

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Пензенский государственный университет
архитектуры и строительства»
(ПГУАС)

М.С. Акимова, Н.Ю. Улицкая

СОВРЕМЕННЫЕ МЕТОДЫ СТАТИСТИЧЕСКОГО АНАЛИЗА КАДАСТРОВЫХ ДАННЫХ

Учебно-методическое пособие
для выполнения расчетно-графической работы
по направлению подготовки 21.04.02 «Землеустройство и кадастры»

Пенза 2016

УДК 330:347.235.11(075.8)

ББК 65.32-5я73

А39

Рекомендовано Редсоветом университета

Рецензент – доктор экономических наук, профессор,
зав. кафедрой «Экономика, организация и
управление производством» Б.Б. Хрусталева
(ПГУАС)

Акимова М.С.

А39 Современные методы статистического анализа кадастровых данных: учеб.-метод. пособие для выполнения расчетно-графической работы по направлению подготовки 21.04.02 «Землеустройство и кадастры» / М.С. Акимова, Н.Ю. Улицкая. – Пенза: ПГУАС, 2016. – 58 с.

Содержит рекомендации по приемам обработки данных, применяемых при выполнении государственного кадастрового учета, регистрации объектов недвижимости, осуществлении кадастровой и оценочной деятельности, а также отображены возможности применения прикладных программ на персональных компьютерах для обработки кадастровой информации.

Учебно-методическое пособие подготовлено на кафедре «Кадастр недвижимости и право» и предназначено для использования студентами, обучающимися по направлению подготовки 21.04.02 «Землеустройство и кадастры», при изучении дисциплины «Современные методы статистического анализа кадастровых данных».

© Пензенский государственный университет
архитектуры и строительства, 2016

© Акимова М.С., Улицкая Н.Ю., 2016

ВВЕДЕНИЕ

Земельно-кадастровая информация – это совокупность земельно-регистрационной, земельно-учетной, земельно-оценочной информации, характеризующей и отражающей процессы управления земельными ресурсами, перераспределения земель и связанные с ними вопросы права собственности на землю, вида разрешенного использования, хозяйственной деятельности собственников земли, землепользователей и арендаторов.

Земельно-кадастровая информация используется при предоставлении и изъятии земельных участков, определении размеров платы за землю, оценке хозяйственной деятельности и осуществлении других мероприятий, связанных с использованием и охраной земель. Работа с большим объемом земельно-кадастровых (статистических) данных требует правильного представления этих данных, а также знания статистических приемов для систематизации и обработки земельно-кадастровой информации, изучения массовых явлений.

Математическая обработка кадастровой информации представляет собой группировку и сводку результатов наблюдений, контроль данных и характеристику ошибок статистического наблюдения, построение статистических таблиц и графическое отображение информации. При этом используются методы выявления связей между статистическими показателями, обработка и изучение динамических рядов, вариации, применение корреляционно-регрессионного метода и другие.

Целью дисциплины «Современные методы статистического анализа кадастровых данных» является освоение студентами методологии статистического исследования: методов сбора, упорядочения, обобщения, оценки достоверности и анализа массовых данных с целью выявления закономерностей и изучения взаимосвязей между явлениями.

Освоение дисциплины направлено на приобретение знаний и формирования у студентов навыков по применению математических методов для анализа и обработки земельно-кадастровой информации.

Задачи дисциплины:

- получение теоретических знаний, изучение статистических методов и приемов обработки данных;
- обоснование научно-технических и организационных решений;
- сбор и анализ исходных данных для схем и проектов землеустройства, планирования использования земель, проектов развития объектов недвижимости;
- изучение прикладных программ для обработки земельно-кадастровой информации;
- ознакомление с методами и приемами обработки данных, применяемыми при управлении земельными ресурсами, ведении кадастровой деятельности, осуществления землеустройства.

– апробация автоматизированных систем проектирования, обработки кадастровой и другой информации, их анализ.

Процесс изучения дисциплины направлен на формирование ряда **общекультурных и профессиональных компетенций:**

– способность к абстрактному мышлению, анализу, синтезу;

– готовность к коммуникации в устной и письменной формах на государственном языке Российской Федерации и иностранном языке для решения задач профессиональной деятельности;

– способность ставить задачи и выбирать методы исследования, интерпретировать и представлять результаты научных исследований в форме отчетов, рефератов, публикаций и публичных обсуждений.

Учебно-методическое пособие подготовлено для студентов факультета «Управление территориями», обучающихся по направлению подготовки 21.04.02 «Землеустройство и кадастры», с целью оказания практической помощи студентам при выполнении расчетно-графической работы по дисциплине «Современные методы статистического анализа кадастровых данных».

Рабочей программой дисциплины предусмотрена расчетно-графическая работа по следующим тематикам:

1. Анализ и прогноз земельных ресурсов Пензенской области.

2. Анализ размеров земельных участков, предназначенных для строительства жилья в муниципальном районе.

3. Отбор факторов, влияющих на стоимость земли.

В учебно-методическом пособии излагаются методы и приемы обработки данных, применяемые при управлении земельными ресурсами, кадастре и землеустройстве, в оценочной деятельности, а так же основные требования по представлению земельно-кадастровой информации и возможность применения прикладных программ для обработки кадастровой информации.

1. АНАЛИЗ И ПРОГНОЗ ИСПОЛЬЗОВАНИЯ ЗЕМЕЛЬНЫХ РЕСУРСОВ ПЕНЗЕНСКОЙ ОБЛАСТИ

Задание: Исходными данными для выполнения работы по анализу и прогнозу использования земельных ресурсов является характеристика земель Пензенской области, представленная в динамике с 2008 по 2013 года в разрезе категорий земель (табл. 1).

Т а б л и ц а 1

Динамика распределения земель Пензенской области по категориям за период с 2008 по 2013 года, га

Категория земель	Годы					
	2008	2009	2010	2011	2012	2013
1. Земли сельскохозяйственного назначения	3074,9	3074,2	3074,1	3073,5	3073,3	3072,9
2. Земли населенных пунктов	227,3	227,5	227,7	228,0	228,2	228,2
3. Земли промышленности, энергетики транспорта, связи, радиовещания, телевидения, информатики, земли для обеспечения космической деятельности, земли обороны, безопасности и земли иного специального назначения	42,1	42,7	42,8	43,1	43,1	43,5
4. Земли особо охраняемых территорий и объектов	9,1	9,1	9,1	9,1	9,1	9,1
5. Земли лесного фонда	964,5	964,5	964,5	964,5	964,5	964,5
6. Земли водного фонда	14,8	14,8	14,8	14,8	14,8	14,8
7. Земли запаса	2,5	2,4	2,2	2,2	2,2	2,2
Итого земель	4335,2	4335,2	4335,2	4335,2	4335,2	4335,2

Основной целью работы является изучение приёмов систематизации и обработки земельно-кадастровой информации, а так же анализ использования земельных ресурсов Пензенской области с применением данных приёмов и разработка прогноза методом экстраполяции.

1.1. Характеристика использования земельных ресурсов Пензенской области

Любая информация должна быть представлена так, чтобы ей удобно было пользоваться. Существует несколько способов представления информации: текстовый, табличный, графический.

Наиболее эффективным способом представления больших объемов земельно-кадастровых (статистических) данных являются таблицы.

Графический метод используется как в дополнение к таблицам, так и самостоятельно. С помощью графиков достигается наглядность взаимосвязи явлений, характеристики структуры и динамики.

Используя текстовый, табличный и графический способы представления информации, возможно, провести полноценный анализ распределения земель Пензенской области в разрезе категорий земель, представив её в табличной форме (табл. 2) и графически (рис. 1) в виде круговой диаграммы.

Т а б л и ц а 2

Распределение земель Пензенской области по категориям

Категории земель	Площадь	
	2013 г.	
	тыс. га.	%
1. Земли сельскохозяйственного назначения	3072,9	70,88
2. Земли населенных пунктов	228,2	5,26
3. Земли промышленности, энергетики транспорта, связи, радиовещания, телевидения, информатики, земли для обеспечения космической деятельности, земли обороны, безопасности и земли иного специального назначения	43,5	1
4. Земли особо охраняемых территорий и объектов	9,1	0,21
5. Земли лесного фонда	964,5	22,25
6. Земли водного фонда	14,8	0,35
7. Земли запаса	2,2	0,05
Итого земель:	4335,2	100

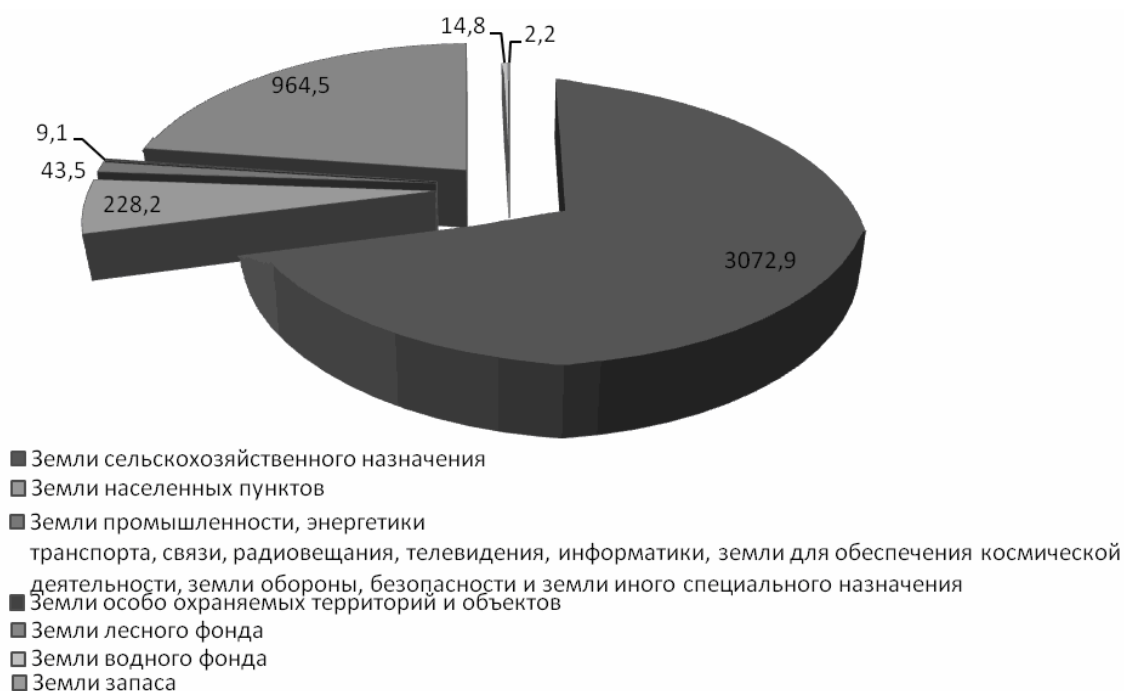


Рис.1. Распределение земель Пензенской области по категориям на 01.01.2014 года

Анализируя представленные данные таблицы 2 и диаграмму, можно сделать вывод, что наибольшая доля в структуре земельных ресурсов области приходится на земли сельскохозяйственного назначения, а именно 3072,9 тыс. га или 70,88 %. Наименьшую площадь занимают земли запаса – 2,2 га или 0,05%.

П р и м е ч а н и е . При построении таблиц следует придерживаться следующих правил:

– цифровые данные таблиц записываются с одной и той же точностью в пределах каждой графы, при этом разряд располагается под разрядом.

– если в клетке нет информации – ставится прочерк, если данные не известны – «сведения отсутствуют».

1.2. Анализ перераспределения земельных ресурсов Пензенской области по видам целевого использования

Одно из основных положений всех исследований – необходимость изучения всех явлений (объектов) в развитии, во времени, т.е. изучение их динамики.

Изучение динамики того или иного явления (объекта) начинается с построения ряда динамики, или временного ряда.

Динамический ряд представляет собой числовые значения определенно-го показателя (y) за последовательные периоды или моменты времени (t). Каждое значение показателя называется уровнем ряда.

Каждый динамический ряд состоит из n -го числа варьирующихся во времени показателей: первый показатель называют начальным уровнем (y_0, y_1); последний – конечным (y_n).

Ряды динамики могут отображать различные процессы и изменения.

Так, например, изменение площади каждой категории земель области в период с 2008 по 2013 годы будет представлять из себя отдельный динамический ряд (табл. 3). В этом случае 2008 год является начальным уровнем ряда, а 2013 год – конечным.

Для наглядности отображения динамического ряда его удобно представлять в виде маркированного графика (рис. 2). Отображая на одном рисунке несколько динамических рядов (рис. 5), можно выявить взаимосимости изменений площадей отдельных категорий земель в области.

Т а б л и ц а 3

Динамика распределения земель Пензенской области
по категориям за период с 2008 по 2013 года, тыс. га

Категория земель	Годы					
	2008	2009	2010	2011	2012	2013
1. Земли сельскохозяйственного назначения	3074,9	3074,2	3074,1	3073,5	3073,3	3072,9
2. Земли населенных пунктов	227,3	227,5	227,7	228,0	228,2	228,2
3. Земли промышленности, энергетики транспорта, связи, радиовещания, телевидения, информатики, земли для обеспечения космической деятельности, земли обороны, безопасности и земли иного специального назначения	42,1	42,7	42,8	43,1	43,1	43,5
4. Земли особо охраняемых территорий и объектов	9,1	9,1	9,1	9,1	9,1	9,1
5. Земли лесного фонда	964,5	964,5	964,5	964,5	964,5	964,5
6. Земли водного фонда	14,8	14,8	14,8	14,8	14,8	14,8
7. Земли запаса	2,5	2,4	2,2	2,2	2,2	2,2
Итого земель	4335,2	4335,2	4335,2	4335,2	4335,2	4335,2

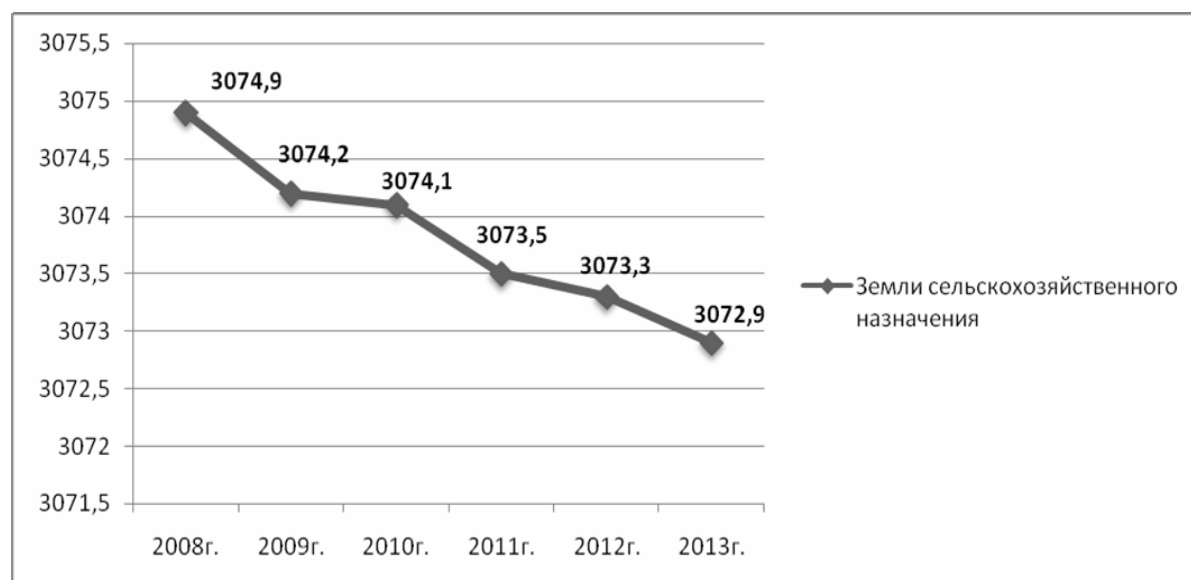


Рис. 2. График изменения площади земель сельскохозяйственного назначения за период 2008–2013 гг.

Процессы изменения показателя во времени, отображают не только графически (рис. 2), но и через числовые показатели, сравнивая значения

начального уровня ряда с конечным (табл. 4). Изменения, происходящие за период исследования, можно определить по формулам:

$$y_n - y_1, \quad (1)$$

$$\frac{y_n}{y_1} \cdot 100 \%, \quad (2)$$

где y_n – конечный уровень ряда (площадь земель в 2013 году), га;
 y_1 – начальный уровень ряда (площадь земель в 2008 году), га.

Так, для земель сельскохозяйственного назначения изменения площади в 2013 году по отношению к 2008 году составят:

$$3072,9 - 3074,9 = -2 \text{ тыс. га};$$

$$T_p = \frac{3072,9}{3074,9} \cdot 100 = 99,9 \%.$$

Т а б л и ц а 4

Изменения распределения земель области в 2013 году
по отношению к 2008 году

Категории земель	Площадь тыс. га.					
	2008 г.		2013 г.		Изменения 2013 г. к 2008 г.	
	тыс.га	%	тыс.га	%	тыс.га	%
1. Земли сельскохозяйственного назначения	3074,9	70,93	3072,9	70,89	-2	99,9
2. Земли населенных пунктов	227,3	5,24	228,2	5,26	+0,9	100,4
3. Земли промышленности, энергетики транспорта, связи, радиовещания, телевидения, информатики, земли для обеспечения космической деятельности, земли обороны, безопасности и земли иного специального назначения	42,1	0,97	43,5	1	+1,4	103,3
4. Земли особо охраняемых территорий и объектов	9,1	0,21	9,1	0,21	0	100
5. Земли лесного фонда	964,5	22,25	964,5	22,25	0	100
6. Земли водного фонда	14,8	0,34	14,8	0,34	0	100
7. Земли запаса	2,5	0,06	2,2	0,05	-0,3	88
Итого земель:	4335,2	100	4335,2	100	0	100

Наглядно изменения, произошедшие за период времени, можно отобразить столбчатой и линейчатой диаграммами (рис. 3 и 4):

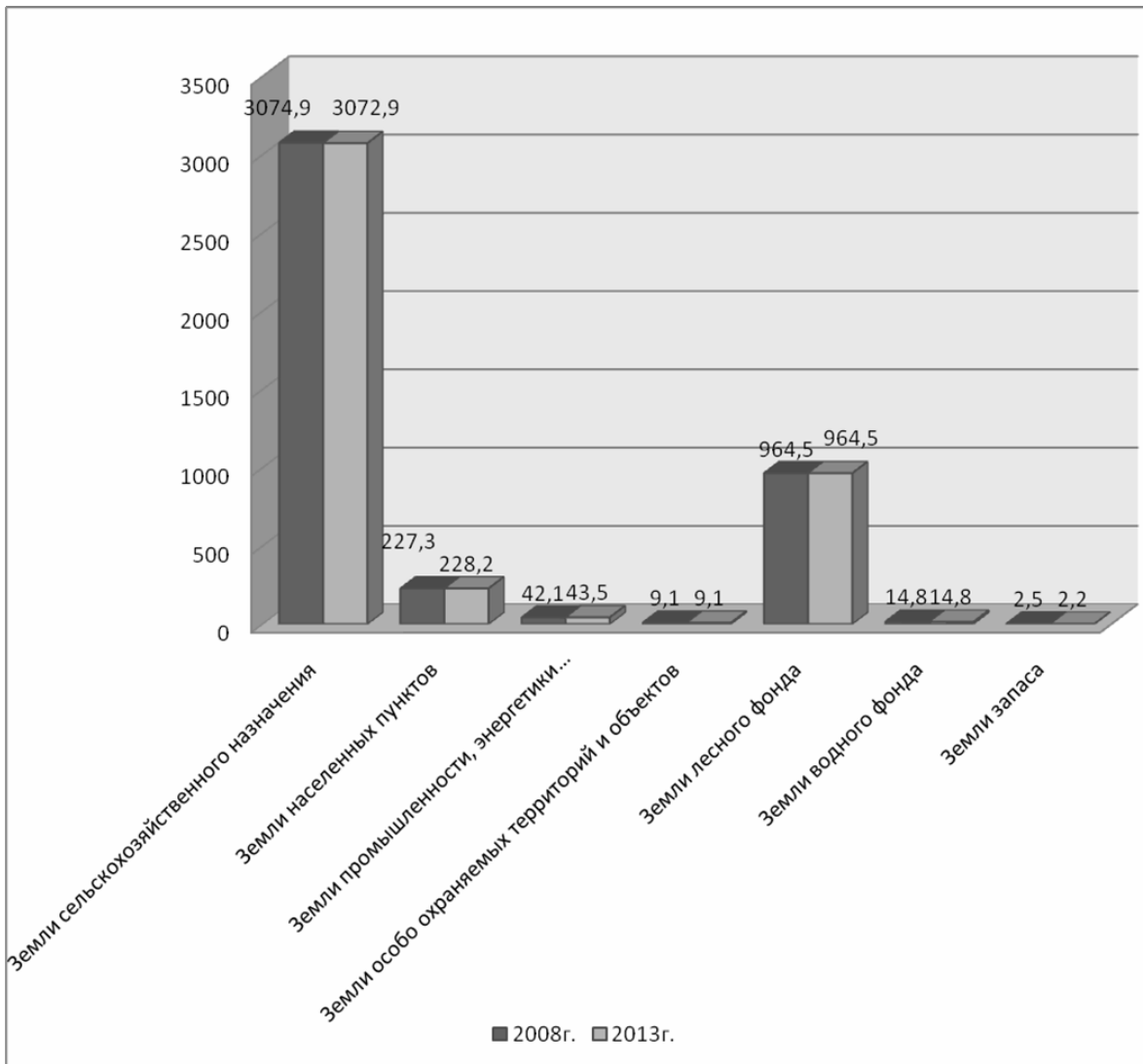


Рис. 3. Сопоставление распределения категорий земель области в 2013 и 2008 годах, тыс. га

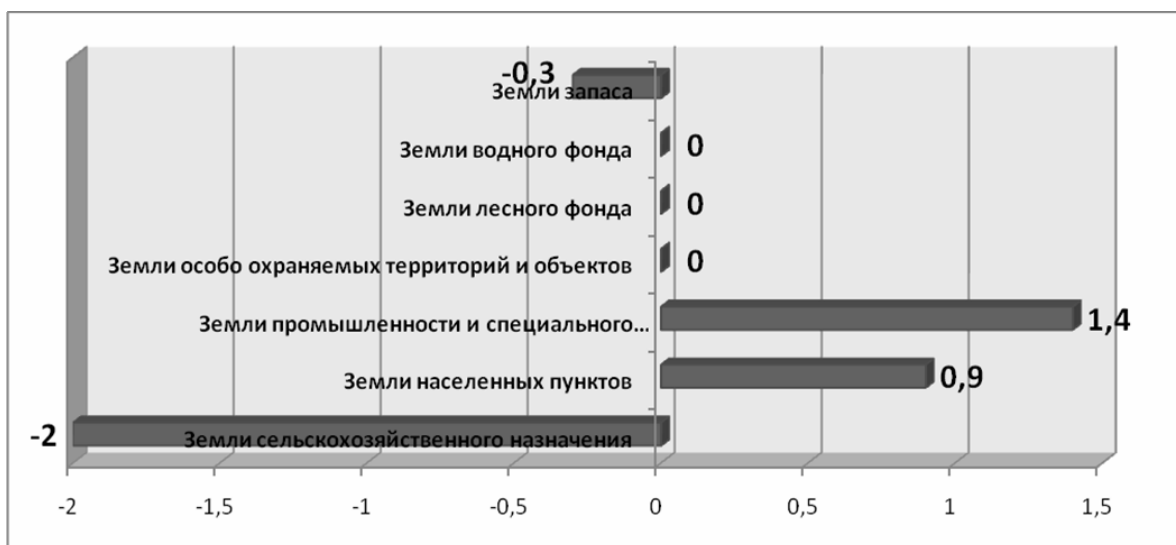


Рис. 4. Изменение структуры категорий земель области в 2013 году по отношению к 2008 году, тыс. га

Представленные расчеты и графики свидетельствуют, что наибольшие изменения по площади в границах области произошли с землями промышленности и землями запаса, первые увеличились на 1,4 тыс. га (103,3 %), вторые уменьшились на 0,3 тыс. га (88 %). По другим видам использования земель произошли не значительные изменения или не произошли совсем.

Изменения площадей по двум выше названным видам можно представить на одном графике, рис. 5.

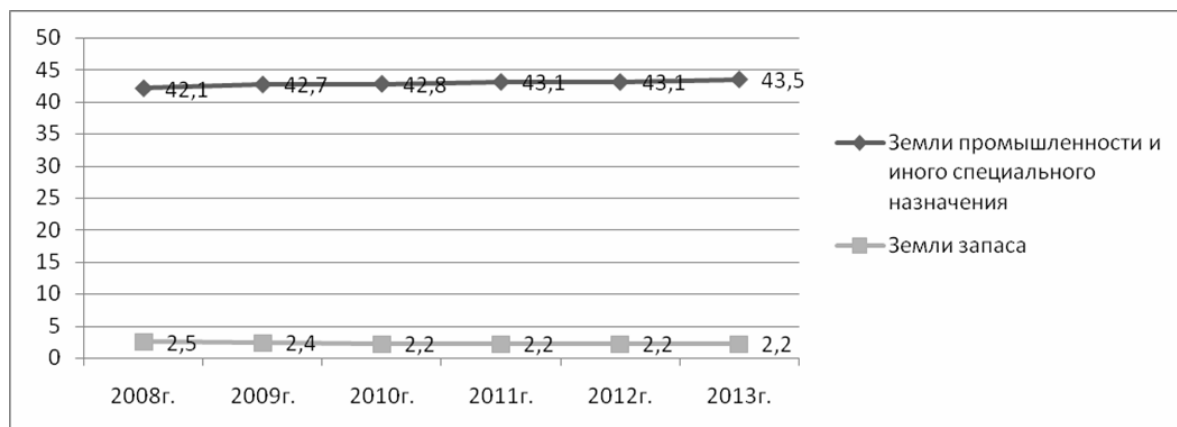


Рис. 5. Сопоставление динамики изменения площади земель промышленности и земель запаса в области за период 2008-2013 гг.

Анализируя графики, представленные на рис. 5 видно, что между изменением площади земель промышленности и земель запаса существует ярко выраженная зависимость – с увеличением площади одного вида использования (площади земель промышленности), определенным образом уменьшается площадь земель другого вида (земли запаса).

При изучении рядов динамики важно проследить так же и за направлением и размером изменения уровней ряда динамики во времени. С этой целью рассчитываются показатели *темпа роста*, *абсолютного прироста* и *темпа прироста*. В качестве обобщенной характеристики динамического ряда выступает показатель среднего уровня ряда или средняя хронологическая (табл. 5).

Расчет средней хронологической для моментного ряда, га, производится по формуле

$$\bar{y} = \frac{y_1 + y_n + \sum_{i=2}^{n-1} y_i}{n - 1}, \quad (3)$$

где y_i – отдельные уровни ряда;

y_n – конечный уровень ряда (площадь земель в 2013 году), тыс. га;

y_1 – начальный уровень ряда (площадь земель в 2008 году), тыс. га.

n – число уровней.

Показатели, отображающие направление и размер изменения уровней во времени вычисляются по следующим формулам (2, 4, 5):

$$\Delta y = \frac{y_n - y_1}{n - 1}, \text{ га}; \quad (4)$$

$$T_{\text{пр}} = T_p - 100, \%, \quad (5)$$

где: T_p – темп роста;

Δy – абсолютный прирост;

$T_{\text{пр}}$ – темп прироста;

y_n – конечный уровень ряда (площадь земель в 2013 году), тыс. га;

y_1 – начальный уровень ряда (площадь земель в 2008 году), тыс. га;

n – число уровней.

Так, для земель сельскохозяйственного назначения значение средней хронологической составит:

$$y = \frac{\frac{3072,9 + 3074,9}{2} + 12295,1}{6 - 1} = 3073,8 \text{ тыс.га.}$$

Показатели, отображающие направление и размер изменения уровней во времени для земель сельскохозяйственного назначения, составят:

$$T_p = \frac{3072,9}{3074,9} \cdot 100 = 99,9 \%;$$

$$\Delta y = \frac{3072,9 - 3074,9}{6 - 1} = -0,4 \text{ тыс.га};$$

$$T_{\text{пр}} = 99,9 - 100 = -0,1\%.$$

Т а б л и ц а 5

Характеристика изменения структуры земельного фонда

Наименование категории	Средняя хронологическая, тыс. га, \bar{y}	Темп роста, %, T_p	Абсолютный прирост, тыс. га Δy	Темп прироста, %, $T_{\text{пр}}$
1	2	3	4	5
1. Земли сельскохозяйственного назначения	3073,8	99,9	-0,4	-0,1
2. Земли населенных пунктов	227,83	100,4	0,18	0,4

Окончание табл. 5

1	2	3	4	5
3. Земли промышленности, энергетики транспорта, связи, радиовещания, телевидения, информатики, земли для обеспечения космической деятельности, земли обороны, безопасности и земли иного специального назначения	42,9	103,3	0,28	3,3
4. Земли особо охраняемых территорий и объектов	9,1	100	0	0
5. Земли лесного фонда	964,5	100	0	0
6. Земли водного фонда	14,8	100	0	0
7. Земли запаса	2,27	88	-0,06	-18
Итого земель	4335,2	100	0	0

Анализ изменений структуры категорий земель области, показал, что наибольший темп роста наблюдается по землям промышленности (103,3 %) и землям населенных пунктов (100,4 %). Наибольшие темпы убыли отмечаются для земель сельскохозяйственного использования (99,9 %), земель запаса (88%), соответственно темпы прироста составляют 3,3%, 0,4%, -0,1 %, -18 %.

Абсолютные показатели ежегодного изменения площадей использования составляют: для земель промышленности ежегодный абсолютный прирост равен 0,28 тыс. га, земель населенных пунктов 0,18 тыс. га, абсолютный показатель ежегодной убыли для земель сельскохозяйственного использования равен - 0,4 тыс. га, земель промышленности - 0,06 тыс. га.

1.3. Прогноз распределения земель Пензенской области методом экстраполяции

Для определения возможных значений признаков в будущем строятся различные прогнозы. В решении этой задачи роль статистических методов очень значительна. Одним из таких методов является расчет прогнозов на основе тренда и колеблемости динамического ряда. Методика такого статистического прогноза базируются на экстраполяции (продлении) рядов динамики.

Экстраполяцию можно осуществить различными способами. Но, независимо от способа, сущность метода заключается в предположении о том,

что закономерности или тенденции изменения изучаемого объекта (явления) в прошлом сохраняются на определенном отрезке времени в будущем.

Считается, что наиболее точный прогноз методом экстраполяции получается на срок, не превышающий $1/3$ базисного периода. Поэтому прогноз методом экстраполяции всегда базируется на тщательном изучении длительных рядов динамики, которые позволяют выявить долгосрочную тенденцию изменения в достаточно стабильных условиях. При этом период прогнозирования не должен превышать базисный период. В действительности тенденция развития может меняться, поэтому данные, получаемые путем экстраполирования, рассматриваются как вероятностные, что требует оценки ошибки метода (табл. 6) и определения доверительных интервалов (табл. 7).

Одним из наиболее распространенных способов экстраполяции рядов динамики является их выравнивание по аналитической формуле, то есть подбор функции, отображающей общую тенденцию изменения. Зная, какая функция отображает тенденцию изменения и, подставляя в нее значения X за пределами динамического ряда, можно рассчитать теоретический уровень с определенной степенью вероятности.

Прогноз использования земель области (табл. 7) разрабатывается на основании данных табл. 1, для каждой категории земель. Подбор функции осуществляется с помощью программы «Excel» путем построения графика с добавлением линии тренда (рис. 6) для каждой категории земель области. Кроме самой линии тренда, диаграмма должна содержать уравнение линии тренда и величину достоверности аппроксимации – коэффициент детерминации $D=R^2$.

Если коэффициент детерминации (R^2) больше 0,70, то функция пригодна для практического применения. Если данное условие не выполняется, необходимо продолжить подбор уравнения.

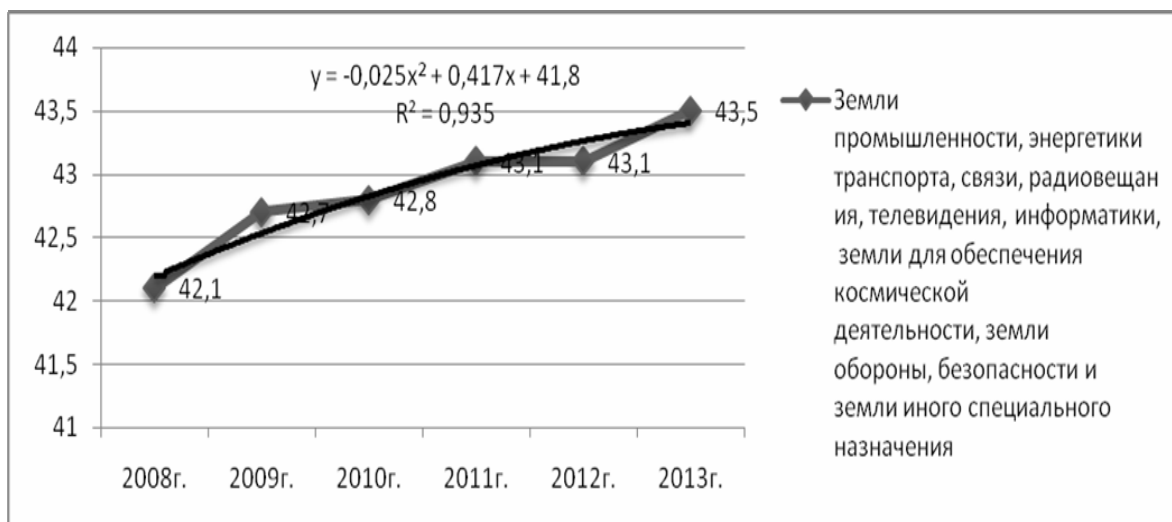


Рис. 6. Подбор функции для прогноза земель транспорта, связи, инженерных коммуникаций методом экстраполяции

Общий вид функции может быть следующим:

- линейная

$$Y_T = A + BX; \quad (6)$$

- полиномиальная

$$Y_T = A + BX + CX^2; \quad (7)$$

- логарифмическая

$$Y_T = A + \ln(X); \quad (8)$$

- степенная

$$Y_T = A \cdot X^B; \quad (9)$$

- экспоненциальная

$$Y_T = A \cdot e^{B \cdot X}, \quad (10)$$

где A – начальный уровень показателя;

B, C – темпы изменения этого показателя;

X – индекс года прогнозирования;

Y_T – теоретический уровень на прогнозируемый период.

Для установления точности прогноза необходимо рассчитать доверительный интервал:

$$Y_d = Y_T \pm M \quad (11)$$

и интервальную оценку:

$$M = G \cdot C, \quad (12)$$

где G – критерий Стьюдента (прил. А);

C – ошибка метода.

Расчет ошибки метода производится по формуле

$$C = \sqrt{\frac{\sum (Y_\phi - Y_T)^2}{n - p - 1}}, \quad (13)$$

где Y_ϕ – фактический уровень показателя;

Y_T – теоретический уровень показателя;

n – количество наблюдений;

p – количество параметров функции (A, B, C).

Теоретический уровень площади определяется по функции, подобранной методом экстраполяции.

Т а б л и ц а 6

Расчет оценки ошибки метода

Уровни динамического ряда	Индекс года	Фактическое значение Y_{ϕ}	Теоретическое значение Y_T	$Y_{\phi} - Y_T$	$(Y_{\phi} - Y_T)^2$	C
2008	1	42,1	42,2	-0,1	0	
2009	2	42,7	42,5	0,2	0,04	
2010	3	42,8	42,8	0	0	
2011	4	43,1	43,1	0	0	
2012	5	43,1	43,3	-0,2	0	
2013	6	43,5	43,4	0,1	0,01	
					0,05	0,025

Таблица 6 рассчитывается для каждой категории земель области, по которой выполняется расчет прогнозной площади.

По результатам подбора функции и расчета оценки ошибки метода для каждой категории земель осуществляется расчет прогнозных площадей на определенный год прогнозирования, одновременно с определением доверительных интервалов (таблица 7); производится общий анализ результатов прогнозирования.

Подобранная функция для прогноза земель промышленности и иного специального назначения методом экстраполяции имеет полиномиальный вид: $Y_T = A + BX + CX^2$, параметры которой равны: $A = 41,8$; $B=0,417$; $C=-0,025$; X – это индекс года прогнозирования:

– для первого периода прогнозирования в 5 лет или для 2018 года, индекс будет равен 11;

– для второго периода в 10 лет или для 2023 года, индекс будет равен 16.

$$Y_{2018} = 41,8 + 0,417 \cdot 11 - 0,025 \cdot 11^2 = 43,4;$$

$$Y_{2023} = 41,8 + 0,417 \cdot 16 - 0,025 \cdot 16^2 = 42,1.$$

По всем другим категориям земель области проводится аналогичный расчет в зависимости от выбранной математической функции. Результаты расчетов сводятся в таблицу 7.

Анализ показал, что при сложившихся тенденциях площадь земель промышленности и иного специального назначения в области будет уменьшаться, и теоретическое значение данного показателя к 2018 году может составить 43,4 га, к 2023 – 42,1 га. Интервальная оценка составит 0,06 тыс. га, следовательно, фактическое значение площади земель промышленности и иного специального назначения будет находиться в интервале:

на первый период прогнозирования – 2018 год от 43,34 тыс. га до 43,46 тыс. га;

на второй период прогнозирования – 2023 год от 42,04 тыс. га до 42,16 тыс. га.

Проведенный прогноз возможного изменения структуры земель промышленности и иного специального назначения области на ближайших 5-10 лет должен выявить положительные и отрицательные тенденции, помочь сделать правильные выводы и принять оптимальные управленческие решения по распоряжению земельными ресурсами.

Контрольные вопросы

1. Формы предоставления информации.
2. Отличительные свойства математических таблиц от статистических.
3. Требования по оформлению статистических таблиц.
4. Виды статистических таблиц.
5. Что такое подлежащее в статистической таблице.
6. Что такое сказуемое в статистической таблице.
7. Отличительные особенности простой таблицы от сложной.
8. Отличительные особенности сложной таблицы от комбинационной.
9. Что характеризует ряд динамики.
10. Как определяется средний уровень интервального ряда динамики.
11. Как определяется средняя хронологическая.
12. Как исчисляется абсолютный прирост.
13. Как исчисляется темп роста.
14. Что представляют собой изменение ряда динамики.
15. Какие способы используются для выявления основной тенденции развития.
16. В каких случаях возникает несопоставимость уровней ряда.

2. АНАЛИЗ РАЗМЕРОВ ЗЕМЕЛЬНЫХ УЧАСТКОВ ПОД СТРОИТЕЛЬСТВО ЖИЛЬЯ В МУНИЦИПАЛЬНЫХ РАЙОНАХ

Задание: Исходными данными для выполнения работы является статистическая информация о размерах земельных участков, предоставленных для строительства жилья в 4 муниципальных районах Пензенской области, объем выборки составляет десять земельных участков в каждом районе (табл. 8).

Основной целью работы является освоение приёмов изучения массовых явлений, а также основ статистической группировки.

2.1. Характеристика и анализ размеров земельных участков, предназначенных для строительства жилья в муниципальных районах Пензенской области

Процесс использования и перераспределения земли, как и любой другой процесс или явление, обладает как общими свойствами и признаками, так и особенными, индивидуальными свойствами.

Различие значений признака у разных единиц совокупности в один и тот же период или момент времени называется вариацией этого признака.

Причиной вариаций являются разные условия существования различных единиц совокупности. Вариация присуща всем без исключения массовым явлениям природы и общества. Так, например, размеры земельных участков, предоставляемых населению для строительства, имеют различия как в целом по России, так и в конкретных субъектах Российской Федерации, в муниципальных районах и в отдельных населенных пунктах (табл. 8). Свойства массовых явлений проявляются как близостью характеристик отдельных явлений и их однородностью, так и определенными различиями.

Т а б л и ц а 8

Размер земельных участков, предназначенных для строительства жилья
в муниципальных районах

Наименование муниципального района	Численность населения, чел.	Номер участка	Площадь, м ²	Наименование муниципального района	Численность населения, чел.	Номер участка	Площадь, м ²
1	2	3	4	5	6	7	8
Сосновоборский район	16800	1	1998	Каменский район	58600	1	1200
		2	1999			2	1375
		3	1690			3	1500
		4	1746			4	1375
		5	1888			5	1650
		6	1871			6	1650
		7	2000			7	1500
		8	1659			8	1500
		9	1875			9	1766
		10	1875			10	1375

Окончание табл. 8

1	2	3	4	5	6	7	8
Кузнецкий район	37691	1	1650	Мокшанский район	27900	1	1375
		2	1500			2	1500
		3	1500			3	1500
		4	1500			4	1888
		5	1050			5	1375
		6	1200			6	1050
		7	1500			7	1050
		8	1375			8	1050
		9	1375			9	2000
		10	2000			10	1500

Вариация – это необходимое условие существования и развития массовых явлений. Однако для каждого массового процесса, явления, необходимо так же соблюдение определенных условий – «допусков» – т.е. специфическая мера вариации их элементов, при которой данный процесс или явление будет существовать оптимально.

Информацию о массовых явлениях можно представить различными способами, например, в виде вариационных рядов. Существует три способа предоставления вариационных рядов: ранжированный, дискретный, интервальный.

Первый способ — это представление информации в виде ранжированного ряда. Ранжированный вариационный ряд – это перечень отдельных единиц совокупности в порядке возрастания или убывания изучаемого признака. Для представления информации по земельным участкам муниципального района в виде ранжированного вариационного ряда необходимо просто упорядочить все значения в порядке возрастания, табл. 9.

Т а б л и ц а 9

Ранжированный ряд

№ п/п	Площадь, м ²	№ п/п	Площадь, м ²	№ п/п	Площадь, м ²	№ п/п	Площадь, м ²
1	1050	11	1375	21	1500	31	1874
2	1050	12	1375	22	1500	32	1875
3	1050	13	1375	23	1500	33	1875
4	1050	14	1500	24	1650	34	1888
5	1200	15	1500	25	1650	35	1888
6	1200	16	1500	26	1650	36	1998
7	1375	17	1500	27	1659	37	1999
8	1375	18	1500	28	1690	38	2000
9	1375	19	1500	29	1746	39	2000
10	1375	20	1500	30	1766	40	2000

Дискретный вариационный ряд – это таблица, где даны конкретные значения варьирующего признака x_i и число единиц совокупности f_i с данным значением признака (f_i – частота проявления i -го значения признака x).

Так, например, значения площади земельных участков, предоставляемых для строительства в муниципальном районе – конкретные значения варьирующего признака x_i , а количество земельных участков с такой площадью – число единиц совокупности с данными значениями признака f_i (табл. 10). Из таблицы видно, что в данных районах наибольшее количество земельных участков имеют размер 1500 м² и 1375 м², небольшое количество земельных участков с площадью 1050 м², 1200 м², 1650 м², 1875 м², 1888 м² и 2000 м². Все остальные земельные участки – это единичные случаи, встречающиеся на территории муниципальных районов.

Т а б л и ц а 10

Дискретный вариационный ряд по площади земельных участков, предназначенных для строительства в муниципальных районах

Размер земельного участка под строительство, м ² , x_i	Количество участков, шт., f_i
1050	4
1200	2
1375	7
1500	10
1650	3
1659	1
1690	1
1746	1
1766	1
1874	1
1875	2
1888	2
1998	1
1999	1
2000	3
Σ	40

Интервальный вариационный ряд – представляет собой таблицу, которая характеризуется величиной интервала варьирующего признака x_i и количеством единиц совокупности попадающих в тот или иной интервал f_i .

Для построения интервального вариационного ряда необходимо определить оптимальное число групп или интервалов и установить длину или шаг интервала. Оптимальное число групп выбирается так, чтобы в достаточной мере отобразить разнообразие значений признаков и в то же время

закономерность распределения и его форма не искажалась случайными колебаниями. Если групп будет слишком мало, не проявится закономерность вариации; если слишком много – случайные скачки частот исказят форму распределения. Число групп должно быть целым, табл. 11.

Т а б л и ц а 11

Характеристика вариационной совокупности

Число наблюдений n	Минимальное значение совокупности x_{\min}	Максимальное значение совокупности x_{\max}	Вариационный размах R	Количество интервалов K	Шаг интервала d
40	1050	2000	950	7	150

Чаще всего число групп в вариационном ряду рассчитывают по формуле Стержесса:

$$k = 1 + 3,321 \cdot \lg(n), \quad (14)$$

где n – число наблюдений в совокупности.

Далее определяется величина интервала (шаг интервала):

$$d = \frac{x_{\max} - x_{\min}}{k}, \quad (15)$$

где d – величина интервала;

$x_{\max} - x_{\min}$ – вариационный размах R .

Таким образом, исходную совокупность данных по земельным участкам, предназначенных для строительства в муниципальных районах необходимо разделить на 8 интервалов с шагом 75 м^2 (табл. 12). Кроме того, полезно определить середину интервала и удельный вес группы относительно всей исследуемой совокупности.

При построении интервалов одно и тоже значение признака выступает и верхней и нижней границей. При таком обозначении может возникнуть вопрос – в какую группу включать единицу объекта, значение которого совпадает с границами интервала.

Если нижняя граница формируется по принципу «включительно», а верхняя «исключительно», то объект со значением, находящимся на границе интервала включается во вторую группу. Если верхняя граница формируется по принципу «включительно», а нижняя «исключительно», то объект со значением, находящимся на границе интервала включается в первую группу.

Т а б л и ц а 12

Характеристика интервального вариационного ряда земельных участков, предназначенных для строительства в муниципальных районах

Группы земельных участков по площади	Число земельных участков в группе f_i	Середина интервала x_i	Удельный вес
1050–1200	4	1125	10
1200–1350	2	1275	5
1350–1500	7	1425	17,5
1500–1650	10	1575	25
1650–1800	7	1725	17,5
1800–1950	5	1875	12,5
1950–2100	5	2025	12,5
Σ	40		100

Анализируя табл. 12, можно увидеть, что 25 % земельных участков, предоставляемых для строительства в муниципальных районах, имеет площадь от 1500 до 1650 м², так же встречается значительное количество участков с площадью от 1350 до 1500 м² – 17,5 %, от 1650 до 1800 м² – 17,5 %. В двух вариационных интервалах (1800–1950; 1950–2100) удельный вес земельных участков составил 12,5 %. В вариационных интервалах 1200–1350; 1050–1200 имеются два и четыре земельных участка соответственно.

Существенную помощь в анализе вариационных рядов оказывает графическое изображение, т.е. построение гистограммы. Интервальный ряд изображается в виде столбчатой диаграммы, в которой на оси абсцисс расположены интервалы значений варьирующего признака, на оси ординат – частота признака. Так наличие засоренности в рассматриваемой совокупности наглядно отображается на гистограмме, рис. 7.

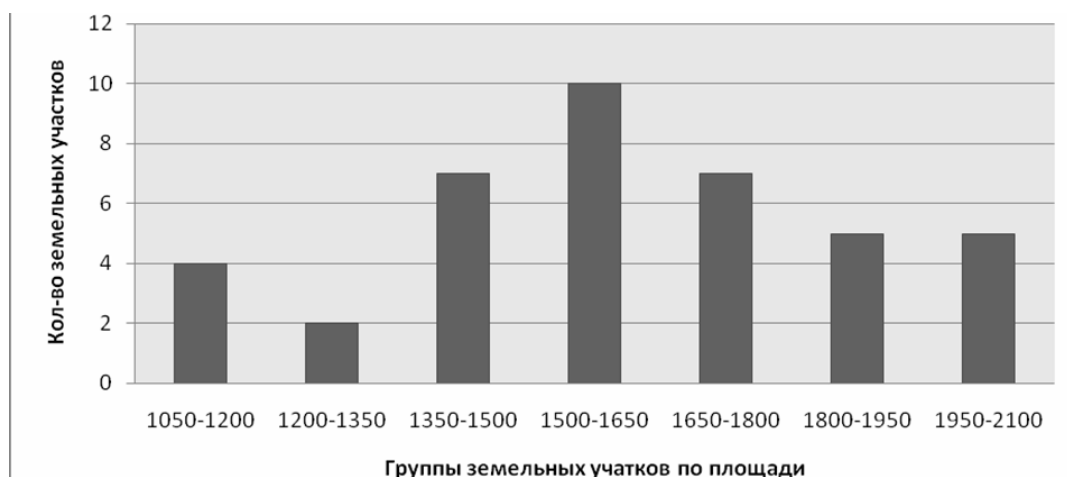


Рис. 7. Интервальный вариационный ряд земельных участков под строительство в муниципальных районах

По гистограмме можно наглядно определить близость распределения признака к нормальному распределению, наличие аномальных значений, так, например, два земельных участка с площадью 1200 м² является аномальным.

2.2. Определение структурных характеристик вариационного ряда

В процессе изучения вариации необходимо вычислять показатели, которые могут отобразить особенности распределения изучаемых признаков.

Все эти показатели можно разделить на 3 группы:

- показатели центра распределения (средняя, средняя арифметическая, мода, медиана);
- показатели степени вариации (вариационный размах, среднее линейное отклонение, дисперсия (среднеквадратическое отклонение));
- показатели типа распределения (структурные характеристики, показатели асимметрии, эксцесса, кривые распределения).

2.2.1. Показатели центра распределения

Для характеристики массового явления используют средние величины, такие как средняя арифметическая и средневзвешенная.

Главное значение средних величин состоит в обобщении функции, то есть замене множества индивидуальных значений признака средней величиной. Средняя величина отражает то общее, что присуще всем единицам совокупности.

Средняя арифметическая величина используется в том случае, если распределение признака достаточно однородное, и рассчитывается по формуле

$$\bar{x} = \frac{\sum x_i}{n}, \quad (16)$$

где $\sum x_i$ – сумма всех значений варьируемого признака в совокупности;
 n – количество единиц в совокупности.

Если совокупность представлена значениями, повторяемость которых различная, лучше использовать средневзвешенную величину. В качестве весов выступает показатель, который отражает число повторяющихся признаков (табл. 13).

Средневзвешенная величина рассчитывается по формуле

$$\bar{x} = \frac{\sum (x_i \cdot f_i)}{\sum f_i}, \quad (17)$$

где x_i – конкретное значение варьируемого признака;
 f_i – количество единиц совокупности с данным значением признака;
 n – количество единиц в совокупности.

Так, для совокупности данных по площади земельных участков, предназначенных для строительства средневзвешенная величина равна:

$$\bar{x} = \frac{62433}{40} = 1561 \text{ м}^2.$$

Кроме средней арифметической и средневзвешенной центр распределения характеризуют медиана и мода (табл. 13).

Медиана – величина варьирующего признака, делящая совокупность на 2 равные части – со значением признака меньше и больше медианы, находящейся в середине упорядоченного ряда.

В ранжированном вариационном ряду с нечетным числом признаков медиана вычисляется по формуле

$$M_e = \frac{n+1}{2}, \quad (18)$$

где n – число совокупности.

С четным числом признаков:

$$M_e = \frac{n}{2}. \quad (19)$$

Для определения численного значения медианы в интервальном ряду следует сначала указать интервал нахождения медианы в интервальном ряду распределения. Медианным называется первый интервал, где сумма накопленных частот превышает половину наблюдений от общего числа всех наблюдений.

Численное значение медианы в интервальном вариационном ряду определяется по формуле

$$M_e = x_0 + \frac{\sum f_i - S_{M_e-1}}{f_{M_e}} \cdot d, \quad (20)$$

где x_0 – низшая граница интервала, в котором находится медиана;

S_{M_e-1} – накопительная частота в интервале, предшествующем медианному;

f_{M_e} – частота медианного интервала;

d – величина интервала;

f_i – число наблюдений.

В вариационном ряду площади земельных участков, предназначенных для строительства в муниципальных районах медиана располагается в интервале от 1500 до 1650 м² (табл. 13). Численное значение, соответственно, равно:

$$M_e = 1500 + \frac{\frac{40}{2} - 13}{10} \cdot 150 = 1605 \text{ м}^2.$$

Мода – это величина признака, который в изучаемом ряду встречается чаще всего. В дискретном вариационном ряду мода определяется как значение признака с наибольшей частотой. В интервальном ряду для определения численного значения моды необходимо сначала определить модальный интервал – интервал, к которому относится наибольшее количество участков изучаемой совокупности. Само численное значение моды определяется по формуле

$$M_0 = x_0 + \frac{(f_{M_e} - f_{M_{e-1}})}{(f_{M_e} - f_{M_{e-1}}) + (f_{M_e} - f_{M_{e+1}})} \cdot d, \quad (21)$$

где x_0 – низшая граница интервала, в котором находится мода;

f_{M_e} – частота в модальном интервале;

$f_{M_{e-1}}$ – частота в интервале, предшествующем модальному;

$f_{M_{e+1}}$ – частота в последующем интервале;

d – величина интервала.

В вариационном ряду площади земельных участков, предназначенных для строительства в муниципальных районах мода, так же как и медиана располагается в интервале от 1500 до 1650 м² (табл. 13). Численное значение, соответственно, равно:

$$M_0 = 1500 + \frac{(10 - 7)}{(10 - 7) + (10 - 7)} \cdot 150 = 1575 \text{ м}^2.$$

Т а б л и ц а 13

Показатели центра распределения

Группы земельных участков	Число земельных участков в группе f_i	Середина интервала, x_i	$f_i \cdot x_i$	Накопленная частота	Средне-взвешенная \bar{x}	Ме-диана M_e	Мода M_0
1050–1200	4	1125	4500	4			
1200–1350	2	1275	2550	6			
1350–1500	7	1425	9975	13			
1500–1650	10	1575	15750	23			
1650–1800	7	1725	12075	30			
1800–1950	5	1875	9375	35			
1950–2100	5	2025	10125	40			
	40		64350				

2.2.2. Показатели степени вариации

Для вариационного ряда важно изучать степень сплоченности всех отдельных признаков вокруг среднего значения и степень разбросанности этих значений, степень колеблемости. Поэтому полезно рассчитать показатели степени вариации.

Амплитуду колебания отражает вариационных размах – показывает насколько велико различие между единицами совокупности, имеющими наименьшее и наибольшее значение, то есть вариационный размах – это разница между наибольшим и наименьшим значением варьирующего признака (см. табл. 11).

Однако вариационный размах имеет существенный недостаток. В общей совокупности могут встречаться значения признаков, которые или очень велики или очень малы по сравнению с основной массой и эти значения обусловлены случайными обстоятельствами. Эти значения являются аномальными, поэтому размах дает искаженную амплитуду колебания признака. Для исключения аномальных значений необходимо совокупность проверить на засоренность, то есть наличие в совокупности случайных показателей.

Так как величина размаха характеризует лишь максимальное и минимальное значения признака, она не может выявлять закономерности вариации во всей совокупности.

Для анализа необходимы и другие показатели, такие как среднее линейное и среднеквадратическое отклонения (табл. 14).

Эти показатели выступают в роли показателей силы вариации и дают обобщенную характеристику степени колеблемости признаков совокупности, т.е. показывают, на сколько, в среднем, отклоняются конкретные значения признака от его среднего значения.

Для интервального ряда среднее линейное отклонение вычисляется по следующей формуле

$$\bar{d} = \frac{\sum (|x_i - \bar{x}| \cdot f_i)}{\sum f_i}. \quad (22)$$

Среднеквадратическое отклонение для интервального ряда вычисляется по формуле

$$G = \sqrt{\frac{\sum ((x_i - \bar{x})^2 \cdot f_i)}{\sum f_i}}, \quad (23)$$

где x_i – значение признака в вариационном ряду;

\bar{x} – средневзвешенное значение;

f_i – количество единиц совокупности с данным значением признака;

$\sum f_i$ – общее количество единиц совокупности.

Среднеквадратическое отклонение по величине в реальной совокупности всегда больше среднего линейного отклонения. Отношение среднеквадратического отклонения к среднему линейному зависит от наличия в совокупности редких выделяющихся отклонений, и это отношение может служить индикатором засоренности совокупности. Для нормального закона распределения соотношение должно быть равно 1,25 или $\leq 1,25$.

Показатели степени вариации

Группы земельных участков по площади	Число земельных участков в группе f_i	Середина интервала, x_i	Средне-взвешенная \bar{x}	$ x_i - \bar{x} $	$ x_i - \bar{x} \cdot f_i$	\bar{d}	G	$\frac{G}{\bar{d}}$
1050–1200	4	1125	1561	436	1744	211	265	1,25
1200–1350	2	1275		286	572			
1350–1500	7	1425		136	952			
1500–1650	10	1575		14	140			
1650–1800	7	1725		164	1148			
1800–1950	5	1875		314	1570			
1950–2100	5	2025		464	2320			
	40				8446			

Таким образом, значение соотношения G к d является индикатором наличия засоренности. Для рассматриваемого примера $G/d = 1,25$. Если соотношение фиксирует засоренность, необходимо найти признаки, которые засоряют совокупность и исключить их, снова проверить совокупность, добиваясь тем самым нормального ряда распределения.

2.2.3. Относительные показатели вариации

Для целей сравнения колеблемости различных признаков в одной и той же совокупности или при сравнении колеблемости одного признака в нескольких совокупностях представляют интерес показатели вариации, выраженные в относительных единицах (табл. 20). Они определяются как отношение ранее определенных показателей силы вариации к средней величине признака:

1. Коэффициент осцилляции

$$V_R = \frac{R}{\bar{x}} \cdot 100\%; \quad (24)$$

$$V_R = \frac{950}{1561} \cdot 100\% = 60,8\%.$$

2. Коэффициент вариации

$$V_G = \frac{G}{\bar{x}} \cdot 100\%; \quad (25)$$

$$V_G = \frac{265}{1561} \cdot 100\% = 16,9\%.$$

3. Линейные коэффициенты вариации

$$V_{\bar{d}} = \frac{\bar{d}}{x} \cdot 100 \% ; \quad (26)$$

$$V_{\bar{d}} = \frac{211}{1561} \cdot 100 \% = 13,5 \% ;$$

$$V_{M_e} = \frac{G}{M_e} \cdot 100 \% ; \quad (27)$$

$$V_{M_e} = \frac{265}{1605} \cdot 100 \% = 16,5 \% .$$

Т а б л и ц а 15

Относительные показатели вариации

Наименование показателя	Буквенное обозначение	Значение
Вариационный размах	R	950
Средневзвешенная	\bar{x}	1561
Среднеквадратическое отклонение	G	265
Среднее линейное отклонение	\bar{d}	211
Медиана	M_e	1605
Коэффициент осцилляции	V_R	60,8
Коэффициент вариации	V_G	16,9
Линейные коэффициенты вариации	$V_{\bar{d}}$	13,5
	V_{M_e}	16,5

Совокупность считается однородной, если коэффициенты вариации V_G , $V_{\bar{d}}$ и $V_{M_e} \leq 33 \%$. После расчета всех показателей, характеризующих вариационный ряд, необходимо сделать вывод об однородности изучаемой совокупности, то есть об отсутствии в ней аномальных значений.

Совокупность считается не засоренной, если выполняются следующие условия:

- значение медианы больше средневзвешенной и модой;
- отношение среднеквадратического отклонения к среднему линейному отклонению $\leq 1,25$;
- коэффициенты вариации $\leq 33 \%$.

Сравнительная характеристика расчетных и нормированных показателей при нормальном распределении представлена в табл. 16.

Таблица 16

Сравнительная характеристика показателей

Показатели	\bar{x}	M_e	M_0	$\frac{G}{d}$	V_G	V_d	V_{Me}
Нормативные				$\leq 1,25$	≤ 33	≤ 33	≤ 33
Расчетные значения	1561	1605	1575	1,25	16,9	13,5	16,5

В случае не выполнения обозначенных условий, необходимо выявить показатели засоряющие совокупность, исключить их из совокупности и произвести повторные расчеты показателей, характеризующих вариацию.

Показатели, засоряющие совокупность хорошо видно на гистограмме, это те значения совокупности, которые находятся или в минимальной зоне гистограммы или в максимальной, при этом они находятся обособленно от основной массы данных. Так, например, на рис. 7 с изображением интервального вариационного ряда земельных участков, предназначенных для строительства в муниципальных районах видно, что интервал с площадью от 1200 до 1350 имеет два объекта и находится в отрыве от основного ряда распределения. Из этого следует, что земельные участки с площадью 1200 м² являются аномальными и засоряют рассматриваемую совокупность, следовательно, этот признак необходимо исключить из совокупности.

2.3. Устранение засоренности совокупности

Проведенные расчеты на нормальный ряд распределения выявили присутствие засоренности в совокупности аномальными признаками. Установлено, что аномальным признаком являются два земельных участка площадью 1200 м². Исключив данные объекты из совокупности, рассчитаем основные показатели вариационной совокупности, табл. 17.

Таблица 17

Характеристика вариационной совокупности

Число наблюдений n	Минимальное значение совокупности x_{\min}	Максимальное значение совокупности x_{\max}	Вариационный размах R	Количество интервалов K	Шаг интервала d
38	1050	2000	950	6	150

В совокупности остается 38 объектов, вариационный размах составил 950 м², шаг интервала 150 м². Представим интервальный вариационный ряд, табл. 18.

Т а б л и ц а 18

Характеристика интервального вариационного ряда
земельных участков, предназначенных для строительства
в муниципальных районах

Группы земельных участков по площади	Число земельных участков в группе f_i	Середина интервала x_i	Удельный вес
1050–1200	4	1125	10,5
1200–1350	0	1275	0
1350–1500	7	1425	18,4
1500–1650	10	1575	26,3
1650–1800	7	1725	18,4
1800–1950	5	1875	13,2
1950–2100	5	2025	13,2
Σ	40		100

Наибольшее количество объектов находится в интервале от 1350 до 1800 м² (24), а самыми распространенными являются земельные участки с площадью от 1500 до 1650 м² – 26,3 %. Рассчитаем показатели центра распределения, табл.19.

Т а б л и ц а 19

Показатели центра распределения

Группы земельных участков	Число земельных участков в группе f_i	Середина интервала, x_i	$f_i \cdot x_i$	Накопленная частота	Средневзвешенная \bar{x}	Медиана M_e	Мода M_0
1050–1200	4	1125	4500	4			
1200–1350	0	1275	0	0			
1350–1500	7	1425	9975	11			
1500–1650	10	1575	15750	21			
1650–1800	7	1725	12075	28			
1800–1950	5	1875	9375	33			
1950–2100	5	2025	10125	38			
	38		61800		1626	1620	1575

Расчеты показателей центра распределения показали, что средневзвешенное значение \bar{x} после устранения засоренности увеличилось и стало равно 1626 м², увеличилось значения $M_e = 1620$ м², мода осталась прежней $M_0 = 1575$ м². Рассчитаем показатели степени вариации, табл. 20.

Таблица 20

Показатели степени вариации

Группы земельных участков по площади	Число земельных участков в группе, f_i	Середина интервала, x_i	Средневзвешенная \bar{x}	$ x_i - \bar{x} $	$ x_i - \bar{x} \cdot f_i$	\bar{d}	G	$\frac{G}{\bar{d}}$
1050–1200	4	1125	1626	501	2004	207	256	1,24
1200–1350	0	1275		351	0			
1350–1500	7	1425		201	1407			
1500–1650	10	1575		51	510			
1650–1800	7	1725		99	693			
1800–1950	5	1875		249	1245			
1950–2100	5	2025		399	1995			
	38				7854			

После устранения засоренности совокупности нормальное распределение приобрело и соотношение $G/d = 1,24$.

Рассчитаем показатели степени вариации, табл. 21.

Таблица 21

Относительные показатели вариации

Наименование показателя	Буквенное обозначение	Значение
Вариационный размах	R	950
Средневзвешенная	\bar{x}	1626
Среднеквадратическое отклонение	G	256
Среднее линейное отклонение	\bar{d}	207
Медиана	M_e	1620
Коэффициент осцилляции	V_R	58,4
Коэффициент вариации	V_G	15,7
Линейные коэффициенты вариации	$V_{\bar{d}}$	12,7
	V_{M_e}	15,8

Сравнительная характеристика расчетных показателей в сопоставлении с нормированными значениями представлена в табл. 22.

Таблица 22

Сравнительная характеристика показателей

Показатели	\bar{x}	M_e	M_0	$\frac{G}{\bar{d}}$	V_G	$V_{\bar{d}}$	V_{M_e}
Нормативные				$\leq 1,25$	≤ 33	≤ 33	≤ 33
Расчетные значения	1626	1620	1575	1,24	15,7	12,7	15,8

Исключение объектов засорявших выборочную совокупность привело ранжированный ряд к нормальному распределению, гистограмма представлена на рис.8. Среднее значение площади земельного участка, предназначенного под строительство составило 1626 м^2 , наиболее распространенными являются земельные участки под строительство площадью 1500 м^2 , соотношение G/d не превышает 1,25, коэффициент вариации признака не превышает 33 %.

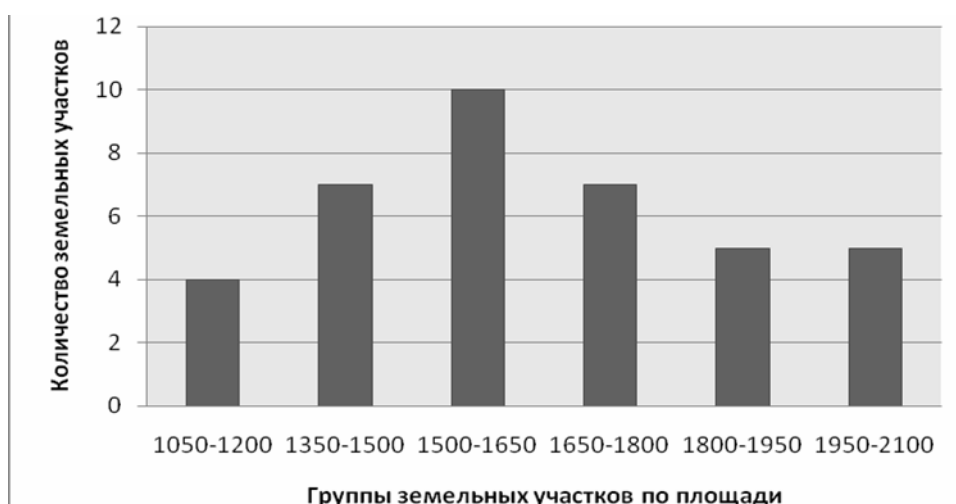


Рис. 8. Интервальный вариационный ряд земельных участков, предназначенных для строительства в муниципальных районах при нормальном распределении

2.4. Анализ зависимости размера земельных участков, предоставляемых для строительства, от численности населения в населенном пункте

Для анализа зависимости среднего размера земельного участка для строительства в муниципальном районе от численности населения, необходимо произвести группировку населенных пунктов муниципального района по численности населения, проживающих в них (табл. 24).

Используя исходные данные по численности (табл. 8), рассчитываем характеристики (табл. 23), необходимые для построения группировки – количество интервалов (формула (15)) и шаг интервала (формула (16)).

Т а б л и ц а 23

Характеристика вариационной совокупности

Число наблюдений n	Минимальное значение совокупности x_{\min}	Максимальное значение совокупности x_{\max}	Вариационный размах R	Количество интервалов, K	Шаг интервала d
4	16800	58600	41800	3	13950

Перед построением группировки шаг интервала необходимо округлить до ближайшего числа кратного 50 или 100. Построение группировки по численности населения (табл. 24) начинается с 0. По получившимся группам населенные пункты распределяются согласно их людности, табл. 24.

Т а б л и ц а 24

Группировка муниципальных районов по численности населения

Группы муниципальных образований по людности, чел.	Наименование муниципального района	Численность населения, чел.	Количество районов
0–13950	-	-	0
13950–27900	Сосновоборский	16800	1
27900–41850	Кузнецкий	37691	2
	Мокшанский	27900	
41850–55800	-	-	0
55800–69750	Каменский	58600	1

Анализируя таблицу 24 можно увидеть, что наибольшее количество муниципальных районов имеет численность населения от 27900 до 41850 человек, в одном муниципальном районе количество жителей более 13950 человек, в другом – более 55800 человек.

По результатам группировки населенных пунктов согласно их численности населения, необходимо каждую группу представить в виде интервального ряда и произвести расчет всех структурных характеристик (по аналогии с табл. 8–16) по соответствующим формулам (12) – (26), определить наличие или отсутствие аномальных показателей в каждой совокупности.

Результатом анализа зависимости среднего размера земельного участка, предназначенного для строительства от численности населения является сопоставление средневзвешенного размера земельного участка и численности населения в муниципальных районах в соответствующей группе (табл. 25).

Т а б л и ц а 25

Зависимость размера земельного участка, предназначенного для строительства от численности населения

Группы муниципальных районов по людности	Количество районов в группе	Средний размер земельного участка в группе
13950–27900	1	1860
27900–41850	2	1445
55800–69750	1	1489

В результате проведенного анализа можно сделать вывод, чем больше населения проживает в населенном пункте, тем меньше размер земельного участка, предоставляемого для строительства.

Контрольные вопросы

1. Что такое вариация?
2. Что характеризует коэффициент вариации?
3. Что такое дисперсия?
4. Что такое коэффициент вариации?
5. Что такое мода?
6. Как рассчитывается мода?
7. Что такое медиана?
8. Как рассчитывается медиана?
9. Как рассчитывается размах вариации?
10. Как рассчитывается среднее линейное отклонение?
11. Этапы обработки статистической информации.
12. Задачи, решаемые при помощи статистических группировок.
13. Виды статистических группировок.
14. Принципы построения статистических группировок.
15. По какой формуле определяют оптимальное число групп?
16. Как рассчитывается величина равного интервала?
17. Дать характеристику интервалам группировки.
18. Что такое группировочный признак?
19. Закрытые, открытые интервалы.
20. Как рассчитывается величина неравного интервала?

3. ОТБОР ФАКТОРОВ, ВЛИЯЮЩИХ НА СТОИМОСТЬ ЗЕМЕЛЬ

Задание: Исходными данными для выполнения работы по отбору факторов, влияющих на кадастровую стоимость земель, является характеристика населенных пунктов муниципального района Пензенской области по основным социально-демографическим и экономическим показателям в баллах и кадастровой стоимости (таблица 26).

Основной целью работы является оценка степени влияния социально-экономических факторов на кадастровую стоимость земель с помощью статистических методов.

3.1. Корреляционный анализ

В ходе статистического исследования объективно существующих связей между явлениями необходимо выявить причинно-следственные зависимости между показателями, то есть определить, насколько изменение одних показателей зависит от изменения других показателей.

При изучении массовых общественно-экономических явлений между признаками проявляется корреляционная связь.

Корреляционной связью называют связь, при которой разным значениям одной переменной соответствуют различные средние значения другой переменной. С изменением значения признака-фактора X_i закономерным образом изменяются средние значения результативного признака Y . В то время как в каждом отдельном случае значение результативного признака может принимать множество различных значений.

Таким образом, корреляционная связь – это связь, где воздействие отдельных факторов на результативный показатель проявляется только как тенденция (в среднем) при массовом наблюдении фактических данных.

Так, например, корреляционная связь может проявляться между показателем кадастровой стоимости земель в населенных пунктах и значениями различных социально-экономических показателей по данным населенным пунктам (табл. 26). В этом случае кадастровая стоимость земель будет выступать в качестве результативного признака Y , а социально экономические показатели – факторных признаков X_i .

Т а б л и ц а 26

Значения социально-экономических показателей
по населенным пунктам муниципального района

Наименование населенных пунктов	Кадастровая стоимость	Баллы по основным социально-демографическим и экономическим показателям						
		1	2	3	4	5	6	7
Александровка	4,58	76,13	16,00	40,00	63,00	10,00	25,00	99,00
Анновка	4,50	64,58	0,01	40,00	72,00	7,25	45,00	99,00
Бакшеевка	4,47	64,61	0,01	0,01	55,00	10,00	45,00	98,00
Бардинка	5,67	76,16	16,50	61,50	54,00	60,00	25,00	99,00
Блохино	4,17	43,29	0,01	0,01	50,00	3,00	100,0	97,00
Вазерки	5,91	69,90	27,00	67,50	74,00	48,00	25,00	98,00
Васильевка	5,04	69,03	10,00	0,01	60,0	34,50	25,00	98,00
Грабово	3,83	37,97	0,01	0,01	7,00	3,00	100,0	80,00
Кижеватово	4,61	69,06	0,01	0,01	60,00	9,25	45,00	99,00
Колос	6,33	65,20	30,00	77,50	60,00	61,88	50,00	99,00
Кропцово	4,74	65,31	0,01	20,00	70,00	6,25	70,00	100,0
Лопатки	4,47	58,00	16,00	40,00	60,00	7,50	70,00	100,0
Лопуховка	5,09	50,91	14,00	67,50	70,00	25,50	50,00	99,00
Мастиновка	4,49	61,66	0,01	15,00	50,00	5,50	70,00	100,0
Николаевка	4,27	54,46	0,01	0,01	50,00	5,50	70,00	99,00
Никольское	4,27	54,46	0,01	20,00	50,00	5,50	70,00	98,00
Новая Жизнь	6,26	75,15	30,00	73,50	76,00	57,50	25,00	98,00
Пазелки	6,09	74,32	0,01	40,00	69,00	60,00	45,00	98,00
Победа	4,52	69,83	0,01	20,00	44,00	8,25	45,00	97,00
Подлесный	4,86	73,41	0,01	20,00	77,00	10,00	45,00	99,00
Полевой	4,86	57,41	13,00	0,01	75,00	37,50	25,00	100,0
Проказна	4,66	69,83	27,00	20,00	61,00	10,00	45,00	100,0
Пыркино	6,09	75,29	15,00	20,00	63,00	60,00	45,00	100,0
Светлополянское лесничество	4,76	74,46	22,00	40,00	62,00	10,00	25,00	100,0
Сергеевка	2,45	0,01	9,00	0,01	66,00	4,00	45,00	99,00
Сосновка	6,16	67,92	36,10	73,50	65,00	57,75	25,00	98,00
Степановка	4,26	53,73	0,01	0,01	65,00	2,50	70,00	100,0
Трофимовка	4,29	57,27	0,01	8,02	50,00	3,00	80,00	100,0
Ухтинка	4,49	53,73	25,00	9,26	50,00	3,00	100,0	90,00
Чемодановка	4,47	58,00	16,00	40,00	60,00	7,50	70,00	97,00

В качестве факторных признаков в данном случае рассматриваются следующие социально-экономические показатели:

– доступность населения к центру населенного пункта, объектам культуры и быта X_1 ;

– обеспеченность централизованным инженерным оборудованием и благоустройство территории и застройки X_2 ;

– уровень развития сферы социального культурно-бытового обслуживания населения микро районного значения X_3 ;

- историческая ценность застройки, эстетическая и ландшафтная ценность территории X_4 ;
- состояние окружающей среды, санитарные и микроклиматические условия X_5 ;
- инженерно-геологические условия строительства и степень подверженности территории разрушительным воздействиям природы X_6 ;
- рекреационная ценность территории X_7 .

При формировании массива данных для установления функциональных связей и регрессионных зависимостей, необходимо представленную совокупность проверить на подчиненность закону нормального распределения, т.е. проверить совокупность на наличие случайных наблюдений – засоренность. Теоретические положения по данному вопросу представлены в разд. 2.

3.1.1. Определение коэффициентов корреляции и анализ тесноты связи

Задача корреляционного метода – определение тесноты связи между признаками. Для установления тесноты связи рассчитывают линейный коэффициент корреляции, который дает возможность определить полезность факторных признаков:

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{G_x \cdot G_y}, \quad (28)$$

$$G_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}, \quad (29)$$

$$G_y = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}}, \quad (30)$$

- где x_i – значение факторного признака;
 y_i – значение результативного признака;
 r – линейный коэффициент корреляции;
 \overline{xy} – среднее произведение признаков;
 \bar{x} – среднее значение факторного признака;
 \bar{y} – среднее значение результативного признака;
 G_x – стандартное отклонение X ;
 G_y – стандартное отклонение Y ;
 n – число наблюдений.

Коэффициент корреляции может принимать значения в пределах от минус единицы до плюс единицы ($-1 \leq r \leq 1$). Положительный знак коэффициента корреляции свидетельствует о прямой зависимости между признаками, т.е. с увеличением факторного признака происходит увеличение результативного признака. Отрицательный знак свидетельствует об обратной зависимости: с увеличением факторного признака происходит уменьшение результативного признака.

Для оценки тесноты связи между признаками используют шкалу значимости коэффициентов корреляции (табл. 27).

Т а б л и ц а 2 7

Шкала значимости линейных коэффициентов корреляции

Значение коэффициента корреляции	Теснота связи
от 0 до $\pm 0,15$	Связь отсутствует
от $\pm 0,16$ до $\pm 0,20$	Связь плохая
от $\pm 0,21$ до $\pm 0,30$	Связь слабая
от $\pm 0,31$ до $\pm 0,40$	Связь умеренная
от $\pm 0,41$ до $\pm 0,60$	Связь средняя
от $\pm 0,61$ до $\pm 0,80$	Связь высокая
от $\pm 0,81$ до $\pm 0,90$	Связь очень высокая
от $\pm 0,91$ до $\pm 1,00$	Связь полная

В современных условиях расчет парных коэффициентов корреляции автоматизирован и может осуществляться при помощи специализированного программного обеспечения. Одним из таких программных продуктов является программный комплекс «STATISTICA».

На основе исходной информации (см. табл. 26) рассчитывается матрица парных коэффициентов корреляции (табл. 28) и поясняется мера тесноты связи между каждым факторным признаком и результативным признаком, исходя из которого, можно сделать вывод о пригодности конкретного факторного признака для дальнейших расчетов и анализа.

Т а б л и ц а 2 8

Матрица парных коэффициентов корреляции

Признаки	Y	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇
Y	1,00	0,74	0,54	0,71	0,38	0,88	-0,49	0,20
X ₁		1,00	0,24	0,41	0,30	0,44	-0,48	0,30
X ₂			1,00	0,88	0,30	0,56	-0,43	0,06
X ₃				1,00	0,38	0,64	-0,46	0,19
X ₄					1,00	0,36	-0,61	0,74
X ₅						1,00	-0,59	0,17
X ₆							1,00	-0,49
X ₇								1,00

Так, рассматривая матрицу парных коэффициентов корреляции (табл. 28) и исходя из шкалы значимости линейных коэффициентов корреляции (табл. 27), можно увидеть, что связь между доступностью населения к центру населенного пункта, объектам культуры и быта (X_1) и кадастровой стоимостью высокая, так как линейный коэффициент корреляции между ними равен 0,74. Так же высокая связь наблюдается между кадастровой стоимостью и уровнем развития сферы социального культурно-бытового обслуживания населения микрорайонного значения (X_3), коэффициент корреляции между которыми равен 0,71. Средняя связь выявлена между кадастровой стоимостью и обеспеченностью централизованным инженерным оборудованием и благоустройством территории и застройки (X_2), коэффициент корреляции равен 0,54, а так же кадастровой стоимостью и инженерно-геологическими условиями строительства и степенью подверженности территории разрушительным воздействиям природы (X_6), коэффициент корреляции равен $-0,49$. Между исторической ценностью застройки, эстетической и ландшафтной ценностью территории (X_4) и кадастровой стоимостью связь умеренная. Между состоянием окружающей среды, санитарными и микроклиматическими условиями (X_5) и кадастровой стоимостью связь очень высокая, а между кадастровой стоимостью и рекреационной ценностью территории (X_7) – плохая, так как коэффициент корреляции между ними равен 0,20

В дальнейшие расчеты не рекомендуется включать факторные признаки, у которых связь с результативным признаком слабая, плохая или отсутствует, то есть коэффициент корреляции с результативным признаком, у которых по абсолютной величине меньше 0,31.

Кроме того, не рекомендуется включать в дальнейшие расчеты факторы, тесно связанные друг с другом. Сложность и взаимное переплетение отдельных факторов может проявляться в так называемой мультиколлинеарности – это теснота зависимости между двумя факторными признаками.

Индикатором, определяющим наличие мультиколлинеарности, является парный коэффициент корреляции между двумя факторными признаками, превышающий величину 0,8 по абсолютной величине. Для устранения мультиколлинеарности необходимо факторный признак преобразовать в новый или один из признаков исключить из дальнейших расчетов.

Выбор, какой из факторных признаков необходимо исключить, решается на основе логического анализа. Для этого сравниваются линейные коэффициенты корреляции данных факторных признаков и результативного признака. Исключается тот факторный признак, у которого линейный коэффициент корреляции с результативным признаком меньше.

Например, анализируя табл. 28, можно сделать вывод о наличии мультиколлинеарности между обеспеченностью централизованным инженерным оборудованием и благоустройством территории и застройки (X_2) и

уровнем развития сферы социального культурно-бытового обслуживания населения микрорайонного значения (X_3) – $r_{x_2x_3} = 0,88 > 0,80$. Сравнивая тесноту связи данных факторных признаков с результативным, исключаем из дальнейших расчетов обеспеченность централизованным инженерным оборудованием и благоустройство территории и застройки, так как данный признак слабее связан с кадастровой стоимостью. То есть $r_{yx_2} = 0,54 < r_{yx_3} = 0,71$.

Таким образом, по результатам анализа тесноты связи между факторными признаками и результативным, а так же между собой, из дальнейших расчетов можно исключить обеспеченность централизованным инженерным оборудованием и благоустройство территории и застройки (X_2) (в связи с мультиколлинеарностью) и рекреационную ценность территории (X_7) (в связи с плохой связью с результативным признаком).

3.1.2. Анализ существенности линейных коэффициентов корреляции

Точность или надежность результатов изучения корреляционной связи зависит от количества сопоставимых данных, число которых очень часто бывает ограничено. Поэтому полезно рассчитать погрешность вычисленного коэффициента корреляции, его существенность (табл. 29) или надежность установленной связи.

Существенность линейного коэффициента корреляции рассчитывают на основе t -критерия Стьюдента.

Для этого определяется расчетное значение t -критерия Стьюдента:

$$t_p = \frac{r}{m_r}, \quad (31)$$

где m_r – средняя ошибка определения линейного коэффициента корреляции.

Средняя ошибка определения линейного коэффициента корреляции, рассчитывается по следующим формулам:

$$m_r = \frac{1 - r^2}{n}, \quad (32)$$

$$m_r = \sqrt{\frac{1 - r^2}{n - 2}}, \quad (33)$$

где n – количество наблюдений;

r – линейный коэффициент корреляции.

Формула (27) используется при объеме выборки более 30 наблюдений, формула (28) – при объеме выборки 30 наблюдений и менее.

Так, для признака доступность населения к центру населенного пункта, объектам культуры и быта (X_1) коэффициент корреляции равен 0,74. Исходя

из этого, можно определить значение средней ошибки и расчетного значения t -критерия Стьюдента:

$$m_r = \sqrt{\frac{1 - 0,74^2}{30 - 2}} = 0,1266;$$

$$t_p = \frac{0,74}{0,1266} = 5,8648.$$

После определения расчетного значения t -критерия Стьюдента для каждого линейного коэффициента корреляции необходимо сопоставить их с табличным значением t -критерия Стьюдента (прил. А). Табличное значение критерия определяется исходя из количества степеней свободы ($n-1$) и принимаемого уровня значимости (95 % или 0,05).

Если расчетное значение t -критерия Стьюдента больше табличного, то существенность линейного коэффициента корреляции доказана, связь между признаками установлена надежно.

Т а б л и ц а 29

Анализ существенности линейных коэффициентов корреляции

Признак	Коэффициент корреляции, r	Средне-квадратическая ошибка, m_r	Расчетное значение критерия Стьюдента, t_p	Табличное значение критерия Стьюдента, t_t	Существенность коэффициента корреляции
X_1	0,74	0,1266	5,8648	2,0452	Существенен
X_3	0,71	0,1335	5,2989	2,0452	Существенен
X_4	0,38	0,1747	2,1830	2,0452	Существенен
X_5	0,88	0,0909	9,6465	2,0452	Существенен
X_6	-0,49	0,1651	-2,9508	2,0452	Существенен

Анализируя таблицу 29, можно сделать вывод, что все коэффициенты корреляции существенны, так как расчетные значения t -критерия Стьюдента для каждого из них больше табличного значения t -критерия Стьюдента для 30 наблюдений.

П р и м е ч а н и е . При сравнении расчетного значения t -критерия Стьюдента с табличным значением знак коэффициента не учитывается.

3.2. Регрессия

Корреляция и регрессия тесно связаны между собой – корреляция оценивает силу связи, а регрессия исследует ее форму. Под формой корреляционной связи понимают тип аналитической формулы или уравнение, выражающее зависимость между изучаемыми признаками.

Регрессия может быть однофакторной (парной) и многофакторной (множественной).

3.2.1. Парная регрессия

Парная (однофакторная) регрессия – это функциональная зависимость результативного признака от одного факторного признака, то есть уравнение, связывающее результативный и факторный показатели (таблица 30).

Степень зависимости результативного признака от конкретного факторного признака можно проанализировать, изучив коэффициент детерминации R^2 . Он характеризует, какая доля вариации результативного признака обусловлена изменением факторного признака, включенного в уравнение парной регрессии.

Процесс подбора уравнения парной регрессии, так же как и значений коэффициентов корреляции (R) и коэффициентов детерминации (R^2), в современных условиях автоматизирован и может осуществляться при помощи специализированного программного обеспечения. Одним из таких программных продуктов является программный комплекс «STATISTICA».

Подбор уравнений осуществляется на основе исходной информации (см. табл. 26).

Т а б л и ц а 3 0

Уравнения парной регрессии

Признак	Уравнение парной регрессии	R	R^2
X_1	$Y=2,31+0,04X_1$	0,74	0,56
X_3	$Y=4,21+0,02X_3$	0,71	0,50
X_4	$Y=3,40+0,02X_4$	0,38	0,15
X_5	$Y=4,12+0,03X_5$	0,88	0,77
X_6	$Y=5,73-0,02X_6$	0,49	0,24

Так, коэффициент детерминации равный 0,56 означает, что на 56 % кадастровая стоимость обусловлена изменением факторного признака доступность населения к центру населенного пункта, объектам культуры и быта (X_1). На 50 % кадастровая стоимость зависит от уровня развития сферы социального культурно-бытового обслуживания населения микрорайонного значения (X_3); на 15 % от исторической ценности застройки, эстетической и ландшафтной ценности территории (X_4); на 77 % от состояния окружающей среды, санитарных и микроклиматических условий (X_5); на 24 % от инженерно-геологических условий строительства и степени подверженности территории разрушительным воздействиям природы (X_6).

3.2.2. Множественная регрессия

В реальном мире на результативный признак Y влияет достаточно большое число факторных признаков X . В этом случае необходимо установить функциональную зависимость между ними. Такая связь носит название множественной или многофакторной регрессии.

Множественная (многофакторная) регрессия – это функциональная зависимость результативного признака от двух и более факторных признаков.

Процесс построения модели множественной регрессии включает в себя 3 этапа. На первом этапе необходимо выбрать форму связи. Выбор заключается в том, что из большого числа вариантов необходимо отобрать ту функцию, которая наиболее оптимально будет описывать связи исследуемого социально-экономического явления.

Практика показывает, что для описания социально-экономических явлений используют 5 основных моделей:

1. Линейное уравнение

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n, \quad (34)$$

где n – число факторных признаков;
 $a_0, a_1, a_2 \dots a_n$ – параметры моделей (коэффициенты регрессии);
 $x_1, x_2, \dots x_n$ – факторные признаки;

2. Степенное уравнение

$$y = a_0x_1^{a_1} + a_2x_2^{a_2} + \dots + a_nx_n^{a_n}. \quad (35)$$

3. Параболическое уравнение

$$y = a_0 + a_1x_1^2 + a_2x_2^2 + \dots + a_nx_n^2. \quad (36)$$

4. Гиперболическое уравнение

$$y = a_0 + \frac{a_1}{x_1} + \frac{a_2}{x_2} + \dots + \frac{a_n}{x_n}. \quad (37)$$

5. Экспоненциальное уравнение

$$y = e^{a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n}. \quad (38)$$

Из них чаще всего используется линейная модель.

Важным этапом построения уже выбранного уравнения множественной регрессии является отбор факторных признаков (т.е. определение размерности модели).

Проблема отбора факторных признаков для построения моделей связи может быть решена на основе эвристических методов (метод экспертных оценок) или с использованием многомерных статистических методов анализа. Наиболее распространенный метод – метод пошаговой регрессии.

Сущность метода пошаговой регрессии заключается в последовательном включении факторных признаков в уравнение регрессии и последующей проверки их значимости.

Для решения этой задачи возможно использование специализированного программного комплекса «STATISTICA». Основой расчетов являются значения социально-экономических показателей и средние значения кадастровой стоимости земель по населенным пунктам муниципального района (табл. 26). Уравнение регрессии фиксируется для каждого этапа, одновременно для каждого уравнения определяется значение множественного коэффициента корреляции (R) (табл. 31).

Множественный коэффициент корреляции рассчитывается при наличии линейной связи между результативным признаком и несколькими парными. Множественный коэффициент корреляции всегда положителен. Приближение множественного коэффициента корреляции (R) к единице свидетельствует о сильной зависимости между признаками.

Т а б л и ц а 31

Отбор признаков методом пошаговой регрессией

Признак	R	Уравнение	R^2	F_p
X_1	0,74	$Y=2,31+0,04X_1$	–	–
X_3	0,86	$Y = 2,56+0,03X_1+0,02X_3$	–	–
X_4	0,87	$Y = 2,44+0,03X_1+0,02X_3+0,01X_4$	–	–
X_5	0,97	$Y = 2,85+0,03X_1+0,02X_3+0,01X_4+0,02X_5$	–	–
X_6	0,99	$Y = 1,60+0,03X_1+0,01X_3+0,01X_4+0,03X_5 - 0,01X_6$	0,98	206,17

При проверке значимости введенного фактора производят анализ изменения величины множественного коэффициента корреляции R и коэффициентов регрессии.

Фактор является не значимым, если его включение в уравнение регрессии изменяет значение коэффициентов регрессии, не увеличивая или уменьшая величины множественного коэффициента корреляции, и наоборот.

Так, анализируя таблицу 31, можно сделать вывод, что каждый факторный признак существенен и должен быть включен в множественное уравнение регрессии, так как значение множественного коэффициента корреляции с каждым шагом увеличивается, а значение коэффициентов регрессии при соответствующих факторных признаках не меняется или меняется не существенно.

Таким образом, в результате проведения корреляционно-регрессионного анализа исходных данных, получено линейное уравнение множественной регрессии, отображающее взаимосвязь между результативным показателем и факторами признаками, оказывающими на него наибольшее влияние:

$$Y = 1,60+0,03X_1+0,01X_3+0,01X_4+0,03X_5 - 0,01X_6, \quad (39)$$

3.2.3. Оценка степени соответствия модели явлению и возможности ее применения

Качество уравнения регрессии зависит от степени достоверности и надежности исходных данных и их количества. Поэтому необходимо провести анализ адекватности уравнения регрессии по F -критерию Фишера-Снедекора или оценку степени соответствия модели изучаемому явлению (процессу).

Для этого необходимо сравнить табличное значение F -критерия Фишера-Снедекора (прил. Б) с расчетным значением. В случае если расчетное значение F -критерия Фишера больше табличного, уравнение считается адекватным изучаемому явлению (процессу), а количество исходных данных достаточным, и наоборот.

Расчетное значение F -критерия Фишера определяется по формуле

$$F_R = \frac{R^2}{1 - R^2} \cdot \frac{n - m}{m - 1}, \quad (40)$$

где R – множественный коэффициент корреляции;
 R^2 – коэффициент детерминации;
 n – количество наблюдений;
 m – количество факторных признаков.

При работе с программным комплексом «STATISTICA» расчетное значение F -критерия Фишера определяется автоматически при подборе уравнения регрессии (табл. 31). Для полученного уравнения регрессии по кадастровой стоимости (формула (34)) расчетное значение F -критерия Фишера $F_R = 206,17$, табличное значение $F_T = 2,64$ (прил. Б). Таким образом, расчетное значение больше табличного, следовательно, уравнение адекватно изучаемому явлению.

При анализе адекватности уравнения регрессии изучаемому явлению (процессу) по F -критерию Фишера возможны следующие варианты его использования:

1. Построенное уравнение на основе его проверки по F -критерию Фишера адекватно, все коэффициенты регрессии значимы – такое уравнение может быть использовано для принятия решений и осуществления прогнозов.

2. Уравнение по F -критерию Фишера адекватно, но часть коэффициентов регрессии не значимы. В этом случае уравнение пригодно для принятия необходимых решений. Для осуществления прогноза оно не пригодно.

Уравнение по F -критерию Фишера неадекватно, коэффициенты регрессии не значимы. В этом случае уравнение считается непригодным. Уравнение регрессии, которое включает в себя основные факторы, влияющие на вариацию результативного признака, и обладает высоким коэффициентом детерминации (не менее 0,5) и значимыми коэффициентами регрессии яв-

ляется Корреляционно-регрессионной моделью (КРМ) системы взаимосвязанных признаков.

Коэффициент детерминации – коэффициент множественной корреляции в квадрате. Он характеризует, какая доля вариации результативного признака обусловлена изменением факторных признаков, входящих в многофакторную регрессионную модель. Коэффициент детерминации больше 0,5 означает, что результативный признак Y должен зависеть более чем на 50 % от совокупности факторных признаков, включенных в уравнение.

Из этого следует, что не любое уравнение регрессии можно считать моделью. Так, полученное для кадастровой стоимости уравнение регрессии (формула (34)), является корреляционно-регрессионной моделью, так как включает в себя основные социально-экономические показатели, оказывающие значительное влияние на кадастровую стоимость, обладает высоким коэффициентом детерминации ($R^2=0,98$) и значащими коэффициентами регрессии. Такая модель может быть использована для принятия управленческих решений и осуществления прогноза.

3.3. Экономический анализ уравнения множественной регрессии

Наиболее сложный этап, который завершает регрессионный анализ – интерпретация коэффициентов уравнения регрессии, т.е. перевод с языка статистики и математики на язык экономики.

Для экономической интерпретации в первую очередь рассматриваются коэффициенты регрессии $a_0, a_1, a_2, \dots, a_n$ при соответствующих факторных признаках (уравнение множественной регрессии – формула (34)). Чем больше величина коэффициента регрессии, тем значительнее влияние данного фактора на результативный показатель.

Особое значение имеет знак перед коэффициентом регрессии. Он говорит о характере влияния на результативный признак. Если коэффициент имеет знак «+», то с увеличением данного фактора результативный признак увеличивается, и наоборот. Если коэффициент имеет знак «-», то с его увеличением факторного признака результативный признак будет уменьшаться, и наоборот.

Так, например: $a_1=0,03$ означает, что при увеличении (уменьшении) факторного признака доступность населения к центру населенного пункта, объектам культуры и быта (X_1) на 1 балл кадастровая стоимость земель в населенном пункте увеличится (уменьшится) на 0,03 рубля. $a_6 = -0,01$ означает, что при увеличении (уменьшении) факторного признака инженерно-геологические условия строительства и степень подверженности территории разрушительным воздействиям природы (X_6) на 1 балл кадастровая стоимость уменьшится (увеличится) на 0,01 рубля.

Кроме коэффициентов регрессии, полезно рассчитать и проанализировать частные коэффициенты эластичности:

$$\mathcal{E}_{X_i} = a_i \cdot \frac{\bar{x}_i}{y}, \quad (41)$$

где a_i – коэффициент регрессии при соответствующем факторном признаке;

\bar{x}_i – среднее значение соответствующего факторного признака;

y – среднее значение результативного признака.

Коэффициент эластичности показывает, на сколько процентов, в среднем, изменится значение результативного признака при изменении факторного признака на 1 %.

Так же, как и коэффициент регрессии, частный коэффициент эластичности может иметь знак «+» и «-». Положительный коэффициент эластичности свидетельствует о прямой зависимости между признаками, отрицательный – об обратной.

Средние значения факторных и результативного признаков определяются на основе исходных данных (табл. 26). Коэффициент регрессии для соответствующего факторного признака берется из уравнения регрессии (формула (34)).

Так для факторного признака доступность населения к центру населенного пункта, объектам культуры и быта (X_1) расчет коэффициента эластичности будет следующим:

$$\mathcal{E}_{x_1} = 0,03 \cdot \frac{61,37}{4,82} = 0,38.$$

Это значит, что при увеличении (уменьшении) факторного признака доступность населения к центру населенного пункта, объектам культуры и быта (X_1) на 1 % кадастровая стоимость увеличится (уменьшится) на 38%.

Коэффициент эластичности для факторного признака инженерно-геологические условия строительства и степень подверженности территории разрушительным воздействиям природы (X_6) будет следующим:

$$\mathcal{E}_{x_6} = 0,01 \cdot \frac{52,5}{4,82} = 0,11.$$

Это значит, что при увеличении (уменьшении) факторного признака инженерно-геологические условия строительства и степень подверженности территории разрушительным воздействиям природы (X_6) на 1 % кадастровая стоимость уменьшится (увеличится) на 11 %.

Контрольные вопросы

1. В каких случаях возникает функциональная связь?
2. В каких случаях возникает статистическая связь?
3. Что понимается под корреляционной связью?
4. Что показывает парный коэффициент корреляции?
5. Что такое мультиколлинеарность?
6. Как определяется оценка значимости параметров модели регрессии?
7. Как определяется оценка значимости уравнения регрессии?
8. В каких случаях возникает парная регрессия?
9. Что такое результативный признак?
10. Что такое факторный признак это?
11. Понятие прямой связи.
12. Понятие обратной связи.
13. Понятие парной регрессии.
14. Понятие множественной регрессии.
15. В чем суть метода пошаговой регрессии?
16. Что показывает коэффициент регрессии?
17. Что показывает коэффициент детерминации?
18. Что показывает коэффициент эластичности?
19. Понятие корреляционно-регрессионной модели.
20. Возможности применения корреляционно-регрессионных моделей.

ЗАДАНИЯ ДЛЯ ВЫПОЛНЕНИЯ РАСЧЕТНО-ГРАФИЧЕСКОЙ РАБОТЫ

Расчетно-графическая работа (РГР) выполняются при ознакомлении с теоретической частью учебно-методического пособия согласно варианту, который выдает преподаватель и включает выполнение следующих разделов.

1. Анализ и прогноз земельных ресурсов Пензенской области.

Примерная структура РГР:

- Содержание.
- Введение.
- Характеристика использования земельных ресурсов муниципального района.
- Анализ перераспределения земельных ресурсов муниципального района по категориям земельного фонда.
- Прогноз распределения земельных ресурсов муниципального района методом экстраполяции.
- Заключение.
- Библиографический список.

Целью работы является ознакомление студентов с простейшими приемами обработки земельно-кадастровой информации.

В качестве индивидуального задания студенту выдается характеристика земельных ресурсов муниципального района в разрезе категорий земель.

Используя статистические методы систематизации и обработки земельно-кадастровой информации, студент должен выполнить следующие задания:

Задание 1. Провести анализ площади земель муниципального района по категориям на последний год наблюдения; представить структуру земель муниципального района в табличной форме и графически; сделать соответствующие выводы.

Задание 2. Провести анализ распределения земель муниципального района по категориям (в гектарах) за рассматриваемый период; представить графики изменения категорий земель муниципального района за рассматриваемый период; провести анализ изменения структуры земель муниципального района в последний год по отношению к начальному периоду; представить изменения графически, используя столбчатую и ленточную диаграммы; определить основные характеристики для каждой категории земель (средняя хронологическая, темп роста, абсолютный прирост, темп прироста); сделать соответствующие выводы.

Задание 3. Разработать прогноз изменения структуры земель муниципального района по категориям методом экстраполяции; отразить подбор аналитической функции для прогноза земель графически; определить оценку ошибки метода, интервальную оценку и доверительные интервалы; вы-

полнить соответствующий анализ прогнозируемого изменения структуры земельных ресурсов района.

Работа выполняется в программе Microsoft Excel. После выполнения задания работа оформляется с помощью программы Microsoft Word в соответствии с правилами оформления работы.

2. Анализ размеров земельных участков, предназначенных для строительства жилья в муниципальном районе.

Примерная структура РГР:

- Введение.
- Характеристика и анализ размеров земельных участков, предназначенных для строительства жилья в муниципальном районе.
- Группировка населенных пунктов муниципального района по численности населения.
- Характеристика и анализ размеров земельных участков, предназначенных для строительства по группам с учетом численности населения.
- Определение зависимости размеров земельных участков, предназначенных для строительства жилья от численности населения.
- Заключение.
- Библиографический список.

Целью работы является ознакомление студентов с приемами изучения массовых явлений.

В качестве индивидуального задания студенту выдаются данные о размерах земельных участков, предназначенных для строительства жилья по населенным пунктам муниципального района Пензенской области.

Используя простейшие методы статистического анализа и освоенные приемы изучения массовых явлений, студент должен выполнить следующие задания:

Задание 1. Провести ранжирование исследуемой совокупности данных по площади земельных участков, предназначенных для строительства жилья; представить совокупность в виде дискретного вариационного ряда и интервального вариационного ряда; представить графическое изображение интервального вариационного ряда.

Задание 2. Рассчитать структурные характеристики интервального вариационного ряда: показатели центра распределения, показатели степени вариации, относительные показатели вариации; провести анализ структурных характеристик и сделать вывод о наличии или отсутствии засоренности. При наличии засоренности исключить засоряющие значения показателей из совокупности и произвести новые расчеты.

Задание 3. Произвести группировку населенных пунктов муниципального района с учетом численности населения.

Задание 4. По каждой группе населенных пунктов представить интервальный вариационный ряд по площади земельных участков, предназна-

ченных для строительства жилья; отразить интервальный вариационный ряд графически; рассчитать структурные характеристики интервального вариационного ряда; провести анализ структурных характеристик и сделать вывод о наличии или отсутствии засоренности. При наличии засоренности исключить засоряющие значения показателя из совокупности и произвести новые расчеты.

Задание 5. Провести анализ зависимости размера земельных участков, предоставляемых для строительства жилья, от численности населения в населенном пункте. Работа выполняется в программе Microsoft Excel. После выполнения задания работа оформляется с помощью программы Microsoft Word в соответствии с правилами оформления работы.

3. Отбор факторов, влияющих на стоимость земли.

Примерная структура РГР:

- Введение.
- Корреляционный анализ.
- Регрессия.
- Экономический анализ уравнения многофакторной регрессии.
- Заключение.
- Библиографический список.

Целью работы является ознакомление студентов с методом корреляционно-регрессионного анализа и возможностями его применения при ведении земельно-кадастровых работ.

В качестве индивидуального задания студенту выдается характеристика населенных пунктов муниципального района Пензенской области по основным социально-демографическим и экономическим факторам.

Используя метод корреляционно-регрессионного анализа, студент должен выполнить следующие задания:

Задание 1. Рассчитать линейные коэффициенты корреляции для всех исследуемых признаков; провести анализ тесноты связи между каждым факторным признаком и результативным; провести анализ тесноты связи между факторными признаками на наличие мультиколлинеарности; сделать соответствующие выводы.

Задание 2. Провести анализ существенности линейных коэффициентов корреляции между каждым факторным признаком и результативным; сделать соответствующий вывод.

Задание 3. Определить форму связи между ценообразующими факторами и стоимостью земли; подобрать уравнения парной регрессии; сделать соответствующие выводы.

Задание 4. Подобрать уравнение множественной регрессии, используя метод пошаговой регрессии; провести оценку адекватности уравнения множественной регрессии и существенности множественного коэффициента корреляции; провести анализ множественного коэффициента детерминации.

Задание 5. Определить, является ли построенное уравнение множественной регрессии корреляционно-регрессионной моделью; провести анализ возможности применения полученного уравнения множественной регрессии (корреляционно-регрессионной модели).

Задание 6. Выполнить экономический анализ коэффициентов уравнения множественной регрессии – коэффициентов регрессии, частных коэффициентов эластичности, нормированных коэффициентов регрессии.

Работа выполняется в программах Statistica и Microsoft Excel. После выполнения задания работа оформляется с помощью программы Microsoft Word в соответствии с правилами оформления работы.

Правила оформления работы

1. Работа должна быть напечатана на стандартном листе писчей бумаги в формате А4. Поля должны оставаться по всем четырём сторонам печатного листа: левое поле – 25 мм, правое – 10 мм, верхнее и нижнее – 20 мм.

2. Шрифт Times New Roman, размер шрифта – 14, интервал – полуторный. Выравнивание текста работы необходимо производить по ширине листа, отступ первой строки абзаца установить 12,5 мм.

3. Каждая новая работа начинается с новой страницы; это же правило относится к другим основным структурным частям работы (введению, заключению, библиографическому списку, приложениям и т.д.).

4. Все страницы работы должны быть пронумерованы сквозной нумерацией арабскими цифрами. Порядковый номер страницы ставится на середине нижнего поля. Первой страницей является титульный лист (номер на этой странице не проставляется). Второй страницей – содержание.

5. Титульный лист оформляется по установленному образцу.

6. Библиографический список оформляется согласно требованиям ГОСТ 7.1–2003.

7. Все ошибки и опечатки должны быть исправлены до того, как работа будет переплетена.

Порядок проведения защиты и критерии оценки РГР

Результат предварительной оценки работы отражается на титульном листе в форме вывода: «К защите» или «Не допускается к защите». Замечания преподавателя указываются в рецензии. Срок предварительной оценки не должен превышать четырех рабочих дней.

К защите допускаются работы с наличием исправлений или ответа на указанные замечания.

Если РГР не допускается к защите, она должна быть переработана студентом в соответствии с замечаниями и вновь представлена руководителю вместе с первым вариантом.

Защита РГР осуществляется в виде итогового собеседования с руководителем. При защите РГР учитываются:

- правильность решения задачи;
- самостоятельность и творческий подход к теме;
- логика аргументации и стройность изложения представленного материала;
- качество выполнения текстового и графического материала;
- полнота, правильность и аргументированность ответов при защите работы;
- своевременность представления работы.

Руководитель на титульном листе делает запись «Зачтено» в случае успешной защиты студентом РГР или «Не зачтено» при неудовлетворительной защите.

В обоих случаях указывается дата и ставится подпись преподавателя. Повторная защита РГР разрешается не ранее, чем через сутки.

Если студент три раза получает неудовлетворительную оценку по результатам защиты, то ему выдается новое задание.

При наличии незащищенной РГР, студент не допускается к сдаче экзамена по дисциплине «Современные методы статистического анализа кадастровых данных».

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Математическая обработка земельно-кадастровой информации [Текст]: учеб.-метод. пособие / Н.П. Шалдунова, Н.С. Денисова, Д.А. Кирик // М-во с.-х. РФ, ФГБОУ ВПО Пермская ГСХА. – Пермь: 2015. – 73 с.
2. Доклад о состоянии и использовании земель в Пензенской области в 2009 году.
3. Доклад о состоянии и использовании земель в Пензенской области в 2010 году.
4. Доклад о состоянии и использовании земель в Пензенской области в 2011 году.
5. Доклад о состоянии и использовании земель в Пензенской области в 2012 году.
6. Доклад о состоянии и использовании земель в Пензенской области в 2013 году.
7. <http://reftrend.ru/970662.html>. [Электронный ресурс]. – Дата обращения: 15.10.2015 г.

ПРИЛОЖЕНИЯ

Приложение А

Значение t -критерия Стьюдента

Число степеней свободы d, f	Уровень значимости P		
	0,10	0,05	0,01
1	6,3138	12,706	63,657
2	2,9200	4,3027	9,9248
3	2,3534	3,1825	5,8409
4	2,1318	2,7764	4,6041
5	2,0150	2,5706	4,0321
6	1,9432	2,4469	3,7074
7	1,8946	2,3646	3,4995
8	1,8595	2,3060	3,3554
9	1,8331	2,2622	3,2498
10	1,8125	2,2281	3,1693
11	1,7959	2,2010	3,1058
12	1,7823	2,1788	3,0545
13	1,7709	2,1604	3,0123
14	1,7613	2,1448	2,9768
15	1,7530	2,1315	2,9467
16	1,7459	2,1199	2,9208
17	1,7396	2,1098	2,8982
18	1,7341	2,1009	2,8784
19	1,7291	2,0930	2,8609
20	1,7247	2,0860	2,8453
21	1,7207	2,0796	2,8314
22	1,7171	2,0739	2,8188
23	1,7139	2,0687	2,8073
24	1,7109	2,0639	2,7969
25	1,7081	2,0595	2,7874
26	1,7056	2,0555	2,7787
27	1,7033	2,0518	2,7707
28	1,7011	2,0484	2,7633
29	1,6991	2,0452	2,7564
30	1,6973	2,0423	2,7500
40	1,6839	2,0211	2,7045
60	1,6707	2,0003	2,6603
120	1,6577	1,9799	2,6174
∞	1,6449	1,9600	2,5758

Приложение Б

Критические значения *F*-критерия Фишера–Снедекора
при уровне значимости 0,05

Количество на- блюдений	Количество факторных признаков							
	1	2	3	4	5	6	8	12
1	161,5	199,5	215,7	224,6	230,2	233,9	238,9	243,9
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92
100	3,94	3,09	2,70	2,46	2,30	2,19	2,03	1,85
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	3
1. АНАЛИЗ И ПРОГНОЗ ИСПОЛЬЗОВАНИЯ ЗЕМЕЛЬНЫХ РЕСУРСОВ ПЕНЗЕНСКОЙ ОБЛАСТИ.....	5
1.1. Характеристика использования земельных ресурсов Пензенской области	5
1.2. Анализ перераспределения земельных ресурсов Пензенской области по видам целевого использования	7
1.3. Прогноз распределения земель Пензенской области методом экстраполяции	13
2. АНАЛИЗ РАЗМЕРОВ ЗЕМЕЛЬНЫХ УЧАСТКОВ ПОД СТРОИТЕЛЬСТВО ЖИЛЬЯ В МУНИЦИПАЛЬНЫХ РАЙОНАХ.....	18
2.1. Характеристика и анализ размеров земельных участков, предназначенных для строительства жилья в муниципальных районах Пензенской области	18
2.2. Определение структурных характеристик вариационного ряда	23
2.3. Устранение засоренности совокупности	29
2.4. Анализ зависимости размера земельных участков, предоставляемых для строительства, от численности населения в населенном пункте	32
3. ОТБОР ФАКТОРОВ, ВЛИЯЮЩИХ НА СТОИМОСТЬ ЗЕМЕЛЬ	35
3.1. Корреляционный анализ	35
3.2. Регрессия	41
3.3. Экономический анализ уравнения множественной регрессии.....	46
ЗАДАНИЯ ДЛЯ ВЫПОЛНЕНИЯ РАСЧЕТНО-ГРАФИЧЕСКОЙ РАБОТЫ.....	49
БИБЛИОГРАФИЧЕСКИЙ СПИСОК	54
ПРИЛОЖЕНИЯ.....	55

Учебное издание

Акимова Мария Сергеевна
Улицкая Наталья Юрьевна

СОВРЕМЕННЫЕ МЕТОДЫ СТАТИСТИЧЕСКОГО АНАЛИЗА КАДАСТРОВЫХ ДАННЫХ

Учебно-методическое пособие
для выполнения расчетно-графической работы
по направлению подготовки 21.04.02 «Землеустройство и кадастры»

В авторской редакции
Верстка Т.А. Лильп

Подписано в печать 10.12.15. Формат 60×84/16.
Бумага офисная «Снегурочка». Печать на ризографе.
Усл.печ.л. 3,37. Уч.-изд.л. 3,63. Тираж 80 экз.
Заказ № 26.

Издательство ПГУАС.
440028, г. Пенза, ул. Германа Титова, 28.