 & 5 & -5 \end{pmatrix}.$$

Вычисление обратной матрицы (МОБР)

Только квадратные матрицы и к тому же только некоторые из квадратных матриц имеют обратные матрицы. Обратная матрица это такая, произведение которой на исходную матрицу равно единичной матрице. В единичной матрице каждая ячейка равна нулю за исключением главной диагонали, содержащей единицы. Например, единичная матрица 3×3 выглядит так:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Матрица, обратная матрице X, обозначается X^{-1}.

Умножение матриц (МУМНОЖ)

Определение. Операция умножения матрицы A на матрицу B определена, когда число столбцов первой матрицы равно числу строк второй. Тогда *произведением матриц* $A \times B$ называется такая матрица C, каждый элемент c_{ij} которой равен сумме произведений элементов i-й строки матрицы A на соответствующие элементы j-го столбца матрицы B.

Решение систем уравнений

Матричная алгебра может быть использована для решения систем уравнений, и именно это свойство позволяет применить ее в вычислении коэффициентов регрессии по методу наименьших квадратов.

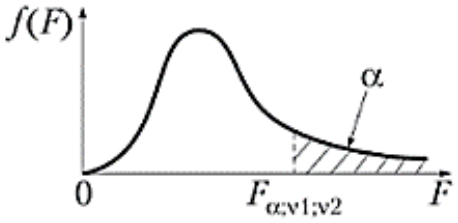
Формула для вектора коэффициентов регрессии выглядит так:

$$b = [X^T X]^{-1} X^T Y,$$

где $[X^T X]^{-1}$ – это обратная к матрице, получающаяся от умножения транспонированной из X матрицы на матрицу X. Эта обратная матрица снова умножается на транспонированную из X, а произведение умножается на вектор Y; b – вектор с компонентами $b_0, b_1, \ldots, b_k$; X – матрица значений объясняющих переменные; Y – вектор значений зависимой переменной.

Приложение Д
Распределение Фишера (*F*-распределение)

Пример:
при $\nu_1 = 6$, $\nu_2 = 5$ $P(F > 3{,}40) = 0{,}1$;
при $\nu_1 = 6$, $\nu_2 = 5$ $P(F > 4{,}95) = 0{,}05$;
при $\nu_1 = 6$, $\nu_2 = 5$ $P(F > 10{,}7) = 0{,}01$.



ν_2	α	ν_1 (число степеней свободы)											
		1	2	3	4	5	6	7	8	9	10	11	12
1	,10	39,9	49,5	53,6	55,8	57,2	58,2	58,9	59,4	59,9	60,2	60,5	60,7
	,05	161	200	216	225	230	234	237	239	241	242	243	244
2	,10	8,53	9,00	9,16	9,24	9,29	9,33	9,35	9,37	9,38	9,39	9,40	9,41
	,05	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4
	,01	98,5	99,2	99,2	99,2	99,3	99,3	99,4	99,4	99,4	99,4	99,4	99,4
3	,10	5,54	5,46	5,39	5,34	5,31	5,28	5,27	5,25	5,24	5,23	5,22	5,22
	,05	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,76	8,74
	,01	34,1	30,8	29,5	28,7	28,2	27,9	27,7	27,5	27,3	27,2	27,1	27,1
4	,10	4,54	4,32	4,19	4,11	4,05	4,01	3,98	3,95	3,94	3,92	3,91	3,90
	,05	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,94	5,91
	,01	21,2	18,0	16,7	16,0	15,5	15,2	15,0	14,8	14,7	14,5	14,4	14,4
5	,10	4,06	3,78	3,62	3,52	3,45	3,40	3,37	3,34	3,32	3,30	3,28	3,27
	,05	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,71	4,68
	,01	16,3	13,3	12,1	11,4	11,0	10,7	10,5	10,3	10,2	10,1	9,96	9,89
6	,10	3,78	3,46	3,29	3,18	3,11	3,05	3,01	2,98	2,96	2,94	2,92	2,90
	,05	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
	,01	13,7	10,9	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72
7	,10	3,59	3,26	3,07	2,96	2,88	2,83	2,78	2,75	2,72	2,70	2,68	2,67
	,05	5,59	4,74	4,35	4,12	3,97	3,87.	3,79	3,73	3,68	3,64	3,60	3,57
	,01	12,2	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,54	6,47
8	,10	3,46	3,11	2,92	2,81	2,73	2,67	2,62	2,59	2,56	2,54	2,52	2,50
	,05	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,31	3,28
	,01	11,3	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,73	5,67
9	,10	3,36	3,01	2,81	2,69	2,61	2,55	2,51	2,47	2,44	2,42	2,40	2,38
	,05	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,10	3,07
	,01	10,6	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,18	5,11
10	,10	3,29	2,92	2,73	2,61	2,52	2,46	2,41	2,38	2,35	2,32	2,30	2,28
	,05	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,91
	,01	10,0	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,77	4,71
11	,10	3,23	2,86	2,66	2,54	2,45	2,39	2,34	2,30	2,27	2,25	2,23	2,21
	,05	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,82	2,79
	,01	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,46	4,40
12	,10	3,18	2,81	2,61	2,48	2,39	2,33	2,28	2,24	2,21	2,19	2,17	2,15
	,05	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,72	2,69
	,01	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,22	4,16

ПРОДОЛЖЕНИЕ ПРИЛОЖЕНИЯ Д

ν_1 (число степеней свободы)												α	ν_2
15	20	24	30	40	50	60	100	120	200	500	∞		
61,2	61,7	62,0	62,3	62,5	62,7	62,8	63,0	63,1	63,2	63,3	63,3	,10	1
246	248	249	250	251	252	252	253	253	254	254	254	,05	
9,42	9,44	9,45	9,46	9,47	9,47	9,47	9,48	9,48	9,49	9,49	9,49	,10	2
19,4	19,4	19,5	19,5	19,5	19,5	19,5	19,5	19,5	19,5	19,5	19,5	,05	
99,4	99,4	99,5	99,5	99,5	99,5	99,5	99,5	99,5	99,5.	99,5	995	,01	
5,20	5,18	5,18	5,17	5,16	5,15	5,15	5,14	5,14	5,14	5,14	5,13	,10	3
8,70	8,66	8,64	8,62	8,59	8,58	8,57	8,55	8,55	8,54	8,53	8,53	,05	
26,9	26,7	26,6	26,5	26,4	26,4	26,3	26,2	26,2	26:2	26,1	26,1	,01	
3,87	3,84	3,83	3,82	3,80	3,80	3,79	3,78	3,78	3,77	3,76	3,76	,10	4
5,86	5,80	5,77	5,75	5,72	5,70	5,69	5,66	5,66	5,65	5,64	5,63	,05	
14,2	14,0	13,9	13,8	13,7	13,7	13,7	13,6	13,6	13,5	13,5	13,5	,01	
3,24	3,21	3,19	3,17	3,16	3,15	3,14	3,13	3,12	3,12	3,11	3,10	,10	5
4,62	4,56	4,53	4,50	4,46	4,44	4,43	4,41	4,40	4,39	4,37	4,36	,05	
9,72	9,55	9,47	9,38	9,29	9,24	9,20	9,13	9,11	9,08	9,04.	9,02	,01	
2,87	2,84	2,82	2,80	2,78	2,77	2,76	2,75	2,74	2,73	2,73	2,72	,10	6
3,94	3,87	3,84	3,81	3,77	3,75	3,74	3,71	3,70	3,69	3,68	3,67	,05	
7,56	7,40	7,31	7,23	7,14	7,09	7,06	6,99	6,97	6,93	6,90	6,88	,01	
2,63	2,59	2,58	2,56	2,54	2,52	2,51	2,50	2,49	2,48	2,48	2,47	,10	7
3,51	3,44	3,41	3,38	3,34	3,32	3,30	3,27	3,27	3,25	3,24	3,23	,05	
6,31	6,16	6,07	5,99	5,91	5,86	5,82	5,75	5,74	5,70	5,67	5,65	,01	
2,46	2,42	2,40	2,38	2,36	2,35	2,34	2,32	2,32	2,31	2,30	2,29	,10	8
3,22	3,15	3,12	3,08	3,04	2,02	3,01	2,97	2,97	2,95	2,94	2,93	,05	
5,52	5,36	5,28	5,20	5,12	5,07	5,03	4,96	4,95	4,91	4,88	4,86	,01	
2,34	2,30	2,28	2,25	2,23	2,22	2,21	2,19	2,18	2,17	2,17	2,16	,10	9
3,01	2,94	2,90	2,86	2,83	2,80	2,79	2,76	2,75	2,73	2,72	2,71	,05	
4,96	4,81	4,73	4,65	4,57	4,52	4,48	4,42	4,40	4,36	4,33	4,31	,01	
2,24	2,20	2,18	2,16	2,13	2,12	2,11	2,09	2,08	2,07	2,06	2,06	,10	10
2,85	2,77	2,74	2,70	2,66	2,64	2,62	2,59	2,58	2,56	2,55	2,54	,05	
4,56	4,41	4,33	4,25	4,17	4,12	4,08	4,01	4,00	3,96	3,93	3,91	,01	
2,17	2,12	2,10	2,08	2,05	2,04	2,03	2,00	2,00	1,99	1,98	1,97	,10	11
2,72	2,65	2,61	2,57	2,53	2,51	2,49	2,46	2,45	2,43	2,42	2,40	,05	
4,25	4,10	4,02	3,94	3,86	3,81	3,78	3,71	3,69	3,66	3,62	3,60	,01	
2,10	2,06	2,04	2,01	1,99	1,97	1,96	1,94	1,93	1,92	1,91	1,90	,10	12
2,62	2,54	2,51	2,47	2,43	2,40	2,38	2,35	2,34	2,32	2,31	2,30	,05	
4,01	3,86	3,78	3,70	3,62	3,57	3,54	3,47	3,45	3,41	3,38	3,36	,01	

ПРОДОЛЖЕНИЕ ПРИЛОЖЕНИЯ Д

ν_2	α	ν_1 (число степеней свободы)											
		1	2	3	4	5	6	7	8	9	10	11	12
13	,10	3,14	2,76	2,56	2,43	2,35	2,28	2,23	2,20	2,16	2,14	2,12	2,10
	,05	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,63	2,60
	,01	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96
14	,10	3,10	2,73	2,52	2.39	2,31	2,24	2,19	2,15	2,12	2,10	2,08	2,05
	,05	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,57	2,53
	,01	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80
15	,10	3,07	2,70	2,49	2,36	2,27	2,21	2,16	2,12	2,09	2,06	2,04	2,02
	,05	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,51	2,48
	,01	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67
16	,10	3,05	2,67	2,46	2,33	2,24	2,18	2,13	2,09	2,06	2,03	2,01	1,99
	,05	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,46	2,42
	,01	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,62	3,55
17	,10	3,03	2,64	2,44	2,31	2,22	2,15	2,10	2,06	2,03	2,00	1,98	1,96
	,05	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,41	2,38
	,01	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,46
18	,10	3,01	2,62	2,42	2,29	2,20	2,13	2,08	2,04	2,00	1,98	1,96	1,93
	,05	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34
	,01	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,43	3,37
19	,10	2,99	2,61	2,40	2,27	2,18	2,11	2,06	2,02	1,98	1,96	1,94	1,91
	,05	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,34	2,31
	,01	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,36	3,30
20	,10	2,97	2,59	2,38	2,25	2,16	2,09	2,04	2,00	1,96	1,94	1,92	1,89
	,05	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,31	2,28
	,01	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,29	3,23
22	,10	2,95	2,56	2,35	2,22	2,13	2,06	2,01	1,97	1,93	1,90	1,88	1,86
	,05	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,26	2,23
	,01	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,18	3,12
24	,10	2,93	2,54	2,33	2,19	2,10	2,04	1,98	1,94	1,91	1,88	1,85	1,83
	,05	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,21	2,18
	,01	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17	3,09	3,03
26	,10	2,91	2,52	2,31	2,17	2,08	2,01	1,96	1,92	1,88	1,86	1,84	1,81
	,05	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,18	2,15
	,01	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09	3,02	2,96
28	,10	2,89	2,50	2,29	2,16	2,06	2,00	1,94	1,90	1,87	1,84	1,81	1,79
	,05	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,15	2,12
	,01	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03	2,96	2,90
30	,10	2,88	2,49	2,28	2,14	2,05	1,98	1,93	1,88	1,85	1,82	1,79	1,77
	,05	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,13	2,09
	,01	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,91	2,84
40	,10	2,84	2,44	2,23	2,09	2,00	1,93	1,87	1,83	1,79	1,76	1,73	1,71
	,05	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,04	2,00
	,01	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80	2,73	2,66

ПРОДОЛЖЕНИЕ ПРИЛОЖЕНИЯ Д

ν_1 (число степеней свободы)												α	ν_2
15	20	24	30	40	50	60	100	120	200	500	∞		
2,05	2,01	1,98	1,96	1,93	1,92	1,90	1,88	1,88	1,86	1,85	1,85	,10	13
2,53	2,46	2,42	2,38	2,34	2,31	2,30	2,26	2,25	2,23	2,22	2,21	,05	
3,82	3,66	3,59	3,51	3,43	3,38	3,34	3,27	3,25	3,22	3,19	3,17	,01	
2,01	1,96	1,94	1,91	1,89	1,87	1,86	1,83	1,83	1,82	1,80	1,80	,10	14
2,46	2,39	2,35	2,31	2,27	2,24	2,22	2,19	2,18	2,16	2,14	2,13	,05	
3,66	3,51	3,43	3,35	3,27	3,22	3,18	3,11	3,09	3,06	3,03	3,00	,01	
1,97	1,92	1,90	1,87	1,85	1,83	1,82	1,79	1,79	1,77	1,76	1,76	,10	15
2,40	2,33	2,29	2,25	2,20	2,18	2,16	2,12	2,11	2,10	2,08	2,07	,05	
3,52	3,37	3,29	3,21	3,13	3,08	3,05	2,98	2,96	2,92	2,89	2,87	,01	
1,94	1,89	1,87	1,84	1,81	1,79	1,78	1,76	1,75	1,74	1,73	1,72	,10	16
2,35	2,28	2,24	2,19	2,15	2,12	2,11	2,07	2,06	2,04	2,02	2,01	,05	
3,41	3,26	3,18	3,10	3,02	2,97	2,93	2,86	2,84	2,81	2,78	2,75	,01	
1,91	1,86	1,84	1,81	1,78	1,76	1,75	1,73	1,72	1,71	1,69	1,69	,10	17
2,31	2,23	2,19	2,15	2,10	2,08	2,06	2,02	2,01	1,99	1,97	1,96	,05	
3,31	3,16	3,08	3,00	2,92	2,87	2,83	2,76	2,75	2,71	2,68	2,65	,01	
1,89	1,84	1,81	1,78	1,75	1,74	1,72	1,70	1,69	1,68	1,67	1,66	,10	18
2,27	2,19	2,15	2,11	2,06	2,04	2,02	1,98	1,97	1,95	1,93	1,92	,05	
3,23	3,08	3,00	2,92	2,84	2,78	2,75	2,68	2,66	2,62	2,59	2,57	,01	
1,86	1,81	1,79	1,76	1,73	1,71	1,70	1,67	1,67	1,65	1,64	1,63	,10	19
2,23	2,16	2,11	2,07	2,03	2,00	1,98	1,94	1,93	1,91	1,89	1,88	,05	
3,15	3,00	2,92	2,84	2,76	2,71	2,67	2,60	2,58	2,55	2,51	2,49	,01	
1,84	1,79	1,77	1,74	1,71	1,69	1,68	1,65	1,64	1,63	1,62	1,61	,10	20
2,20	2,12	2,08	2,04	1,99	1,97	1,95	1,91	1,90	1,88	1,86	1,84	,05	
3,09	2,94	2,86	2,78	2,69	2,64	2,61	2,54	2,52	2,48	2,44	2,42	,01	
1,81	1,76	1,73	1,70	1,67	1,65	1,64	1,61	1,60	1,39	1,58	1,37	,10	22
2,15	2,07	2,03	1,98	1,94	1,91	1,89	1,85	1,84	1,82	1,80	1,78	,05	
2,98	2,83	2,75	2,67	2,58	2,53	2,50	2,42	2,40	2,36	2,33	2,31	,01	
1,78	1,73	1,70	1,67	1,64	1,62	1,61	1,58	1,57	1,56	1,54	1,53	,10	24
2,11	2,03	1,98	1,94	1,89	1,86	1,84	1,80	1,79	1,77	1,75	1,73	,05	
2,89	2,74	2,66	2,58	2,49	2,44	2,40	2,33	2,31	2,27	2,24	2,21	,01	
1,76	1,71	1,68	1,65	1,61	1,59	1,58	1,35	1,54	1,53	1,51	1,50	,10	26
2,07	1,99	1,95	1,90	1,85	1,82	1,80	1,76	1,75	1,73	1,71	1,69	,05	
2,81	2,66	2,58	2,50	2,42	2,36	2,33	2,25	2,23	2,19	2,16	2,13	,01	
1,74	1,69	1,66	1,63	1,59	1,57	1,56	1,53	1,52	1,50	1,49	1,48	,10	28
2,04	1,96	1,91	1,87	1,82	1,79	1,77	1,73	1,71	1,69	1,67	1,65	,05	
2,75	2,60	2,52	2,44	2,35	2,30	2,26	2,19	2,17	2,13	2,09	2,06	,01	
1,72	1,67	1,64	1,61	1,57	1,55	1,54	1,51	1,50	1,48	1,47	1,46	,10	30
2,01	1,93	1,89	1,84	1,79	1,76	1,74	1,70	1,68	1,66	1,64	1,62	,05	
2,70	2,55	2,47	2,39	2,30	2,25	2,21	2,13	2,11	2,07	2,03	2,01	,01	
1,66	1,61	1,57	1,54	1,51	1,48	1,47	1,43	1,42	1,41	1,39	1,38	,10	40
1,92	1,84	1,79	1,74	1,69	1,66	1,64	1,59	1,58	1,55	1,53	1,51	,05	
2,52	2,37	2,29	2,20	2,11	2,06	2,02	1,94	1,92	1,87	1,83	1,80	,01	

ОКОНЧАНИЕ ПРИЛОЖЕНИЯ Д

ν_2	α	ν_1 (число степеней свободы)												
		1	2	3	4	5	6	7	8	9	10	11	12	
60	,10	2,79	2,39	2,18	2,04	1,95	1,87	1,82	1,77	1,74	1,71	1,68	1,66	
	,05	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,95	1,92	
	,01	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,56	2,50	
80	,01	2,77	2,37	2,16	2,02	1,93	1,85	1,80	1,75	1,72	1,69	1,65	1,63	
	,05	3,96	3,11	2,72	2,48	2,33	2,21	2,12	2,05	1,99	1,95	1,91	1,88	
	,01	6,96	4,88	4,04	3,56	3,25	3,04	2,87	2,74	2,64	2,55	2,48	2,41	
100	,10	2,76	2,36	2,14	2,00	1,91	1,83	1,78	1,73	1,70	1,67	1,63	1,61	
	,05	3,94	3,09	2,70	2,46	2,30	2,19	2,10	2,03	1,97	1,92	1,88	1,85	
	,01	6,90	4,82	3,98	3,51	3,20	2,99	2,82	2,69	2,59	2,51	2,43	2,36	
120	,10	2,75	2,35	2,13	1,99	1,90	1,82	1,77	1,72	1,68	1,65	1,62	1,60	
	,05	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96	1,91	1,87	1,83	
	,01	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56	2,47	2,40	2,34	
200	,10	2,73	2,33	2,11	1,97	1,88	1,80	1,75	1,70	1,66	1,63	1,60	1,57	
	,05	3,89	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93	1,88	1,84	1,80	
	,01	6,76	4,71	3,88	3,41	3,11	2,89	2,73	2,60	2,50	2,41	2,34	2,27	
∞	,10	2,71	2,30	2,08	1,94	1,85	1,77	1,72	1,67	1,63	1,60	1,57	1,55	
	,05	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,79	1,75	
	,01	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32	2,25	2,18	

	ν_1 (число степеней свободы)												α	ν_2
	15	20	24	30	40	50	60	100	120	200	500	∞		
	1,60	1,54	1,51	1,48	1,44	1,41	1,40	1,36	1,35	1,33	1,31	1,29	,10	60
	1,84	1,75	1,70	1,65	1,59	1,56	1,53	1,48	1,47	1,44	1,41	1,39	,05	
	2,35	2,20	2,12	2,03	1,94	1,88	1,84	1,75	1,73	1,68	1,63	1,60	,01	
	1,58	1,52	1,49	1,45	1,41	1,38	1,36	1,31	1,31	1,29	1,27	1,25	,10	80
	1,77	1,70	1,65	1,60	1,54	1,51	1,47	1,42	1,40	1,38	1,34	1,32	,05	
	2,24	2,11	2,03	1,94	1,84	1,78	1,76	1,65	1,63	1,57	1,52	1,49	,01	
	1,56	1,50	1,47	1,43	1,39	1,36	1,34	1,29	1,28	1,26	1,24	1,22	,10	100
	1,75	1,68	1,63	1,57	1,51	1,48	1,45	1,39	1,38	1,34	1,30	1,28	,05	
	2,19	2,06	1,98	1,89	1,79	1,73	1,70	1,59	1,57	1,51	1,46	1,43	,01	
	1,55	1,48	1,45	1,41	1,37	1,34	1,32	1,27	1,26	1,24	1,21	1,19	,10	120
	1,75	1,66	1,61	1,55	1,50	1,46	1,43	1,37	1,35	1,32	1,28	1,25	,05	
	2,19	2,03	1,95	1,86	1,76	1,70	1,66	1,56	1,53	1,48	1,42	1,38	,01	
	1,52	1,46	1,42	1,38	1,34	1,31	1,28	1,24	1,22	1,20	1,17	1,14	,10	200
	1,72	1,62	1,57	1,52	1,46	1,41	1,39	1,32	1,29	1,26	1,22	1,19	,05	
	2,13	1,97	1,89	1,79	1,69	1,63	1,58	1,48	1,44	1,39	1,33	1,28	,01	
	1,49	1,42	1,38	1,34	1,30	1,26	1,24	1,18	1,17	1,13	1,08	1,00	,10	∞
	1,67	1,57	1,52	1,46	1,39	1,35	1,32	1,24	1,22	1,17	1,11	1,00	,05	
	2,04	1,88	1,79	1,70	1,59	1,52	1,47	1,36	1,32	1,25	1,15	1,00	,01	

Приложение Е
Распределение Дарбина – Уотсона (*Durbin – Watson)*

Для ответа на вопрос: какие значения *DW* можно считать статистически близкими к двум, разработаны специальные таблицы критических точек статистики Дарбина – Уотсона (прилож. Е), позволяющие при данном числе наблюдений n, количестве объясняющих переменных m и заданном уровне значимости α определять границы приемлемости (критические точки) наблюдаемой статистики *DW*.

Критические точки d_1 и d_u при уровне значимости α = 0,05 (n – объем выборки, m – число объясняющих переменных в уравнении регрессии).

n	$m = 1$		$m = 2$		$m = 3$		$m = 4$	
	d_1	d_u	d_1	d_u	d_1	d_u	d_1	d_u
6	0,610	1,400						
7	0,700	1,356	0,467	1,896				
8	0,763	1,332	0,359	1,777	0,368	2,287		
9	0,824	1,320	0,629	1,699	0,435	2,128	0,296	2,388
10	0,879	1,320	0,697	1,641	0,525	2,016	0,376	2,414
11	0,927	1,324	0,658	1,604	0,595	1,928	0,444	2,283
12	0,971	1,331	0,812	1,579	0,658	1,864	0,512	2,177
13	1,010	1,340	0,861	1,562	0,715	1,816	0,574	2,094

	$m = 5$		$m = 6$		$m = 7$		$m = 8$		$m = 9$	
	d_1	d_u	d_1	d_u	d_1	d_u	d_1	d_u	d_1	d_u
	0,243	2,822								
	0,316	2,645	0,203	3,005						
	0,379	2,506	0,268	2,832	0,171	3,149				
	0,445	2,390	0,328	2,692	0,230	2,985	0,147	3.266		

ПРОДОЛЖЕНИЕ ПРИЛОЖЕНИЯ Е

n	$m = 1$		$m = 2$		$m = 3$		$m = 4$	
	d_1	d_u	d_1	d_u	d_1	d_u	d_1	d_u
14	1,045	1,330	0,905	1,551	0,767	1,779	0,632	2,030
15	1,077	1,361	0,946	1,543	0,814	1,750	0,685	1,977
16	1,106	1,371	0,982	1,539	0,857	1,728	0,734	1,935
17	1,133	1,381	1,015	1,536	0,897	1,710	0,779	1,900
18	1,158	1,391	1,046	1,535	0,933	1,696	0,820	1,872
19	1,180	1,401	1,074	1,536	0,967	1,685	0,859	1,848
20	1,201	1,411	1,100	1,537	0,998	1,676	0,894	1,828
21	1,221	1,420	1,125	1,538	1,026	1,669	0,927	1,812
22	1,239	1,429	1,147	1,541	1,053	1,664	0,958	1,797
23	1,257	1,437	1,168	1,543	1,078	1,660	0,986	1,785
24	1,273	1,446	1,188	1,546	1,101	1,656	1,013	1,775
25	1,288	1,454	1,206	1,550	1,123	1,654	1,038	1,767
26	1,302	1,461	1,224	1,553	1,143	1,652	1,062	1,759
27	1,316	1,469	1,240	1,556	1,162	1,651	1,084	1,753
28	1,328	1,476	1,255	1,560	1,181	1,650	1,104	1,747
29	1,341	1,483	1,270	1,563	1,198	1,650	1,124	1,743
30	1,352	1,489	1,284	1,567	1,214	1,650	1,143	1,739
31	1,363	1,496	1,297	1,570	1,229	1,650	1,160	1,735
32	1,373	1,502	1,309	1,574	1,244	1,650	1,177	1,732
33	1,383	1,508	1,321	1,577	1,258	1,651	1,193	1,730
34	1,393	1,514	1,333	1,580	1,271	1,652	1,208	1,728
35	1,402	1,519	1,343	1,584	1,283	1,653	1,222	1,726
36	1,411	1,525	1,354	1,587	1,295	1,654	1,236	1,724
37	1,419	1,530	1,364	1,590	1,307	1,655	1,249	1,723
38	1,427	1,535	1,373	1,594	1,318	1,656	1,261	1,722
39	1,435	1,540	1,382	1,597	1,328	1,658	1,273	1,722
40	1,442	1,544	1,391	1,600	1,338	1,659	1,285	1,721
45	1,475	1,566	1,430	1,615	1,383	1,666	1,336	1,720
50	1,503	1,585	1,462	1,628	1,421	1,674	1,378	1,721
55	1,528	1,601	1,490	1,641	1,452	1,681	1,414	1,724
60	1,549	1,616	1,514	1,652	1,480	1,689	1,444	1,727
65	1,567	1,629	1,536	1,662	1,503	1,696	1,471	1,731
70	1,583	1,641	1,554	1,672	1,525	1,703	1,494	1,735
75	1,598	1,65	1,571	1,680	1,543	1,709	1,515	1,739
80	1,611	1,662	1,586	1,688	1,560	1,715	1,534	1,743
85	1,624	1,671	1,600	1,696	1,575	1,721	1,550	1,747
90	1,635	1,679	1,612	1,703	1,589	1,726	1,566	1,751
95	1,645	1,687	1,623	1,709	1,602	1,732	1,579	1,755
100	1,654	1,694	1,634	1,715	1,613	1,736	1,592	1,758
150	1,720	1,746	1,706	1,760	1,693	1,774	1,679	1,788
200	1,758	1,778	1,748	1,789	1,738	1,799	1,728	1,810

ПРОДОЛЖЕНИЕ ПРИЛОЖЕНИЯ Е

	$m = 5$		$m = 6$		$m = 7$		$m = 8$		$m = 9$	
	d_1	d_u	d_1	d_u	d_1	d_u	d_1	d_u	d_1	d_u
	0,505	2,296	0,389	2,572	0,286	2,848	0,200	3,111	0,127	3,360
	0,562	2,220	0,447	2,472	0,343	2,727	0,251	2,979	0,175	3,216
	0,615	2,157	0,502	2,388	0,398	2,624	0,304	2,860	0,222	3,090
	0,664	2,104	0,554	2,318	0,451	1.537	0,356	2,757	0,272	2,975
	0,710	2,060	0,603	2,257	0,502	2,461	0,407	2,667	0,321	2,873
	0,752	2,023	0,649	2,206	0,549	2,396	0,456	2,589	0,369	2,783
	0,792	1,991	0,692	2,162	0,595	2,339	0,502	2,521	0,416	2,704
	0,829	1,964	0,732	2,124	0,637	2,290	0,547	2,460	0,461	2,633
	0,863	1,940	0,769	2,090	0,677	2,246	0,588	2,407	0,504	2,571
	0,895	1,920	0,804	2,061	0,715	2,208	0,628	2,360	0,545	2,514
	0,925	1,902	0,837	2,035	0,751	2,174	0,666	2,318	0,584	2,464
	0,953	1,886	0,868	2,012	0,784	2,144	0,702	2,280	0,621	2,419
	0,979	1,873	0,897	1,992	0,816	2,117	0,735	2,246	0,657	2,379
	1,004	1,861	0,925	1,974	0,845	2,093	0,767	2,216	0,691	2,342
	1,028	1,850	0,951	1,958	0,874	2,071	0,798	2,188	0,723	2,309
	1,050	1,841	0,975	1,944	0,900	2,052	0,826	2,164	0,753	2,278
	1,071	1,833	0,998	1,931	0,926	2,034	0,854	2,141	0,782	2,251
	1,090	1,825	1,020	1,920	0,950	2,018	0,879	2,120	0,810	2,226
	1,109	1,819	1,041	1,909	0,972	2,004	0,904	2,102	0,836	2,203
	1,127	1,813	1,061	1,900	0,994	1,991	0,927	2,085	0,861	2,181
	1,144	1,808	1,080	1,891	1,015	1,979	0,950	2,069	0,885	2,162
	1,160	1,803	1,097	1,884	1,034	1,967	0,971	2,054	0,908	2,144
	1,175	1,799	1,114	1,877	1,053	1,957	0,991	2,041	0,930	2.127
	1,190	1,795	1,131	1,870	1,071	1,948	1,011	2,029	0,951	2,112
	1,204	1,792	1,146	1,864	1,088	1,939	1,029	2,017	0,970	2,098
	1,218	1,789	1,161	1,859	1,104	1,932	1,047	2,007	0,990	2,085
	1,230	1,786	1,175	1,854	1,120	1,924	1,064	1,997	1,008	2,072
	1,287	1,776	1,238	1,835	1,189	1,895	1,139	1,958	1,089	2,022
	1,335	1,771	1,291	1,822	1,246	1,875	1,201	1,930	1,156	1,986
	1,374	1,768	1,334	1,814	1,294	1,861	1,253	1,909	1,212	1,959
	1,408	1,767	1,372	1,808	1,335	1,850	1,298	1,894	1,260	1,939
	1,438	1,767	1,404	1,805	1,370	1,843	1,336	1,882	1,301	1,923
	1,464	1,768	1,433	1,802	1,401	1,837	1,369	1,873	1,337	1,910
	1,487	1,770	1,458	1,801	1,428	1,834	1,399	1,867	1,369	1,901
	1,507	1,772	1,480	1,801	1,453	1,831	1,425	1,861	1,397	1,893
	1,525	1,774	1,500	1,801	1,474	1,829	1,448	1,857	1,422	1,886
	1,542	1,776	1,518	1,801	1,494	1,827	1,469	1,854	1,445	1,881
	1,557	1,778	1,535	1,802	1,512	1,827	1,489	1,852	1,465	1,877
	1,571	1,780	1,550	1,803	1,528	1,826	1,506	1,850	1,484	1,874
	1,665	1,802	1,651	1,817	1,637	1,832	1,622	1,847	1,608	1,862
	1,718	1,820	1,707	1,831	1,697	1,841	1,686	1,852	1,675	1,863

ПРОДОЛЖЕНИЕ ПРИЛОЖЕНИЯ Е

Критические точки d_1 и d_u при уровне значимости α = 0,01 (n – объем выборки, m – число объясняющих переменных в уравнении регрессии).

n	$m = 1$		$m = 2$		$m = 3$		$m = 4$	
	d_1	d_u	d_1	d_u	d_1	d_u	d_1	d_u
6	0,390	1,142						
7	0,433	1,036	0,294	1,676				
8	0,497	1,003	0,343	1,489	0,229	2,102		
9	0,554	0,998	0,408	1,389	0,279	1,873	0,183	2,433
10	0,604	1,001	0,466	1,333	0,340	1,733	0,230	2,193
11	0,633	1,010	0,319	1,297	0,396	1,640	0,286	2,030
12	0,697	1,023	0,369	1,274	0,449	1,373	0,339	1,913
13	0,738	1,038	0,616	1,261	0,499	1,326	0,391	1,826
14	0,776	1,034	0,660	1,234	0,347	1,490	0,441	1,737
15	0,811	1,070	0,700	1,232	0,391	1,464	0,488	1,704
16	0,844	1,086	0,737	1,232	0,633	1,446	0,332	1,663
17	0,874	1,102	0,772	1,233	0,672	1,432	0,374	1,630
18	0,902	1,118	0,803	1,239	0,708	1,422	0,613	1,604
19	0,928	1,132	0,833	1,263	0,742	1,413	0,630	1,384
20	0,932	1,147	0,863	1,271	0,773	1,411	0,683	1,367
21	0,973	1,161	0,890	1,277	0,803	1,408	0,718	1,334
22	0,997	1,174	0,914	1,284	0,831	1,407	0,748	1,343
23	1,018	1,187	0,938	1,291	0,838	1,407	0,777	1,334
24	1,037	1,199	0,960	1,298	0,882	1,407	0,803	1,328
25	1,033	1,211	0,981	1,303	0,906	1,409	0,831	1,323
26	1,072	1,222	1,001	1,312	0,928	1,411	0,833	1,318
27	1,089	1,233	1,019	1,319	0,949	1,413	0,878	1,313
28	1,104	1,244	1,037	1,323	0,969	1,413	0,900	1,313
29	1,119	1,234	1,034	1,332	0,988	1,418	0,921	1,312
30	1,133	1,263	1,070	1,339	1,006	1,421	0,941	1,311
31	1,147	1,273	1,083	1,343	1,023	1,423	0,960	1,310
32	1,160	1,282	1,100	1,332	1,040	1,428	0,979	1,310
33	1,172	1,291	1,114	1,338	1,033	1,432	0,996	1,310
34	1,184	1,299	1,128	1,364	1,070	1,433	1,012	1,311
35	1,193	1,307	1,140	1,370	1,083	1,439	1,028	1,312
36	1,206	1,313	1,133	1,376	1,098	1,442	1,043	1,313
37	1,217	1,323	1,163	1,382	1,112	1,446	1,038	1,314
38	1,227	1,330	1,176	1,388	1,124	1,449	1,072	1,313
39	1,237	1,337	1,187	1,393	1,137	1,433	1,083	1,317
40	1,246	1,344	1,198	1,398	1,148	1,437	1,098	1,318
45	1,288	1,376	1,243	1,423	1,201	1,474	1,136	1,328
50	1,324	1,403	1,283	1,446	1,243	1,491	1,203	1,338
55	1,336	1,427	1,320	1,466	1,284	1,306	1,247	1,348
60	1,383	1,449	1,330	1,484	1,317	1,320	1,283	1,338
65	1,407	1,468	1,377	1,300	1,346	1,334	1,313	1,368
70	1,429	1,483	1,400	1,313	1,372	1,346	1,343	1,378

ПРОДОЛЖЕНИЕ ПРИЛОЖЕНИЯ Е

	$m = 5$		$m = 6$		$m = 7$		$m = 8$		$m = 9$	
	d_1	d_u	d_1	d_u	d_1	d_u	d_1	d_u	d_1	d_u
	0,130	2,690								
	0,193	2,433	0,124	2,892						
	0,244	2,280	0,164	2,663	0,103	3,033				
	0,294	2,130	0,211	2,490	0,140	2,838	0,090	3,182		
	0,343	2,049	0,237	2,334	0,183	2,667	0,122	2,981	0,078	3,287
	0,391	1,967	0,303	2,244	0,226	2,330	0,161	2,817	0,107	3,101
	0,437	1,900	0,349	2,133	0,269	2,416	0,200	2,681	0,142	2,944
	0,480	1,847	0,393	2,078	0,313	2,319	0,241	2,366	0,179	2,811
	0,322	1,803	0,433	2,013	0,333	2,238	0,282	2,467	0,216	2,697
	0,361	1,767	0,476	1,963	0,396	2,169	0,322	2,381	0,233	2,397
	0,398	1,737	0,313	1,918	0,436	2,110	0,362	2,308	0,294	2,310
	0,633	1,712	0,332	1,881	0,474	2,039	0,400	2,244	0,331	2,434
	0,667	1,691	0,387	1,849	0,310	2,013	0,437	2,188	0,368	2,367
	0,698	1,673	0,620	1,821	0,343	1,977	0,473	2,140	0,404	2,308
	0,728	1,638	0,632	1,797	0,378	1,944	0,307	2,097	0,439	2,233
	0,736	1,643	0,682	1,776	0,610	1,913	0,340	2,039	0,473	2,209
	0,783	1,633	0,711	1,739	0,640	1,889	0,372	2,026	0,303	2,168
	0,808	1,626	0,738	1,743	0,669	1,867	0,602	1,997	0,336	2,131
	0,832	1,618	0,764	1,729	0,696	1,847	0,630	1,970	0,366	2,098
	0,833	1,611	0,788	1,718	0,723	1,830	0,638	1,947	0,393	2,068
	0,877	1,606	0,812	1,707	0,748	1,814	0,684	1,923	0,622	2,041
	0,897	1,601	0,834	1,698	0,772	1,800	0,710	1,906	0,649	2,017
	0,917	1,397	0,836	1,690	0,794	1,788	0,734	1,889	0,674	1,993
	0,936	1,394	0,876	1,683	0,816	1,776	0,737	1,874	0,698	1,973
	0,934	1,391	0,896	1,677	0,837	1,766	0,779	1,860	0,722	1,937
	0,971	1,389	0,914	1,671	0,837	1,737	0,800	1,847	0,744	1,940
	0,988	1,388	0,932	1,666	0,877	1,749	0,821	1,836	0,766	1,923
	1,004	1,386	0,930	1,662	0,893	1,742	0,841	1,823	0,787	1,911
	1,019	1,383	0,966	1,638	0,913	1,733	0,860	1,816	0,807	1,899
	1,034	1,384	0,982	1,633	0,930	1,729	0,878	1,807	0,826	1,887
	1,048	1,384	0,997	1,632	0,946	1,724	0,893	1,799	0,844	1,876
	1,111	1,384	1,063	1,643	1,019	1,704	0,974	1,768	0,927	1,834
	1,164	1,387	1,123	1,639	1,081	1,692	1,039	1,748	0,997	1,803
	1,209	1,392	1,172	1,638	1,134	1,683	1,093	1,734	1,037	1,783
	1,249	1,398	1,214	1,639	1,179	1,682	1,144	1,726	1,108	1,771
	1,283	1,604	1,231	1,642	1,218	1,680	1,186	1,720	1,133	1,761
	1,313	1,611	1,283	1,643	1,233	1,680	1,223	1,716	1,192	1,734

ОКОНЧАНИЕ ПРИЛОЖЕНИЯ Е

n	$m = 1$		$m = 2$		$m = 3$		$m = 4$	
	d_1	d_u	d_1	d_u	d_1	d_u	d_1	d_u
75	1,448	1,301	1,422	1,329	1,393	1,337	1,368	1,387
80	1,466	1,313	1,441	1,341	1,416	1,368	1,390	1,393
85	1,482	1,328	1,438	1,333	1,433	1,378	1,411	1,603
90	1,496	1,340	1,474	1,363	1,432	1,387	1,429	1,611
95	1,310	1,332	1,489	1,373	1,468	1,396	1,446	1,618
100	1,322	1,362	1,303	1,383	1,482	1,604	1,462	1,623
150	1,611	1,637	1,398	1,631	1,384	1,663	1,371	1,679
200	1,664	1,684	1,633	1,693	1,643	1,704	1,633	1,713

	$m = 5$		$m = 6$		$m = 7$		$m = 8$		$m = 9$	
	d_1	d_u	d_1	d_u	d_1	d_u	d_1	d_u	d_1	d_u
	1,340	1,617	1,313	1,649	1,284	1,682	1,236	1,714	1,227	1,748
	1,364	1,624	1,338	1,633	1,312	1,683	1,283	1,714	1,239	1,743
	1,386	1,630	1,362	1,637	1,337	1,683	1,312	1,714	1,287	1,743
	1,406	1,636	1,383	1,661	1,360	1,687	1,336	1,714	1,312	1,741
	1,423	1,642	1,403	1,666	1,381	1,690	1,338	1,713	1,336	1,741
	1,441	1,647	1,421	1,670	1,400	1,693	1,378	1,717	1,337	1,741
	1,337	1,693	1,343	1,708	1,330	1,722	1,313	1,737	1,301	1,732
	1,623	1,723	1,613	1,733	1,603	1,746	1,392	1,737	1,382	1,768

Для заданных α, n, m в таблице (прил. Е) указываются два числа: d_1 – нижняя граница и d_u – верхняя граница. Для проверки гипотезы об отсутствии автокорреляции остатков используется числовой отрезок, изображенный на рисунке Е. 1.

Выводы осуществляются но следующей схеме.

Если $DW < d_1$, то автокорреляция остатков положительна.

Если $DW > 4 - d_1$, то автокорреляция остатков отрицательна.

При $d_u < DW < 4 - d_u$ гипотеза об отсутствии автокорреляции остатков принимается.

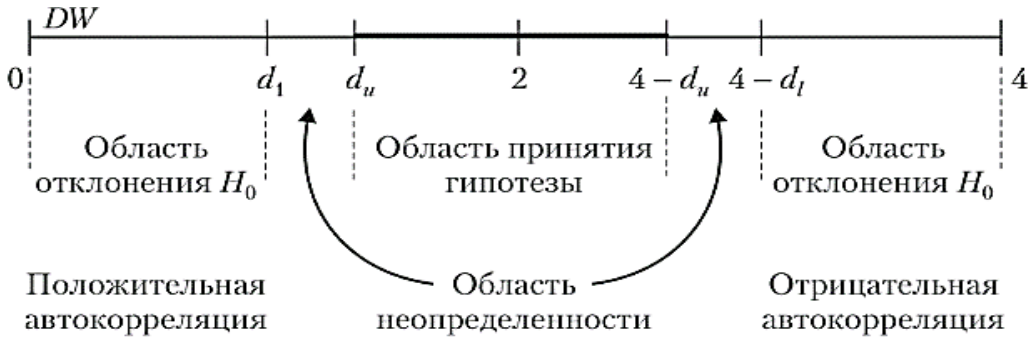


Рисунок Е. 1 – Числовой отрезок Дарбина – Уотсона DW для проверки гипотезы об отсутствии автокорреляции остатков

Если критерий Дарбина – Уотсона не приводит к успешному решению проблемы автокорреляции используют процедуру Оркулта – Кокроуна.

Приложение Ж
Исходные данные организаций для выполнения прикладного статистического анализа

Таблица Ж. 1 – Исходные данные использования основных средств организаций для выполнения прикладного статистического анализа

№	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13
1	0,01573	0,253	0,464	1,698	0,763	0,940	0,544	0,967	0,434	0,010	0,008	0,107	1,149	0,697
2	0,00014	0,169	0,322	2,261	0,728	0,980	0,525	0,935	0,439	0,006	0,005	0,039	1,232	0,999
3	0,02727	0,362	0,643	2,087	0,756	0,898	0,563	0,931	0,396	0,028	0,010	0,014	1,191	0,728
4	0,11356	1,355	2,608	0,454	0,615	0,741	0,519	0,944	0,449	0,068	0,046	0,010	1,742	0,608
5	0,09679	0,609	0,885	1,255	0,763	0,990	0,689	0,973	0,292	0,019	0,012	0,009	2,278	0,339
6	0,06789	1,447	2,622	0,439	0,631	0,728	0,552	0,943	0,415	0,021	0,007	0,006	1,405	0,547
7	0,02681	0,428	0,657	1,924	0,594	0,751	0,657	0,624	0,341	0,042	0,025	0,016	1,381	0,718
8	0,15355	1,699	2,908	0,403	0,669	0,683	0,584	0,973	0,399	0,085	0,043	0,038	1,734	0,102
9	0,23133	1,826	3,278	0,381	0,668	0,730	0,557	0,922	0,395	0,081	0,068	0,021	1,391	0,748
10	0,35100	1,718	3,145	0,365	0,613	0,685	0,546	0,954	0,428	0,066	0,06	0,011	1,319	0,571
11	0,04102	0,441	0,662	1,173	0,628	0,785	0,573	0,725	0,325	0,051	0,029	0,018	1,286	0,671
12	0,01734	0,138	0,260	3,421	0,702	0,881	0,551	0,741	0,341	0,032	0,014	0,011	1,374	0,724
13	0,03812	0,357	0,587	0,833	0,714	0,757	0,686	0,732	0,339	0,027	0,020	0,009	1,293	0,737
14	0,03184	0,514	0,618	1,538	0,738	0,891	0,602	0,699	0,317	0,025	0,018	0,021	1,156	0,619
15	0,03967	0,658	0,732	0,796	0,654	0,745	0,637	0,679	0,405	0,030	0,019	0,017	1,301	0,586
16	0,02408	0,367	0,557	1,001	0,653	0,908	0,574	0,695	0,394	0,011	0,007	0,015	1,112	0,708
17	0,06927	0,538	0,680	0,785	0,756	0,704	0,632	0,688	0,321	0,072	0,041	0,022	1,247	0,694
18	0,02211	0,123	0,236	1,492	0,612	0,727	0,521	0,741	0,443	0,009	0,006	0,013	1,081	0,681
19	0,02347	0,290	0,481	1,668	0,739	0,912	0,56	0,961	0,418	0,033	0,017	0,112	1,338	0,674
20	0,00018	0,193	0,375	2,623	0,714	0,956	0,515	0,932	0,447	0,030	0,027	0,082	1,420	0,999
21	0,03544	0,438	0,797	1,713	0,750	0,892	0,549	0,925	0,406	0,050	0,028	0,056	1,547	0,703
22	0,09924	0,854	1,621	0,734	0,627	0,753	0,527	0,945	0,443	0,048	0,039	0,018	1,503	0,589
23	-0,00613	0,282	0,412	2,672	0,749	0,990	0,684	0,972	0,296	0,027	0,015	0,036	0,176	0,395
24	0,03197	1,261	2,333	0,495	0,614	0,706	0,541	0,934	0,421	0,027	0,014	0,016	1,584	0,518
25	0,03806	0,494	0,671	1,709	0,583	0,738	0,675	0,637	0,346	0,034	0,022	0,017	1,291	0,702
26	0,29284	1,839	2,836	0,318	0,497	0,601	0,648	0,975	0,344	0,415	0,231	0,016	2,287	0,062
27	0,12724	1,632	2,977	0,407	0,655	0,721	0,548	0,917	0,402	0,046	0,029	0,022	1,624	0,675
28	0,28017	1,572	2,843	0,401	0,618	0,697	0,550	0,956	0,427	0,032	0,028	0,018	1,127	0,629
29	0,03521	0,436	0,627	1,184	0,647	0,804	0,557	0,703	0,347	0,048	0,022	0,027	1,118	0,648
30	0,00963	0,129	0,104	3,507	0,731	0,893	0,539	0,718	0,351	0,051	0,019	0,031	1,239	0,701
31	0,03244	0,342	0,305	0,871	0,724	0,774	0,661	0,695	0,348	0,034	0,012	0,024	1,161	0,695
32	0,02791	0,528	0,532	1,302	0,714	0,863	0,615	0,712	0,307	0,037	0,024	0,018	1,308	0,587
33	0,03287	0,643	0,616	0,864	0,671	0,751	0,608	0,653	0,424	0,026	0,011	0,015	1,201	0,553
34	0,04617	0,371	0,391	0,892	0,618	0,897	0,597	0,721	0,378	0,043	0,021	0,022	1,101	0,691
35	0,05306	0,526	0,495	0,803	0,772	0,738	0,611	0,669	0,339	0,059	0,018	0,031	1,317	0,674
36	0,02190	0,118	0,109	1,527	0,629	0,749	0,508	0,725	0,465	0,017	0,007	0,012	1,082	0,669
37	0,00729	0,154	0,286	1,779	0,740	0,934	0,538	0,969	0,443	0,025	0,006	0,108	1,273	0,725
38	0,00011	0,118	0,222	2,143	0,727	0,913	0,531	0,938	0,434	0,003	0,001	0,022	1,186	0,937
39	0,01624	0,295	0,520	2,505	0,740	0,886	0,568	0,932	0,390	0,013	0,009	0,007	1,046	0,691
40	0,10166	1,051	2,050	0,590	0,620	0,748	0,512	0,513	0,456	0,067	0,044	0,022	1,251	0,627
41	0,00124	0,658	0,935	1,407	0,751	0,980	0,708	0,704	0,273	0,328	0,104	0,023	1,386	0,798
42	0,10233	1,415	2,566	0,455	0,624	0,715	0,551	0,552	0,417	0,074	0,038	0,008	1,647	0,574
43	0,02371	0,384	0,621	2,041	0,618	0,775	0,646	0,617	0,337	0,031	0,017	0,011	1,403	0,753
44	0,12430	1,468	2,496	0,495	0,703	0,725	0,588	0,588	0,396	0,071	0,053	0,024	1,486	0,147
45	0,16733	1,447	2,562	0,441	0,632	0,712	0,565	0,565	0,391	0,028	0,013	0,008	1,275	0,704
46	0,28663	1,502	2,746	0,398	0,591	0,674	0,547	0,547	0,429	0,025	0,021	0,008	1,081	0,538
47	0,03417	0,421	0,657	1,162	0,637	0,792	0,661	0,714	0,301	0,042	0,025	0,022	1,471	0,654
48	0,01482	0,136	0,253	3,402	0,716	0,895	0,614	0,721	0,351	0,011	0,007	0,016	1,294	0,781
49	0,0254	0,347	0,583	0,826	0,721	0,874	0,629	0,675	0,334	0,019	0,011	0,013	1,179	0,714
50	0,03307	0,521	0,623	1,541	0,732	0,886	0,637	0,704	0,323	0,023	0,014	0,017	1,031	0,642
51	0,0438	0,673	0,748	0,804	0,649	0,736	0,653	0,692	0,304	0,037	0,021	0,015	1,282	0,617
52	0,02741	0,385	0,561	1,022	0,763	0,894	0,581	0,713	0,382	0,017	0,009	0,012	1,108	0,721
53	0,06504	0,532	0,674	0,781	0,604	0,728	0,625	0,671	0,336	0,053	0,031	0,028	1,438	0,677
54	0,02718	0,127	0,241	1,506	0,647	0,714	0,538	0,765	0,423	0,014	0,007	0,009	1,171	0,736

Приложение И
Распределение t-критерия Стьюдента
при уровне значимости 0,10, 0,05 и 0,01

ν	α			ν	α		
	0,10	**0,05**	**0,01**		**0,10**	**0,05**	**0,01**
1	6,3138	12,706	63,657	18	1,7341	2,1009	2,8784
2	2,9200	4,3027	9,9248	19	1,7291	2,0930	2,8609
3	2,3534	3,1825	5,8409	20	1,7247	2,0860	2,8453
4	2,1318	2,7764	4,6041	21	1,7207	2,0796	2,8314
5	2,0150	2,5706	4,0321	22	1,7171	2,0739	2,8188
6	1,9432	2,4469	3,7074	23	1,7139	2,0687	2,8073
7	1,8946	2,3646	3,4995	24	1,7109	2,0639	2,7969
8	1,8595	2,3060	3,3554	25	1,7081	2,0595	2,7874
9	1,8331	2,2622	3,2498	26	1,7056	2,0555	2,7757
10	1,8125	2,2281	3,1693	27	1,7033	2,0518	2,7707
11	1,7959	2,2010	3,1058	28	1,7011	2,0484	2,7633
12	1,7823	2,1788	3,0545	29	1,6991	2,0452	2,7564
13	1,7709	2,1604	3,0123	30	1,6973	2,0423	2,7500
14	1,7613	2,1448	2,9768	40	1,6839	2,0211	2,7045
15	1,7530	2,1315	2,9467	60	1,6707	2,0003	2,6603
16	1,7459	2,1199	2,9208	120	1,6577	1,9799	2,6174
17	1,7396	2,1098	2,8982	∞	1,6449	1,9600	2,5758

ν – степени свободы вариации

Приложение К
Исходные данные коммерческой организации для выполнения линейного множественного корреляционно-регрессионного анализа

Таблица К.1 – Исходные данные организации

Выручка от реализации продукции, млн ден. ед.	Год	Расходы на рекламу, млн ден. ед.	Цена товара, ден. ед.	Средняя цена товара у конкурентов, ден. ед.	Индекс потребительских расходов, %
Y	X1	X2	X3	X4	X5
126	1	4	15	17	100
137	2	4,8	14,8	17,3	98,4
148	3	3,8	15,2	16,8	101,2
191	4	8,7	15,5	16,2	103,5
274	5	8,2	15,5	16	104,1
370	6	9,7	16	18	107
432	7	14,7	18,1	20,2	107,4
445	8	18,7	13	15,8	108,5
367	9	19,8	15,8	18,2	108,3
367	10	10,6	16,9	16,8	109,2
321	11	8,6	16,3	17	110,1
307	12	6,5	16,1	18,3	110,7
331	13	12,6	15,4	16,4	110,3
345	14	6,5	15,7	16,2	111,8
364	15	5,8	16	17,7	112,3
384	16	5,7	15,1	16,2	112,9

Учебное издание

ПРИКЛАДНОЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ ЭКОНОМИЧЕСКИХ ДАННЫХ

Методические указания к практическим занятиям

Составитель:
Быков Константин Романович

Редактор *Р.А. Никифорова*
Корректор *А.С. Прокопюк*
Компьютерная верстка *К.Р. Быков*

Подписано к печати 03.02.2025. Усл. печ. листов 6,9.
Уч.-изд. листов 8,6. Заказ № 29.

Учреждение образования «Витебский государственный технологический университет»
210038, г. Витебск, Московский пр., 72.
Отпечатано на ризографе учреждения образования
«Витебский государственный технологический университет».
Свидетельство о государственной регистрации издателя, изготовителя,
распространителя печатных изданий № 1/172 от 12 февраля 2014 г.
Свидетельство о государственной регистрации издателя, изготовителя,
распространителя печатных изданий № 3/1497 от 30 мая 2017 г.