

8.1. Введение в факторный анализ

Факторный анализ — это математический инструмент для понижения пространства признаков, широко применяющийся в экономике, социологии, психологии.

Информативность многомерного описания объекта возрастает с увеличением количества используемых признаков. Однако очень трудно сразу выбрать и существенные, и независимые друг от друга характеристики. Как правило, аналитик начинает с заведомо избыточного количества признаков, и в процессе работы сталкивается с необходимостью адекватной интерпретации большого объема полученных данных и их компактной визуализации. Возникает вопрос в том, что многие признаки, вероятно, в некоторой степени дублируют друг друга, а вся полученная информация в целом избыточна. За связанными друг с другом (коррелирующими) переменными, по-видимому, стоит влияние некоторой скрытой переменной (фактора), с помощью которой можно объяснить наблюдаемое сходство полученных оценок. Выделение факторов, как переменных более общего, более высокого порядка, позволяет по-новому взглянуть на полученные данные, заметить те связи между переменными, которые ранее не были очевидны.

В обработчике *Факторный анализ* для факторизации корреляционной матрицы используется *метод главных компонент*. Он сводится к выбору новой ортогональной системы координат в пространстве наблюдений. В качестве первой главной компоненты избирают направление, вдоль которого массив данных имеет наибольший разброс. Выбор каждой главной последующей компоненты происходит так, чтобы разброс данных вдоль нее был максимальным, и чтобы эта главная компонента была ортогональна другим главным компонентам, выбранным прежде. В результате получают несколько главных компонент, каждая следующая из которых несет все меньше информации из исходного набора. Следующим шагом является выбор наиболее информативных главных компонент, которые будут использоваться в дальнейшем анализе.

Таким образом, факторный анализ решает две главные задачи:

- понижение размерности числа используемых переменных за счет их объяснения меньшим числом факторов;
- группировка и структурирование полученных данных.

Факторы имеют две характеристики: долю объясняемой дисперсии и нагрузки. Результат процедуры факторизации заключается в формировании матрицы факторных нагрузок.

На практике аналитикам чаще всего интересен факторный анализ с ортогональным вращением осей, когда при повороте осей координат угол между факторами остается прямым. Цель исследователя заключается в поиске простой структуры или попытке объяснить большее число переменных меньшим числом факторов. «Простота» хорошего факторного решения заключается в том, что

каждая переменная имеет наиболее простое факторное объяснение, то есть характеризуется преобладающим влиянием некоторого одного фактора, и в меньшей степени связана с другими факторами. И наоборот: один фактор должен быть специфическим образом связан с одной группой переменных и не связан с другими переменными.

В узле реализовано два метода вращения:

- *варимакс* — наиболее часто используемый на практике метод, цель которого — минимизировать количество переменных, имеющих высокие нагрузки на данный фактор, что способствует упрощению описания фактора за счет группировки вокруг него только тех переменных, которые с ним связаны в большей степени, чем с остальными.

- *квартимакс* противоположен варимаксу, поскольку минимизирует количество факторов, необходимых для объяснения данной переменной. Квартимакс приводит к выделению одного из общих факторов с достаточно высокими нагрузками на большинство переменных.

После расчета факторных нагрузок для каждой переменной доступны два показателя: собственное значение и объём объясняемой дисперсии в %, а также суммарный процент дисперсии. Собственное значение фактора – это его вклад в дисперсию переменных, объясняемую влиянием общих факторов. Считается, что те факторы, у которых этот показатель меньше 1,0, не вносят значительного вклада в объяснение результата.

Второй расчетный показатель – процент объясняемой дисперсии переменных. Принято считать, что при хорошем факторном решении выбирают столько факторов, чтобы они в сумме объясняли не менее 70–75 %. В отдельных случаях этот показатель может достигать 85–90 %.

Главной проблемой факторного анализа является выделение и интерпретация главных факторов.

В обработчике *Факторный анализ* помимо вида метода (варимакс, квартимакс, без вращения) выбирается число выделяемых факторов.

Можно задать непосредственно число факторов в диапазоне от 1 до общего числа переменных, или задать долю дисперсии, описанной выделяемыми факторами по отношению к общей дисперсии.

После расчета факторных нагрузок количество выявленных факторов можно изменить, уточняя порог значимости или количество факторов.

8.2. Методические указания

8.2.1. Снижение размерности данных

В файле *Задача 8.1. Руководители и специалисты.xlsx* имеются данные распределения областей Центрального федерального округа РФ по профессионально-квалификационной структуре руководителей и специалистов сельскохозяйственных предприятий. Для характеристики профессионально-квалификационной структуры было отобрано девять показателей (рис. 8.1).

	A	B	C	D	E	F	G	H	I	J
1	Область	Обеспеченность кадрами от штатной потребности	Обеспеченность кадрами с высшим и средним образованием от штатной потребности	Доля кадров с высшим и средним образованием в общей численности	Доля кадров с высшим образованием в общей численности	Доля практиков в общей численности	Доля кадров, повысивших квалификацию в общей численности	Доля принятых на работу кадров в общей численности	Доля принятых на работу выпускников высших и средних учебных заведений в общей численности	Доля уволенных с работы кадров в общей численности
2	Белгородская	95,8	85,4	89,1	40	10,9	4,8	11,5	1,2	13,1
3	Брянская	85,2	72,4	84,9	25,8	15,1	3,3	7,8	1,2	8,7
4	Владимирская	92,5	80,5	87	34,5	13	3,3	8,7	0,8	10,6
17	Тульская	87,9	77,5	88,2	27,5	11,8	14,3	7,4	1,1	8,3
18	Ярославская	90,9	79,2	87,1	28,3	12,9	17,3	11,1	1,4	13,2

Рис. 8.1

Требуется:

- 1) построить матрицу факторных нагрузок, используя метод главных компонент;
- 2) выделить главные факторы, определяющие уровень обеспеченности кадрами;
- 3) найти численные значения главных компонент для всех объектов наблюдения.

Решение

Создадим новый пакет *Факторный анализ*. Выполним импорт исходных данных. Для этого создадим узел сценария, выполняющий действие импорта (рис. 8.2).

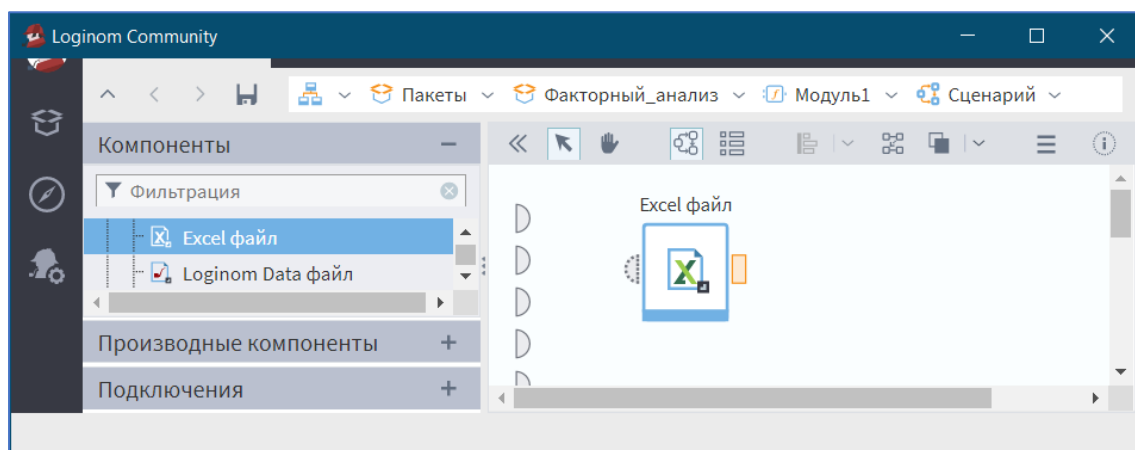


Рис. 8.2

Вызовем *Мастер настройки*. Пройдем шаги мастера, указав в описании узла метку *Задача 8.1. Руководители и специалисты*.

Добавим визуализаторы *Таблица* и *Статистика* (рис. 8.3).

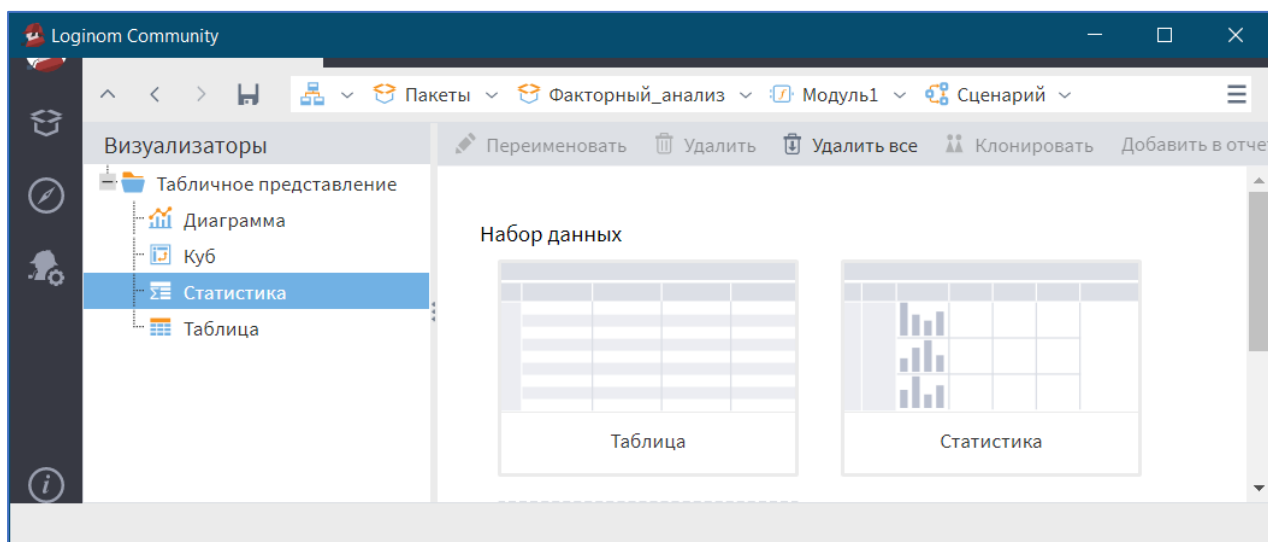


Рис. 8.3

Визуализаторы с исходными данными и статистическими показателями имеют вид (рис. 8.4–8.5).

#	ab Область	9.0 Обеспеченность кадрами от штат...	9.0 Обеспеченность кадрами с выс...	9.0 Доля кадров с выс...
1	Белгородская	95,8	85,4	89,1
2	Брянская	85,2	72,4	84,9
3	Владимирская	92,5	80,5	87
4	Воронежская	96,5	85,8	88,9
5	Ивановская	92,9	82,6	88,9

Рис. 8.4

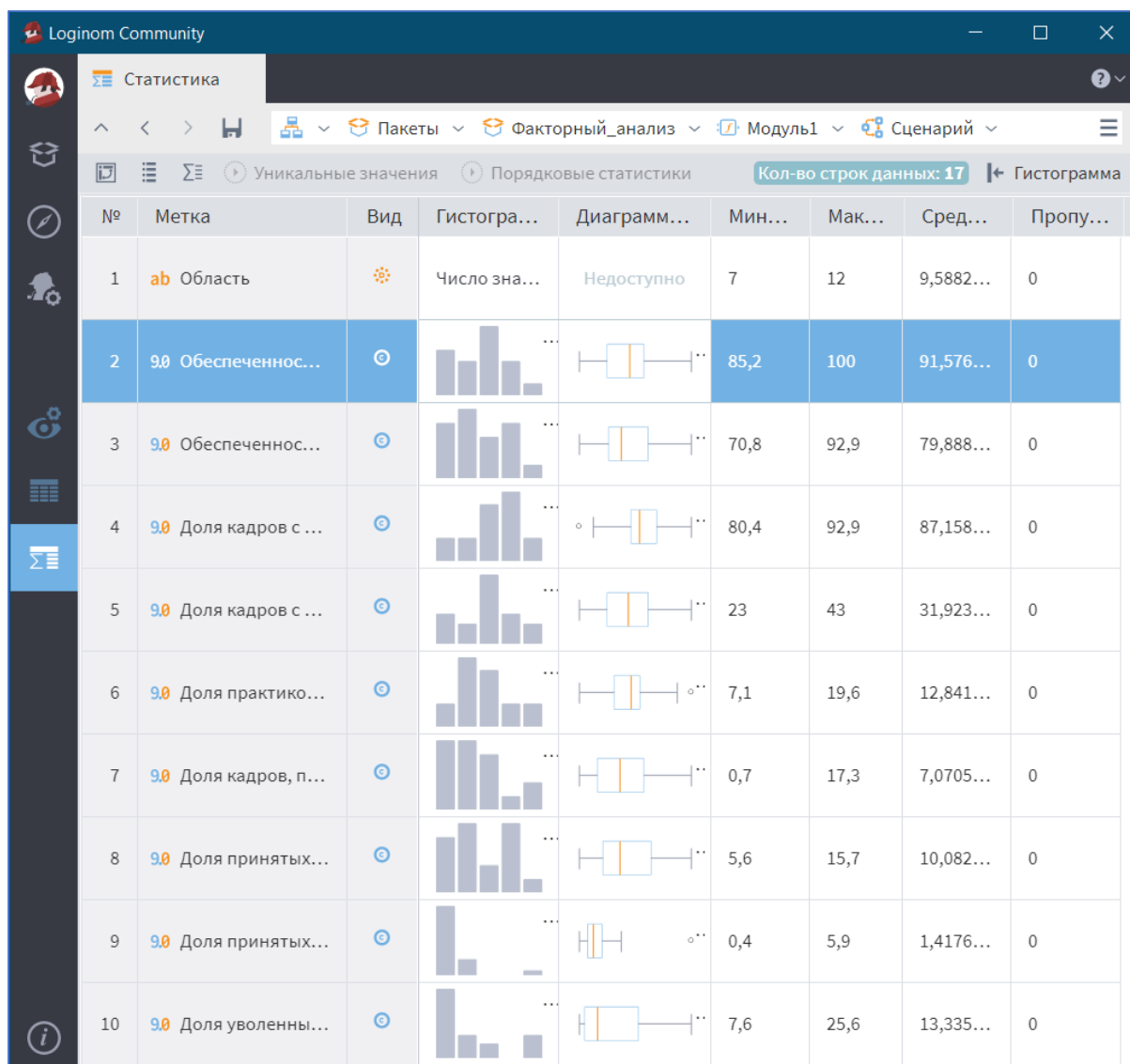


Рис. 8.5

Проведем факторный анализ на основе импортированных данных. Для этого переместим компонент *Факторный анализ* в рабочую область сценария. Последовательность обработки данных задастся соединением выходного порта узла импорта с входным портом факторного анализа (рис. 8.6).

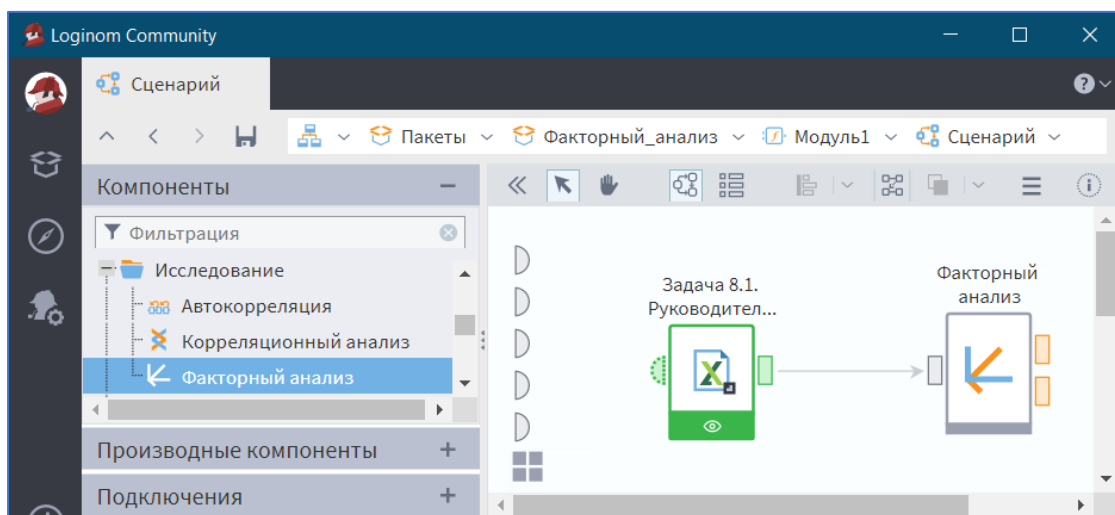


Рис. 8.6

Пройдем шаги *Мастера настройки*. На шаге *Настройка входных столбцов* настроим назначение исходных столбцов данных. Столбец *Область* зададим как *Не задано*, остальные столбцы — как *Используемое* (рис. 8.7).

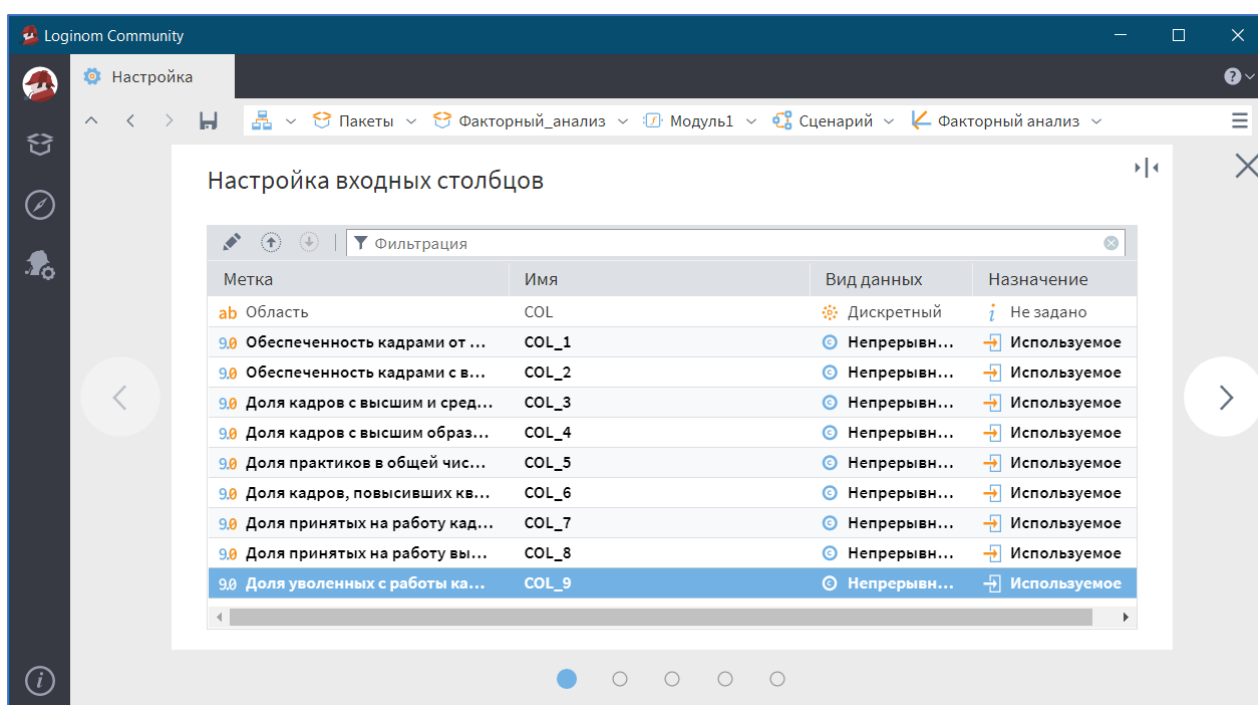


Рис. 8.7

На шаге *Факторный анализ* оставим стандартные параметры по умолчанию (рис. 8.8).

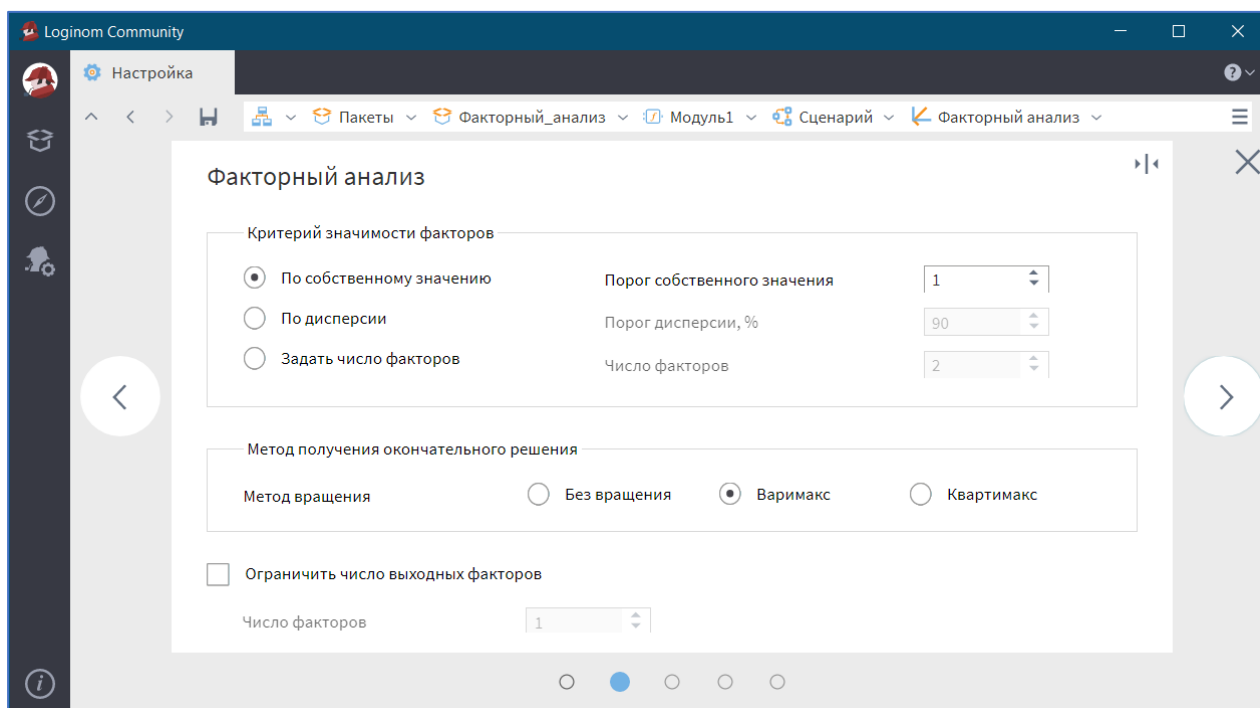


Рис. 8.8

Переобучим узел *Факторный анализ* и перейдем к настройкам визуализаторов (рис. 8.9).

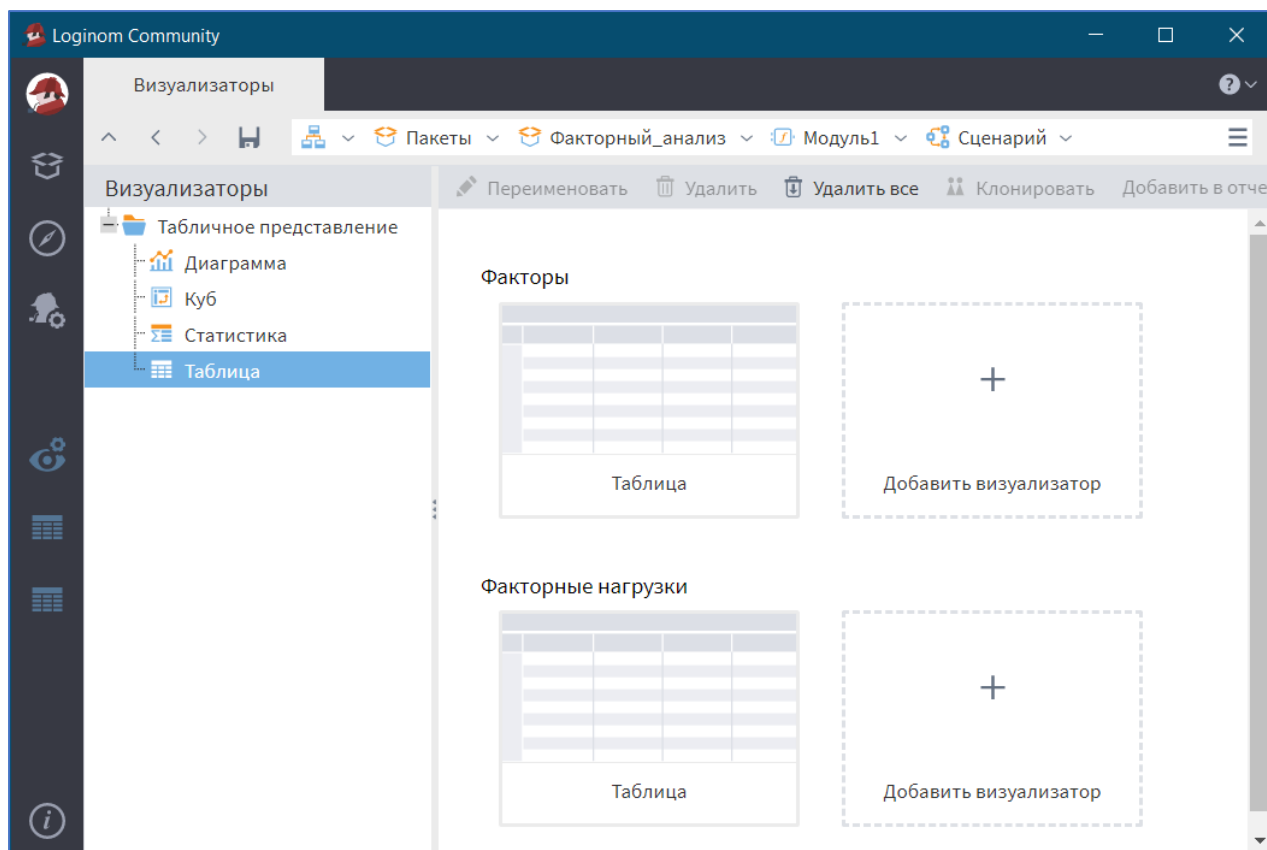
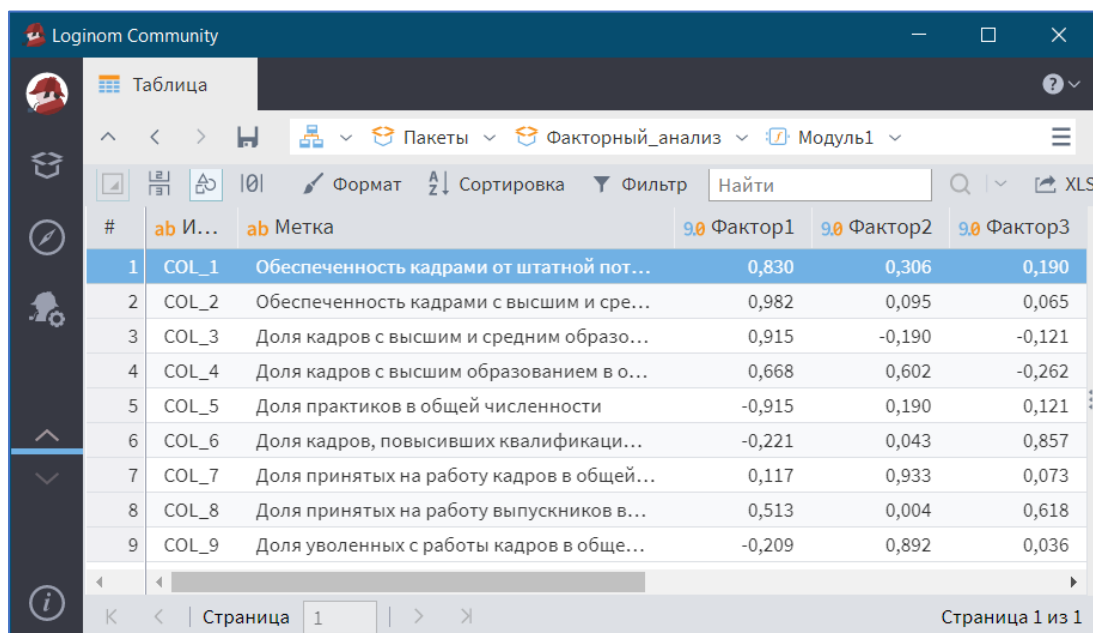


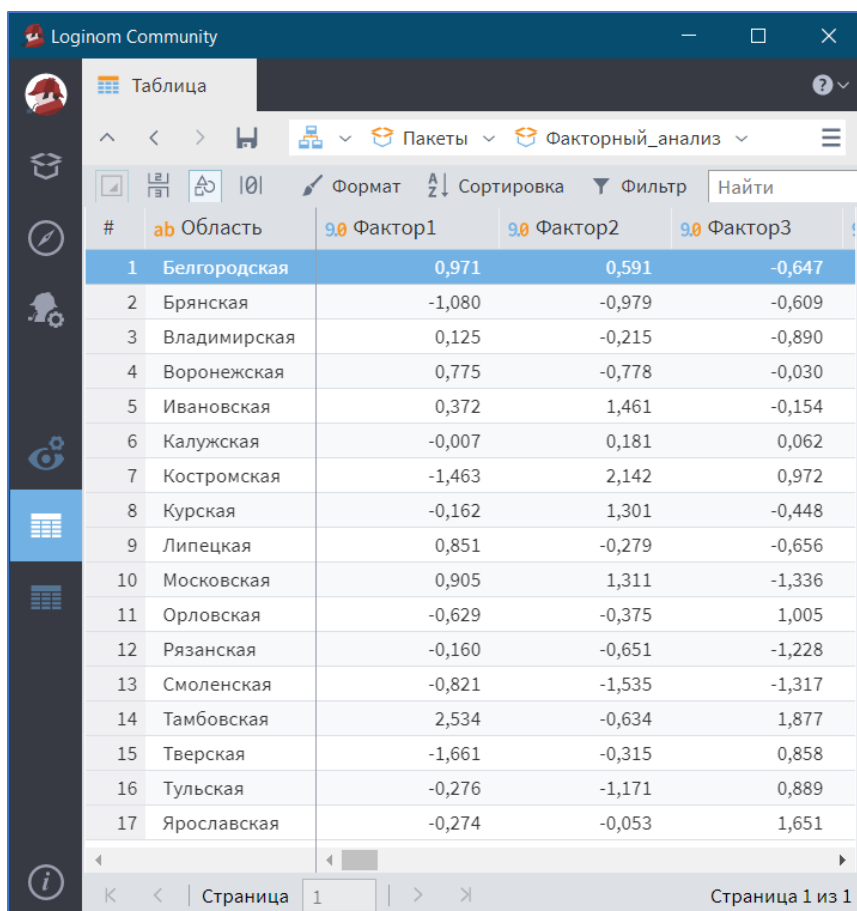
Рис. 8.9

В визуализаторе *Факторные нагрузки* представлена матрица факторных нагрузок (рис. 8.10), а в визуализаторе *Факторы*— численные значения выделенных факторов (рис. 8.11). Всего выделено три фактора.



#	ab И...	ab Метка	9.0 Фактор1	9.0 Фактор2	9.0 Фактор3
1	COL_1	Обеспеченность кадрами от штатной пот...	0,830	0,306	0,190
2	COL_2	Обеспеченность кадрами с высшим и сре...	0,982	0,095	0,065
3	COL_3	Доля кадров с высшим и средним образо...	0,915	-0,190	-0,121
4	COL_4	Доля кадров с высшим образованием в о...	0,668	0,602	-0,262
5	COL_5	Доля практиков в общей численности	-0,915	0,190	0,121
6	COL_6	Доля кадров, повысивших квалификаци...	-0,221	0,043	0,857
7	COL_7	Доля принятых на работу кадров в общей...	0,117	0,933	0,073
8	COL_8	Доля принятых на работу выпускников в...	0,513	0,004	0,618
9	COL_9	Доля уволенных с работы кадров в обще...	-0,209	0,892	0,036

Рис. 8.10



#	ab Область	9.0 Фактор1	9.0 Фактор2	9.0 Фактор3
1	Белгородская	0,971	0,591	-0,647
2	Брянская	-1,080	-0,979	-0,609
3	Владимирская	0,125	-0,215	-0,890
4	Воронежская	0,775	-0,778	-0,030
5	Ивановская	0,372	1,461	-0,154
6	Калужская	-0,007	0,181	0,062
7	Костромская	-1,463	2,142	0,972
8	Курская	-0,162	1,301	-0,448
9	Липецкая	0,851	-0,279	-0,656
10	Московская	0,905	1,311	-1,336
11	Орловская	-0,629	-0,375	1,005
12	Рязанская	-0,160	-0,651	-1,228
13	Смоленская	-0,821	-1,535	-1,317
14	Тамбовская	2,534	-0,634	1,877
15	Тверская	-1,661	-0,315	0,858
16	Тульская	-0,276	-1,171	0,889
17	Ярославская	-0,274	-0,053	1,651

Рис. 8.11

Из таблицы на рис. 8.10 видно, что первый фактор наиболее тесно связан с такими показателями, как обеспеченность руководителями и специалистами от штатной потребности, в том числе с высшим и средним образованием, доля руководителей и специалистов с высшим и средним образованием в общей численности, в том числе с высшим образованием, доля практиков в общей численности. Причем связь его с таким показателем как доля практиков в общей численности руководителей и специалистов является отрицательной. Данные показатели характеризуют как обеспеченность кадрами в целом, так и их квалификацию (наличие высшего и среднего образования), поэтому первый фактор целесообразно назвать *Обеспеченность кадрами*.

Второй фактор можно определить как *Текучесть кадров*, поскольку он имеет наибольшие положительные нагрузки на показатели доля руководителей и специалистов принятых на работу и уволенных с работы в общей их численности.

Третий фактор характеризует рост образовательного уровня кадров, так как определяется долей руководителей и специалистов, повысивших квалификацию и долей принятых на работу выпускников высших и средних учебных заведений. Поэтому данный фактор можно назвать *Подготовка кадров*.

Таким образом, девять исходных показателей были сведены к трем факторам, которые характеризуют основные стороны профессионально-квалификационной структуры кадров управления АПК.

Данные таблицы на рис. 8.11 позволяют определить, какое место занимают исследуемые области по уровню обеспеченности кадрами руководителей и специалистов сельскохозяйственных предприятий, текущесть кадров и их подготовке.

Так, более высокий уровень обеспеченности руководителями и специалистами в Тамбовской области. В ней наиболее высокая обеспеченность кадрами от штатной потребности, в том числе с высшим и средним образованием, самый большой удельный вес руководителей и специалистов с высшим и средним образованием и самая низкая доля практиков (рис. 8.12). Также данной области принадлежит первое место по уровню подготовки кадров (рис. 8.13), а по уровню текущесть кадров она занимает лишь двенадцатое место (8.14). Наиболее высокий уровень текущесть кадров в Костромской области, по значениям таких показателей как доля принятых на работу и уволенных с работы она занимает соответственно первое и третье места.

Loginom Community

Таблица

Пакеты

Формат A Z Сортировка Фильтр На

#	ab Область	9.0 Фактор1	9.0 Фактор2	9.0 Фактор3
1	Тамбовская	2,534	-0,634	1,877
2	Белгородская	0,971	0,591	-0,647
3	Московская	0,905	1,311	-1,336
4	Липецкая	0,851	-0,279	-0,656
5	Воронежская	0,775	-0,778	-0,030
6	Ивановская	0,372	1,461	-0,154
7	Владимирская	0,125	-0,215	-0,890
8	Калужская	-0,007	0,181	0,062
9	Рязанская	-0,160	-0,651	-1,228
10	Курская	-0,162	1,301	-0,448
11	Ярославская	-0,274	-0,053	1,651
12	Тульская	-0,276	-1,171	0,889
13	Орловская	-0,629	-0,375	1,005
14	Смоленская	-0,821	-1,535	-1,317
15	Брянская	-1,080	-0,979	-0,609
16	Костромская	-1,463	2,142	0,972
17	Тверская	-1,661	-0,315	0,858

Страница 1 из 1

Рис. 8.12

Loginom Community

Таблица

Пакеты

Формат A Z Сортировка Фильтр На

#	ab Область	9.0 Фактор1	9.0 Фактор2	9.0 Фактор3
1	Тамбовская	2,534	-0,634	1,877
2	Ярославская	-0,274	-0,053	1,651
3	Орловская	-0,629	-0,375	1,005
4	Костромская	-1,463	2,142	0,972
5	Тульская	-0,276	-1,171	0,889
6	Тверская	-1,661	-0,315	0,858
7	Калужская	-0,007	0,181	0,062
8	Воронежская	0,775	-0,778	-0,030
9	Ивановская	0,372	1,461	-0,154
10	Курская	-0,162	1,301	-0,448
11	Брянская	-1,080	-0,979	-0,609
12	Белгородская	0,971	0,591	-0,647
13	Липецкая	0,851	-0,279	-0,656
14	Владимирская	0,125	-0,215	-0,890
15	Рязанская	-0,160	-0,651	-1,228
16	Смоленская	-0,821	-1,535	-1,317
17	Московская	0,905	1,311	-1,336

Страница 1 из 1

Рис. 8.13

#	ab Область	9.0 Фактор1	9.0 Фактор21	9.0 Фактор3
1	Костромская	-1,463	2,142	0,972
2	Ивановская	0,372	1,461	-0,154
3	Московская	0,905	1,311	-1,336
4	Курская	-0,162	1,301	-0,448
5	Белгородская	0,971	0,591	-0,647
6	Калужская	-0,007	0,181	0,062
7	Ярославская	-0,274	-0,053	1,651
8	Владимирская	0,125	-0,215	-0,890
9	Липецкая	0,851	-0,279	-0,656
10	Тверская	-1,661	-0,315	0,858
11	Орловская	-0,629	-0,375	1,005
12	Тамбовская	2,534	-0,634	1,877
13	Рязанская	-0,160	-0,651	-1,228
14	Воронежская	0,775	-0,778	-0,030
15	Брянская	-1,080	-0,979	-0,609
16	Тульская	-0,276	-1,171	0,889
17	Смоленская	-0,821	-1,535	-1,317

Рис. 8.14

8.2.2. Снижение размерности данных для регрессионной модели

В файле *Задача 8.2. Урожайность картофеля.xlsx* имеются данные об урожайности картофеля по 36 хозяйствам и факторах, оказывающих на нее влияние (рис. 8.15).

	A	B	C	D	E	F	G
1	№ хозяйства	Урожайность картофеля с 1 га, ц	Высажено картофеля на 1 га, ц	Внесено органических удобрений на 1 га, т	Внесено минеральных удобрений на 1 га, ц	Удельный вес минеральных удобрений, внесенных в подкормку, %	Доля посадок картофеля, размещенных по лучшим предшественникам, %
2	1	428	26,7	34	4,8	34	100
3	2	288	29,1	17	3,2	25	76
4	3	432	37,9	31	5,5	34	65
36	35	288	28,4	14	2,4	21	47
37	36	324	25,5	18	3,6	26	52

Рис. 8.15

Требуется, используя метод главных компонент, сократить число факторов, оказывающих влияние на урожайность картофеля, и построить регрессионную модель зависимости урожайности от выделенных факторов.

Решение

Выполним импорт исходных данных. Для этого создадим узел сценария, выполняющий действие импорта (рис. 8.16).

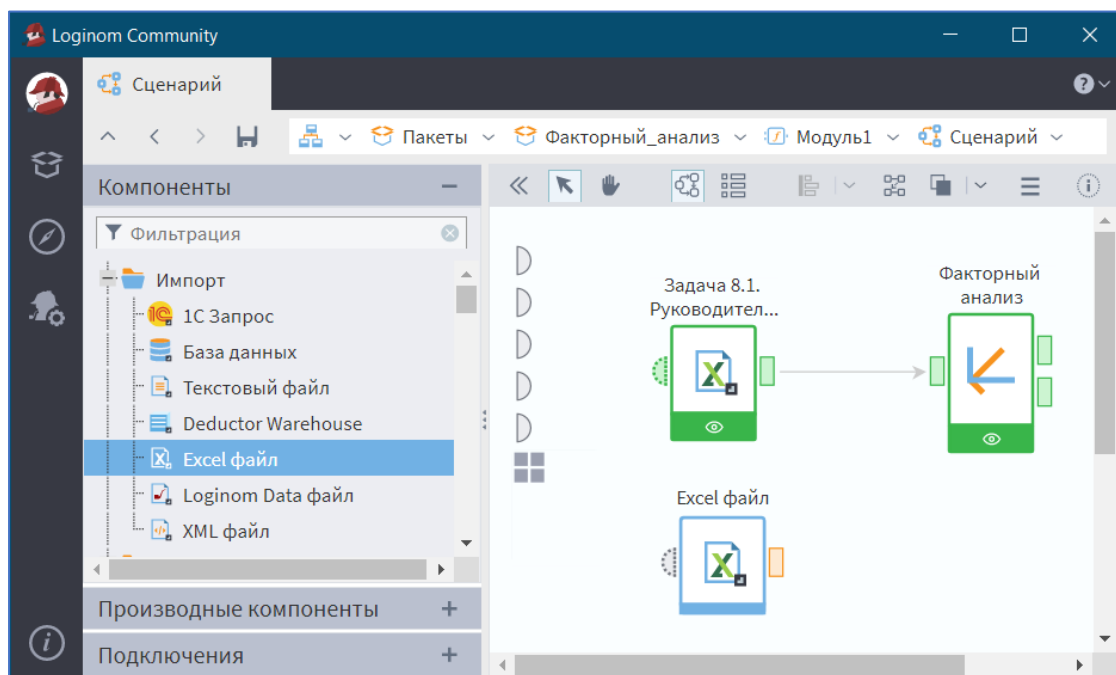


Рис. 8.16

Вызовем *Мастер настройки*. Пройдем шаги мастера. На шаге *Настройка полей* изменим тип данных по колонкам COL_1– COL_6 на *Вещественный* (рис. 8.17), указав в описании узла метку *Задача 8.2. Урожайность картофеля*.

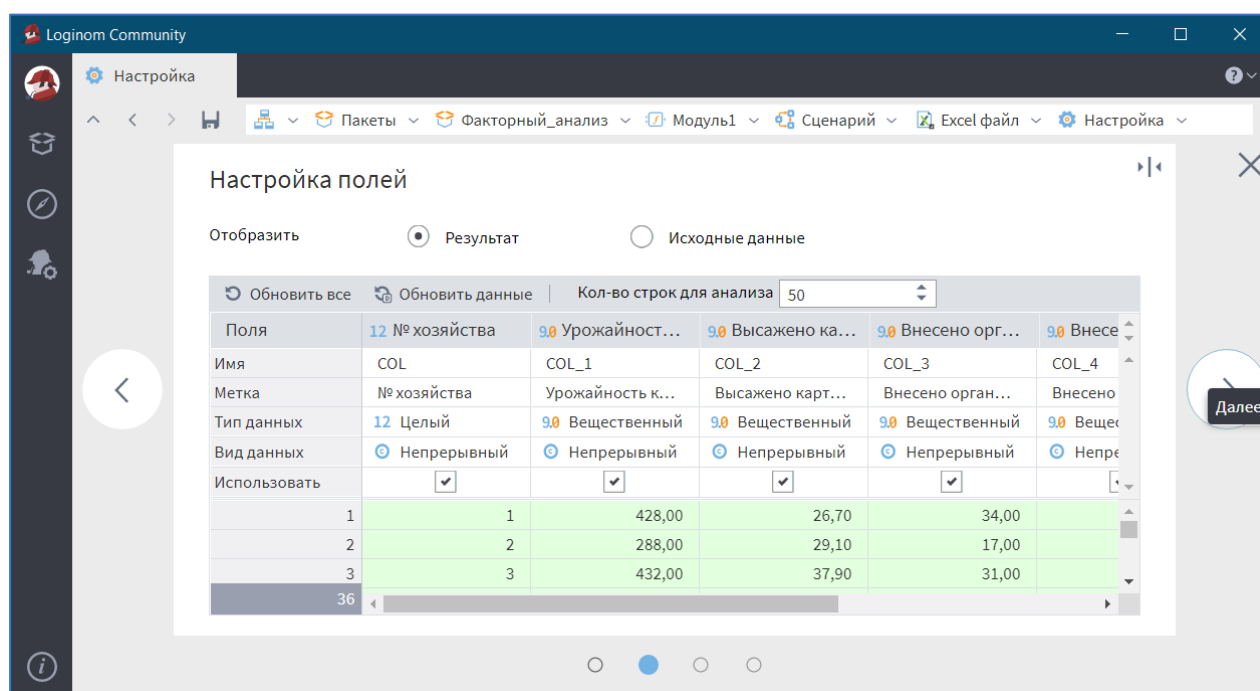


Рис. 8.17

Добавим визуализаторы *Таблица* и *Статистика* (рис. 8.18).

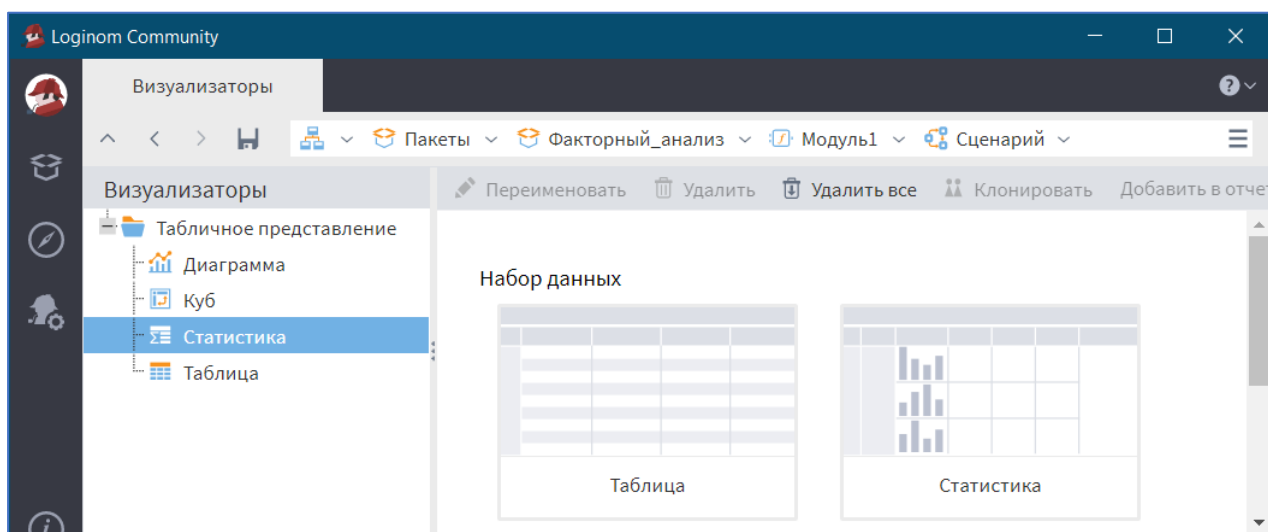


Рис. 8.18

Визуализаторы с исходными данными и статистическими показателями имеют вид (рис. 8.19–8.20).

#	№ хоз...	Урожай...	Высаж...	Внесен...	Внесен...	Удельн...	Доля п...
1	1	428	26,7	34	4,8	34	100
2	2	288	29,1	17	3,2	25	76
3	3	432	37,9	31	5,5	34	65
4	4	348	29,5	24	3,7	21	88
5	5	358	28,4	21	3,8	23	100

Рис. 8.19

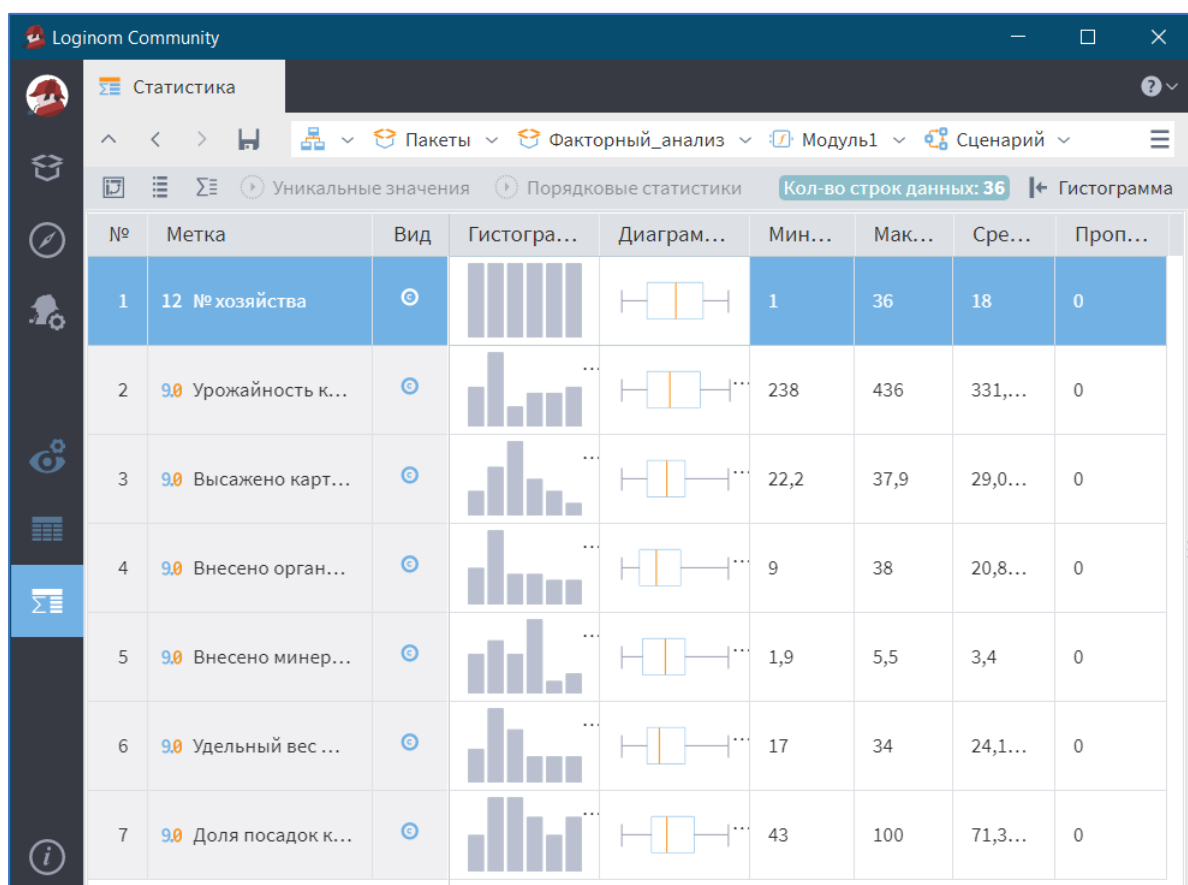


Рис. 8.20

Проведем факторный анализ на основе импортированных данных. Для этого переместим компонент *Факторный анализ* в рабочую область сценария. Последовательность обработки данных задается соединением выходного порта узла импорта с входным портом факторного анализа (рис. 8.21).

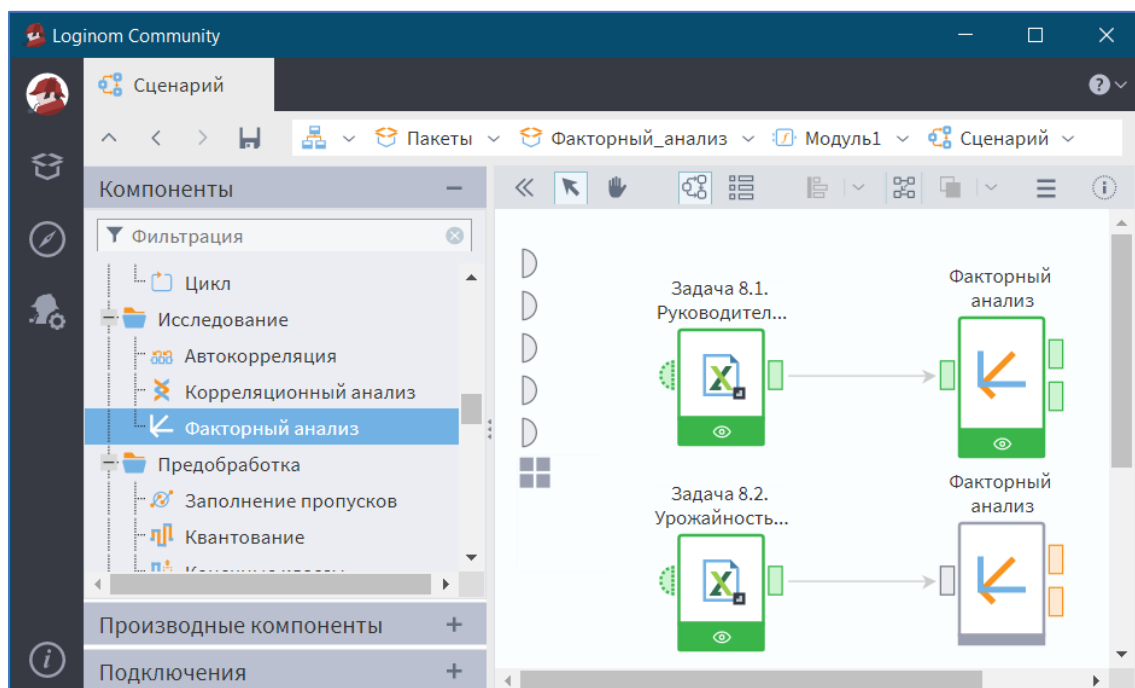


Рис. 8.21

Пройдем шаги *Мастера настройки*. На шаге *Настройка входных столбцов* настроим назначение исходных столбцов данных. Столбцы № хозяйства и Урожайность картофеля с 1 га, ц зададим как *Не задано*, остальные столбцы — как *Используемое* (рис. 8.22).

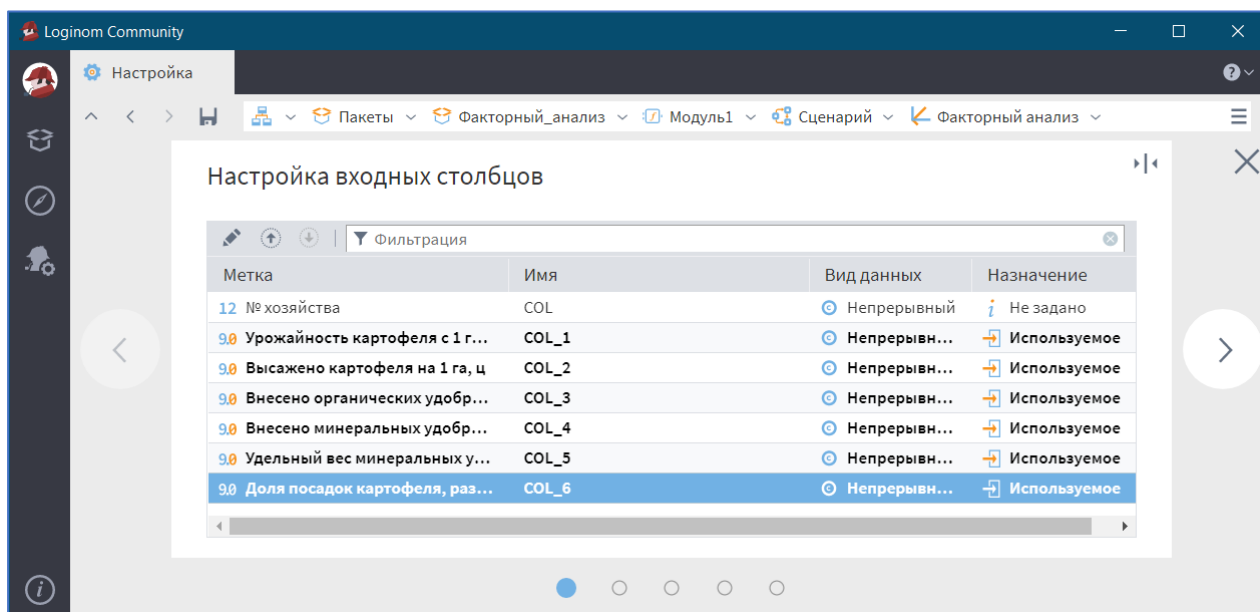


Рис. 8.22

На шаге *Факторный анализ* зададим критерий значимости факторов *По дисперсии*, назначив порог в 90% (рис. 8.23).

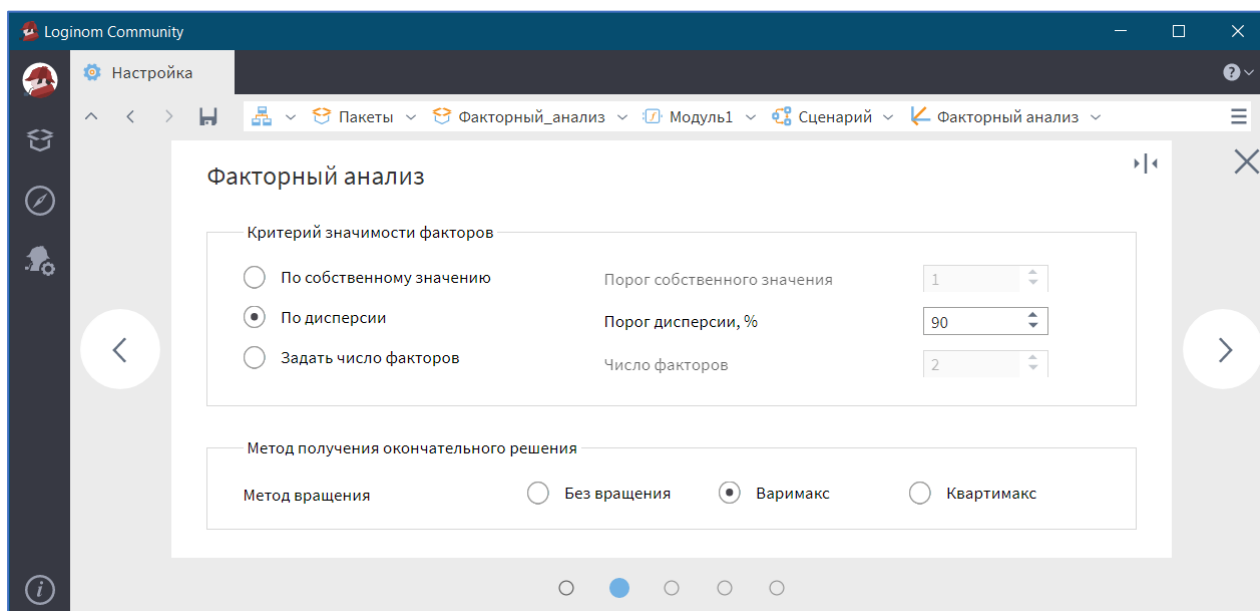


Рис. 8.23

Переобучим узел *Факторный анализ* и перейдем к настройкам визуализаторов (рис. 8.24).

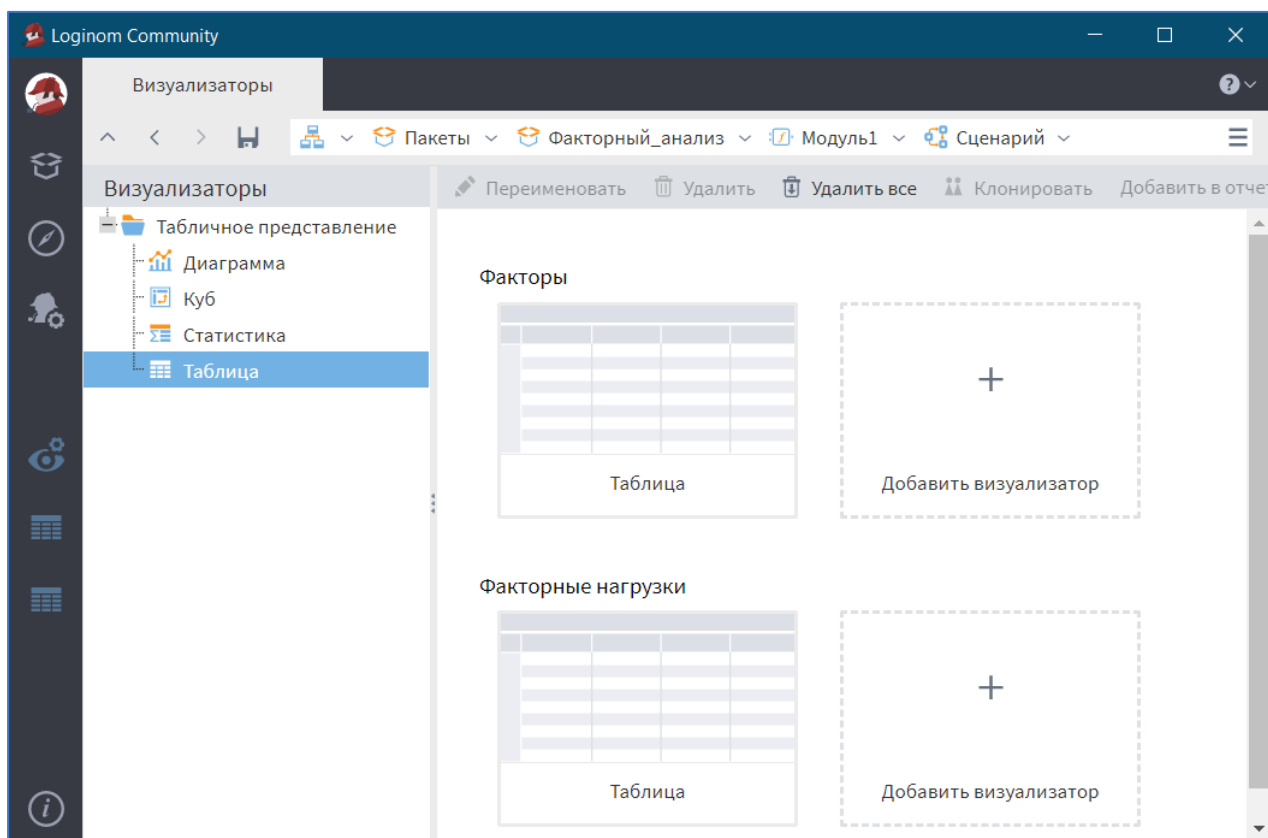


Рис. 8.24

В визуализаторе *Факторные нагрузки* представлена матрица факторных нагрузок (рис. 8.25), а в визуализаторе *Факторы*— численные значения выделенных факторов (рис. 8.26).

#	ab И...	ab Метка	9.0 Фактор1	9.0 Фактор2	9.0 Фактор3
1	COL_1	Урожайность картофеля с 1 га, ц	0,918	0,277	0,196
2	COL_2	Высажено картофеля на 1 га, ц	0,217	0,001	0,975
3	COL_3	Внесено органических удобрений на 1 ...	0,912	0,184	0,228
4	COL_4	Внесено минеральных удобрений на 1 г...	0,905	0,225	0,178
5	COL_5	Удельный вес минеральных удобрений,...	0,852	0,220	0,086
6	COL_6	Доля посадок картофеля, размещенны...	0,314	0,949	-0,001

Рис. 8.25

#	№ хозяйст...	9.0 Фактор1	9.0 Фактор2	9.0 Фактор3
1	1	1,855	1,176	-1,098
2	2	-0,491	0,442	0,067
3	3	1,940	-1,017	1,979
4	4	-0,224	1,113	0,266
5	5	-0,282	1,913	-0,079

Рис. 8.26

Из таблицы видно, что первый фактор наиболее тесно связан с такими показателями, как внесено органических удобрений на 1 га, внесено минеральных удобрений на 1 га и удельный вес минеральных удобрений, внесенных в подкормку, поэтому первый фактор целесообразно назвать *Внесено удобрений*.

Второй и третий факторы можно определить, соответственно, как *Доля посадок по лучшим предшественникам* и *Высажено картофеля на 1 га*, поскольку они тесно связаны с одноименными показателями.

Построим линейную модель на основе выделенных факторов. Для этого переместим компонент *Линейная регрессия* в рабочую область сценария. Последовательность обработки данных задается соединением выходного порта узла факторного анализа *Факторы* с входным портом линейной регрессии (рис. 8.27).

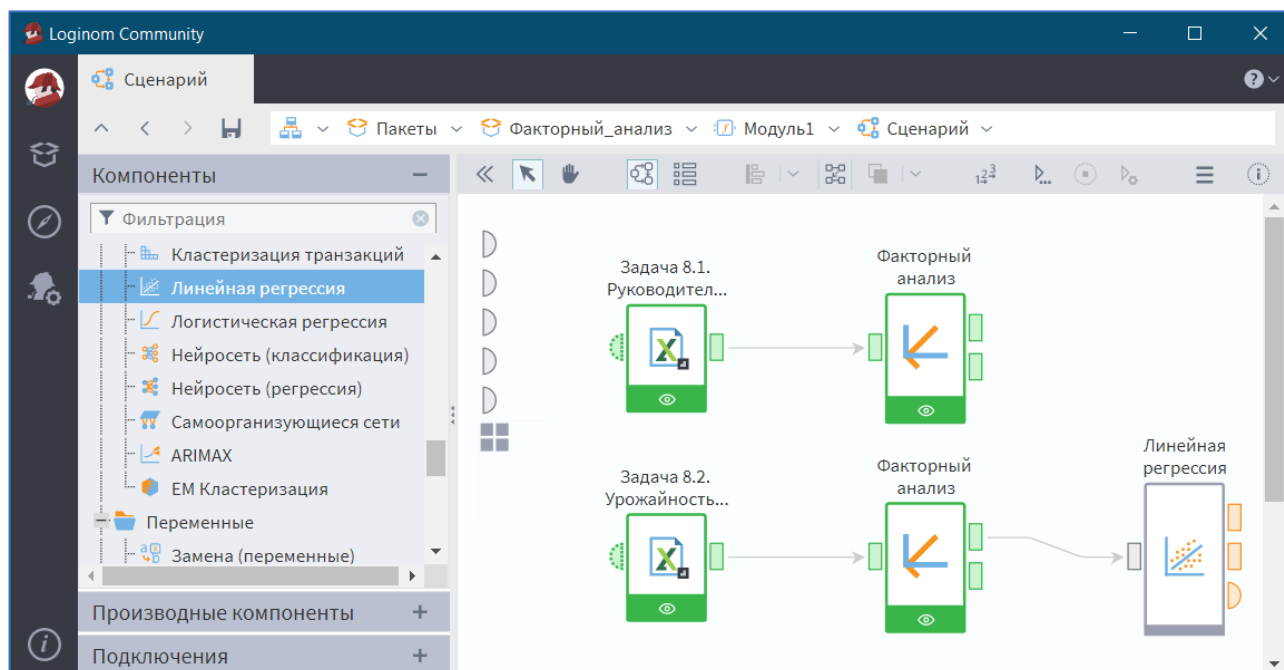


Рис. 8.27

Пройдем шаги *Мастера настройки*. На шаге *Настройка входных столбцов* настроим назначение исходных столбцов данных. Столбец *Урожайность картофеля с 1 га*, и зададим как выходной, столбцы *Фактор1*, *Фактор2* и *Фактор3* — как входные (рис. 8.28).

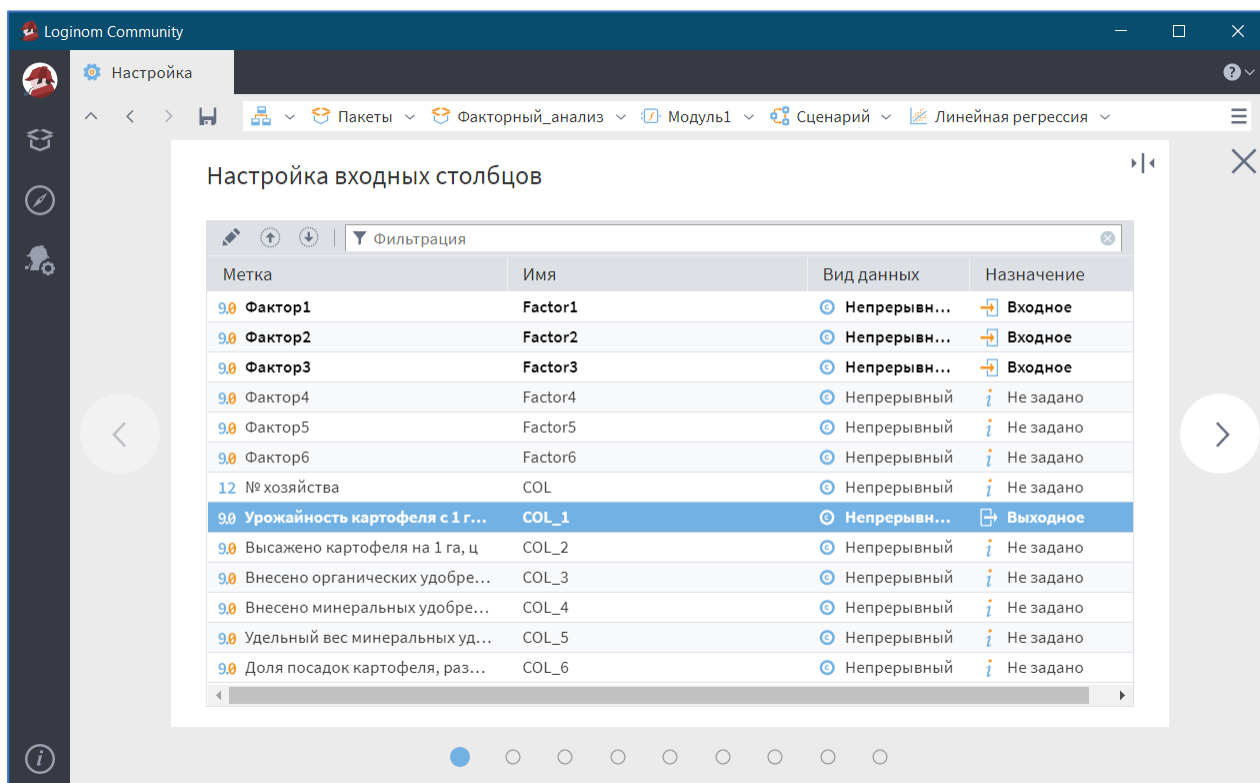


Рис. 8.28

На шагах *Настройка нормализации* и *Разбиение на множества* оставим стандартные параметры по умолчанию.

На шаге *Настройка линейной регрессии* снимем флажок с параметра *Автоматическая настройка*, выберем в качестве метода отбора факторов и защиты от переобучения *Принудительное включение* (рис. 8.29).

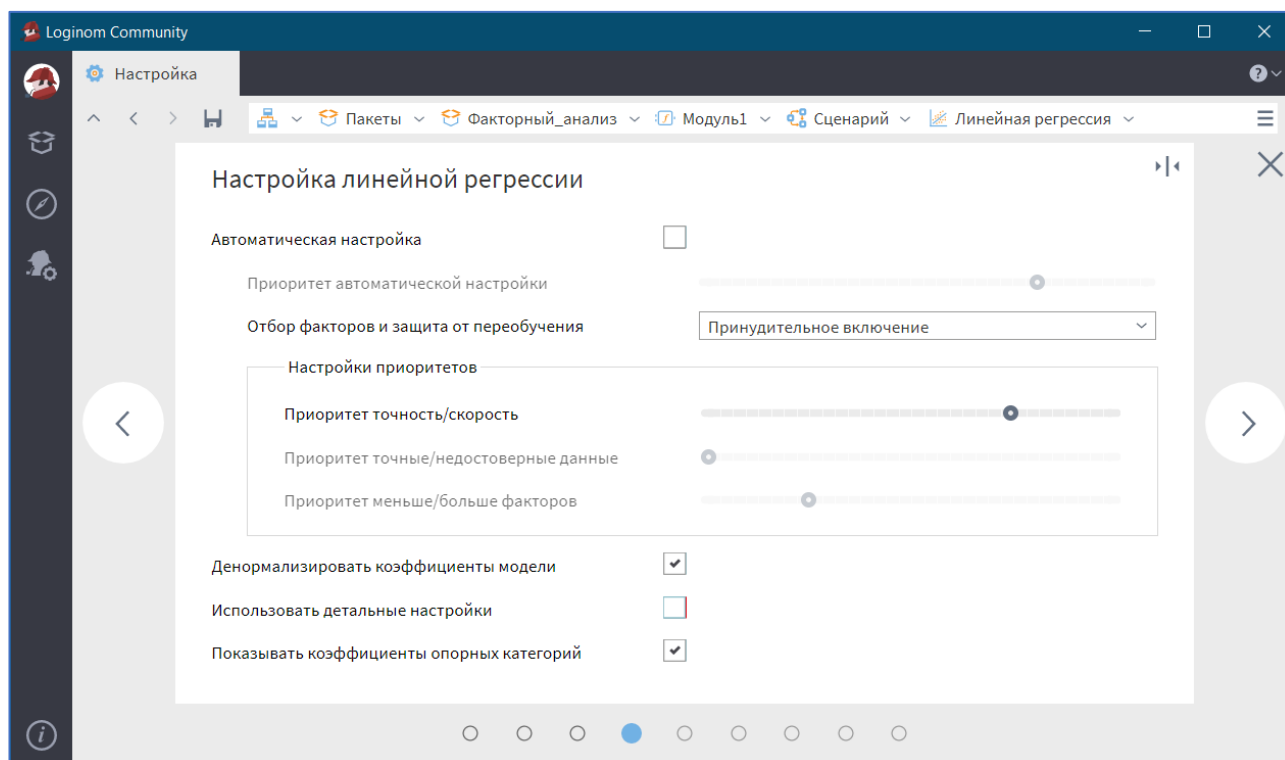


Рис. 8.29

На шаге *Описание узла* оставим стандартные параметры по умолчанию.

Переобучим узел *Линейная регрессия* и перейдем к настройкам визуализаторов (рис. 8.30).

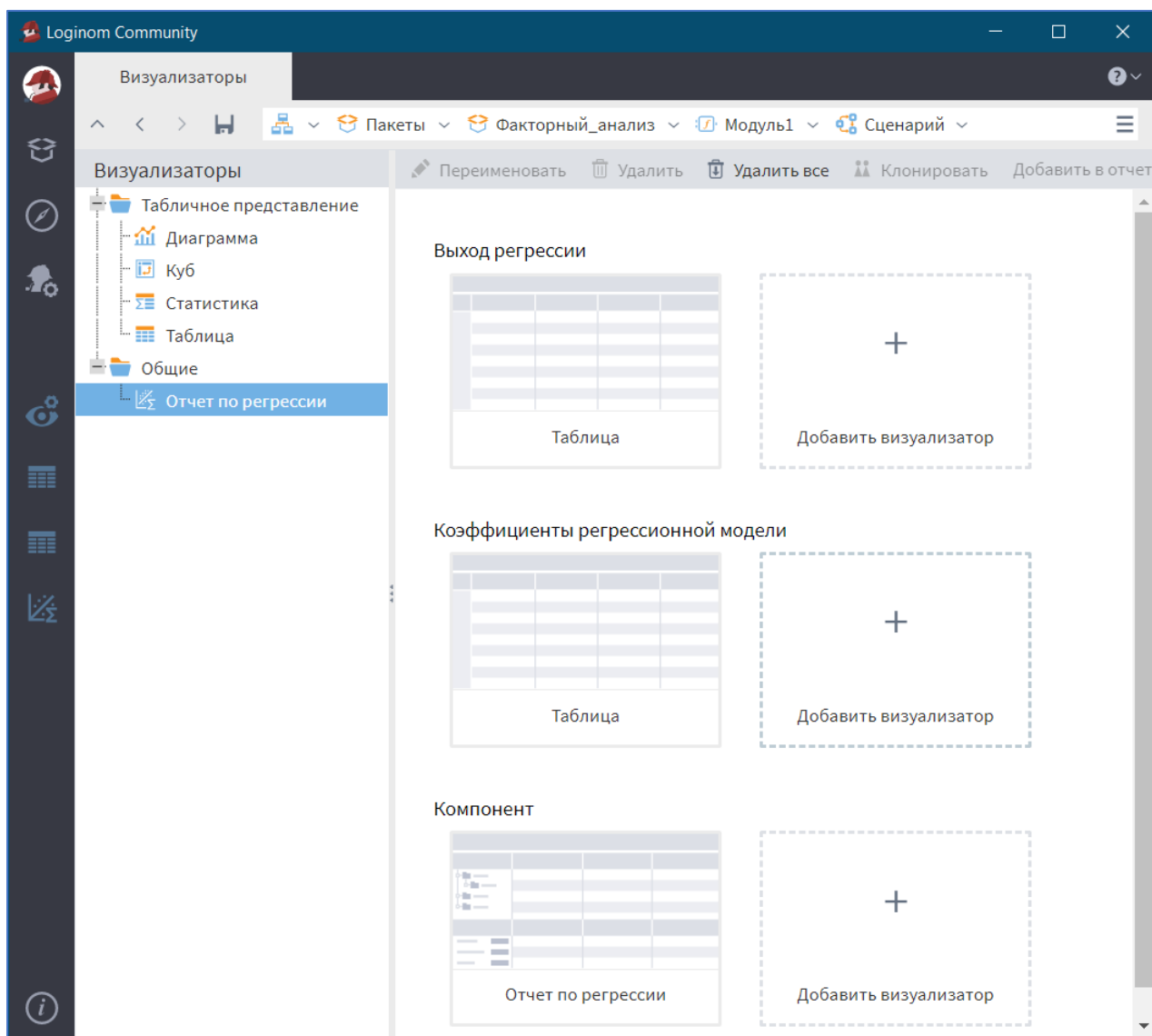


Рис. 8.30

В визуализаторе *Выход регрессии* представлены расчетные значения результативного признака (рис. 8.31).

#	№ хозяйст...	Урожайность кар...	Фактор1	Фактор2	Фактор3
1	1	440,4	1,855	1,176	-1,098
2	2	312,9	-0,491	0,442	0,067
3	3	444,8	1,940	-1,017	1,979
4	4	341,1	-0,224	1,113	0,266
5	5	347,1	-0,282	1,913	-0,079

Рис. 8.31

В визуализаторе *Коэффициенты регрессионной модели* приведены регрессионные коэффициенты, их интервальная оценка и уровень статистической значимости (рис. 8.32).

#	ab	ab	Метки в...	ab	9.0	Урожайнос...	9.0	Урожайность картофеля с 1 га, ц Стд. откл.
1	<...				331,833333		2,184057829	
2	F...	Фактор1			54,94775767		2,184057829	
3	F...	Фактор2			16,57283125		2,184057829	
4	F...	Фактор3			11,74604165		2,184057829	

Рис. 8.32

Визуализатор *Отчет по регрессии* отображает параметры и результаты статистических тестов для анализа регрессионных моделей. Откроем данный визуализатор (рис. 8.33).

Модель	Фит	Знач...
Пок...		
Конс...	true	
-2 Lo...	287,4...	
R^2	0,957...	
R^2 к...	0,953...	
Числ...	32,00	
Знач...	0,000...	

Атрибут	Коэфф...	Стан...	Т-стат...	Значи...
9.0 Константа	331,833...	2,184...	151,93...	0,000000
9.0 Фактор1	54,947758	2,184...	25,158...	0,000000
9.0 Фактор2	16,572831	2,184...	7,588092	0,000000
9.0 Фактор3	11,746042	2,184...	5,378082	0,000007

Рис. 8.33

Коэффициент детерминации $R^2 = 0,957$ свидетельствует о высоком качестве регрессионной модели. Полученная модель объясняет около 95,7 % дисперсии результативной переменной. Данные дисперсионного анализа показывают, что модель статистически значима при заданном уровне значимости 0,05.

Построенная регрессионная модель имеет вид:

$$y = 331,833 + 54,948x_1 + 16,573x_2 + 11,746x_3.$$

8.3. Задания для самостоятельной работы

Задание 8.1

В файле *Задача 8.3. Высшее образование.xlsx* имеются данные распределения регионов Российской Федерации по обеспеченности высшими образовательными учреждениями. Для этого были отобраны пять показателей (рис. 8.34).

	A	B	C	D	E	F
1	Регион	1. Число образовательных организаций высшего образования и филиалов на начало учебного года (2015-2016 гг.) на 1 млн. чел. населения	2. Численность студентов, обучающихся по программам бакалавриата, специалитета, магистратуры на начало учебного года (2015-2016 гг.) на 10 тыс. чел. населения	3. Прием на обучение по программам бакалавриата, специалитета, магистратуры (2015 г.) на 10 тыс. чел. населения	4. Выпуск бакалавров, специалистов, магистров (2015 г.) на 10 тыс. чел. населения	5. Численность профессорско-преподавательского персонала образовательных организаций высшего образования на 1000 студентов на начало учебного года (2015-2016 гг.)
2	Белгородская область	9,7	342,6	81,3	95,5	50,7
3	Брянская область	15,5	281,4	59,5	83,2	41,7
4	Владимирская область	12,2	239,8	62,3	55,8	47,3
82	Республика Крым	5,8	238,1	65,5	38,3	67,0
83	г. Севастополь	16,8	365,4	98,6	55,3	71,3

Рис. 8.34

Требуется:

- 1) построить матрицу факторных нагрузок, используя метод главных компонент;
- 2) выделить главные факторы, определяющие уровень обеспеченности кадрами;
- 3) найти численные значения главных компонент для всех объектов наблюдения.

Задание 8.2

В файле *Задача 8.4. Услуги связи.xlsx* имеются данные об объеме услуг связи, оказанных населению, по Приволжскому федеральному округу РФ и факторах, оказывающих на него влияние (рис. 8.35).

	A	B	C	D	E	F	G
1	№ п/п	Регион	Объем услуг связи, оказанных населению, на одного жителя, руб. (y)	Наличие квартирных телефонных аппаратов сети общего пользования на 1000 человек населения (x ₁)	Число подключенных абонентских устройств подвижной радиотелефонной связи на 1000 человек населения (x ₂)	Число активных абонентов фиксированного широкополосного доступа к сети Интернет на 100 человек населения (x ₃)	Число активных абонентов подвижной радиотелефонной связи, использующих широкополосный доступ к сети Интернет на 100 человек населения (x ₄)
2	1	Республика Башкортостан	3951,0	135,4	1737,9	20,2	71,5
3	2	Республика Марий Эл	4240,8	138,6	1846,6	17,2	70,3
4	3	Республика Мордовия	4238,1	186,3	1590,6	17,9	58,9
14	13	Саратовская область	4845,9	155,4	1798,3	19,8	70,7
15	14	Ульяновская область	4509,5	170,6	1936,6	19,5	67,3

Рис. 8.35

Требуется, используя метод главных компонент, сократить число факторов, оказывающих влияние на объем услуг связи, и построить регрессионную модель зависимости объема услуг связи от выделенных факторов.