

И.Н. Анисимова, канд. физ.-мат. наук;
Н.П. Баринов, канд. техн. наук;
С.В. Грибовский, д-р экон. наук,
г. Санкт-Петербург

Учет разнотипных ценообразующих факторов в многомерных регрессионных моделях оценки недвижимости

При решении задач индивидуальной оценки на развивающемся отечественном рынке недвижимости все большее значение приобретают методы сравнительного подхода, среди которых наиболее перспективными представляются связанные с применением многомерных регрессионных моделей [1-3].

Такие методы предполагают возможность построения статистических моделей, описывающих количественную зависимость исследуемого результирующего признака (стоимость, ставка аренды) от характеристик (влияющих признаков, ценообразующих факторов) объектов-аналогов, для которых известна рыночная ценовая информация. Решив проблему выбора достаточного числа объектов-аналогов [4] из имеющихся рыночных данных, оценщик должен преобразовать характеристики оцениваемого объекта и аналогов к виду, пригодному для построения многомерной регрессионной модели.

ПРЕОБРАЗОВАНИЕ ЗНАЧЕНИЙ ВЛИЯЮЩИХ ПРИЗНАКОВ

Характеристики (признаки) объектов недвижимости, выступающие в роли ценообразующих факторов регрессионной модели, могут иметь разнообразную природу. Часть из них являются количественными характеристиками (площадь объекта, соотношение общей и полезной площадей, коэффициент застройки и др.), другие – дискретными (год постройки, количество комнат, этаж), третьи носят качественный характер (наличие парковки, тип здания, особенности планировки квартиры и т.п.).

Каждому типу признаков соответствуют свой тип шкал измерений (количественные или неколичественные – порядковые, номинальные), группа допустимых преобразований значений шкалы и подмножество корректных методов обработки величин [5-8], применение которых не изменяет результата статистического моделирования.

Методы регрессионного анализа являются методами обработки количественных (числовых) величин. При этом разделение количественных признаков на непрерывные и дискретные в некоторой степени условно, поскольку из-за ограничений точности измерений даже непрерывные по своей природе показатели (например, площадь или расстояние) могут рассматриваться как дискретные. Однако с точки зрения практики решения оценочных задач, наоборот, допустимо рассматривать дискретный количественный признак как непрерывный, если число принимаемых им значений достаточно велико. Более того, во многих случаях методы, предназначенные для обработки непрерывных количественных данных, можно эффективно применять и для обработки дискретных признаков с небольшим числом градаций [5]. В частности, теория линейных регрессионных моделей с ненулевым свободным членом не накладывает никаких ограничений на характер значений количественных признаков – они могут быть непрерывными и дискретными, в том числе бинарными [9]. Примечательно, что оценки линейной регрессионной модели с ненулевым свободным членом инвариантны относительно линейных преобразований значений влияющих факторов, то есть для измерения значений признаков могут быть использованы произвольная точка отсчета шкалы и масштаб [10].

Шкалы количественных значений являются числовыми, то есть позволяют количественно измерять степень проявления некоторого свойства (признака) при

заданной единице измерения. Они применяются, когда имеется возможность численно оценить величину различий между значениями признака у разных объектов.

Если же эксперт в состоянии лишь разбить объекты на несколько классов по данному признаку, либо упорядочить объекты по интенсивности какого-либо их свойства без задания количественных соотношений между ними, то используются неколичественные шкалы – порядковые или номинальные.

Порядковые шкалы используются для отражения отношений между объектами. Порядковый признак обычно отражает различную степень проявления некоторого свойства, но не дает количественной меры для его выражения. Значения таких признаков могут быть заданы в баллах – когда объекты можно разбить на заранее известное число упорядоченных классов, или в рангах – когда объекты ранжируются, упорядочиваются – выстраиваются в ряд по степени проявления свойства, точная количественная мера для которого не определена. Ранг – это порядковый номер объекта в таком ряду.

Номинальные шкалы (шкалы наименований, классов) используются, когда эксперт может разбить объекты на классы – группы объектов, однородные по свойству, отражаемому некоторым признаком, но не в состоянии задать никакого естественного упорядочения между самими классами. Бинарный признак, принимающий всего две градации (например, 0 и 1) может быть рассмотрен как частный случай номинальной переменной.

Применение количественных шкал для измерения признаков требует, как правило, значительно более полной информации об объекте по сравнению с неколичественными шкалами – порядковой и, тем более, номинальной. На практике, когда имеющаяся рыночная информация оказывается недостаточно полной для точного определения значений количественного по природе признака, зачастую используются неколичественные шкалы. В этом случае количественный по своей сути признак можно рассматривать на качественном уровне (то есть как порядковый), приписывая каждой градации группу (диапазон) его значений.

Например, в [11] для решения задачи оценки объектов недвижимости рассматривается количественный показатель «удельные затраты на улучшение», хотя авторы замечают, что даже при выполненных ремонтных работах не всегда имеются объективные данные о величине понесенных затрат. Более того, не всегда затраты соответствуют вкладу в рыночную стоимость. При недостатке точной информации логичнее было бы рассмотреть этот фактор как дискретный количественный или качественный с произвольным числом градаций, отражающих некоторые интервалы удельных затрат на необходимое (или уже проведенное) улучшение. Если же и в этом случае имеющиеся рыночные данные окажутся недостаточными для точного задания значений признака для всех объектов, можно перейти к его заместителю – непосредственно наблюдаемому качественному признаку «состояние объекта». Такой переход оправдан тем, что выполнение неотделимых улучшений приводит к улучшению состояния объекта, то есть рассматриваемые факторы взаимозависимы. Обычно имеется достаточно информации, что определить состояние объекта на уровне самых общих градаций, например: «неудовлетворительное», «удовлетворительное», «хорошее» и «отличное».

Следует отметить, что такой перевод непрерывных количественных признаков в дискретные или качественные может сопровождаться потерей информации и не улучшает точность оценки. Поэтому этот прием оправдан лишь в случае угрозы ошибочного определения значений признака, которое может привести к значительно худшим последствиям при построении регрессионных моделей, нежели частичная потеря информации.

Для количественных показателей полагают адекватными математические модели непрерывных (случайных или не случайных) величин и используют «классический» аппарат многомерного статистического анализа. Для номинальных и порядковых шкал

корректны иные математические модели и методы, см., например, [12-14]. Номинальные признаки могут быть сведены без потери информации к системе бинарных признаков, для которых применимы как методы алгебры логики, так и обычные методы анализа количественных признаков. С порядковыми признаками с определенной осторожностью также можно обращаться как с количественными показателями, причем эффективность этого подхода во многом зависит от выбора числовых меток для градаций признака, с которыми далее можно оперировать как с числами (см. ниже).

Наибольшее прикладное значение имеет случай сочетания разнотипных признаков. Подходы к преобразованию информации при наличии разнотипных признаков для общего случая рассмотрены в [8]. Для задач индивидуальной оценки недвижимости, где наиболее предпочтительным является применение количественных методов регрессионного анализа, допустимы следующие преобразования исходной информации:

- номинальные признаки сводят к совокупности бинарных, а затем используют методы анализа количественных признаков. При этом некоторые порядковые признаки могут приниматься как квазиколичественные, другие – приводиться к системе бинарных переменных. Следует иметь в виду, что использование больших совокупностей бинарных признаков может затруднить интерпретацию модели, а также приводит к искусственному увеличению размерности пространства признаков.
- все признаки приводят к количественному типу [15], обычно с помощью оптимизационных процедур. Основным недостатком подхода считается то, что получаемые модели могут не иметь четкой экономической интерпретации.

Хороший результат, с точки зрения рассматриваемых задач, может дать совместное использование в регрессионных моделях количественных, квазиколичественных (полученных после оптимизации порядковых) и небольших совокупностей бинарных признаков. Данный подход позволяет в наибольшей степени учесть статистическую связь всех признаков, свести к минимуму потери информации при преобразованиях и использовать мощный аппарат классических методов анализа числовых данных. При этом в максимальной степени удастся сохранить степени свободы получаемых регрессионных уравнений, что особенно важно при ограниченном объеме рыночной информации.

ОЦИФРОВКА НЕКОЛИЧЕСТВЕННЫХ ПРИЗНАКОВ

Признаки нечисловой природы для учета их в регрессионной модели приводятся к квазиколичественному типу процедурой оцифровки, то есть путем присвоения их значениям некоторых числовых меток. Оцифрованные неколичественные признаки описываются обычно с помощью дискретных шкал с фиксированным количеством градаций. При этом качество регрессионной модели, построенной для разнотипных признаков, во многом зависит от выбранной процедуры оцифровки.

Из номинальных на практике чаще всего встречаются бинарные признаки, описывающие наличие/отсутствие какого-либо качества (отдельного входа, парковки, подъездных путей и т.п.). Бинарные признаки могут быть оцифрованы произвольным образом, однако в линейных регрессионных моделях из соображений наглядности их градациям чаще всего присваивают значения 0 и 1. Номинальный признак, имеющий более двух градаций, может быть описан системой бинарных переменных, хотя при этом происходит потеря степеней свободы регрессионного уравнения.

Порядковые величины предпочтительно измерять в баллах, поскольку в этом случае удобнее производить оцифровку. Следует отметить, что эксперт-оценщик обычно в состоянии высказать экономическую гипотезу о характере влияния признака на оцениваемую величину, хотя и не может дать четкого количественного выражения этого влияния. Поэтому, как правило, имеется возможность упорядочить классы объектов по их влиянию на значения результирующего показателя, то есть перейти от номинального к рассмотрению порядкового признака.

Например, номинальный признак «тип жилого дома» с градациями «хрущевка», «современный панельный», «современный кирпичный», «старый фонд» можно преобразовать в порядковый, высказав экономическую гипотезу о влиянии (в среднем) типа домов на цены квартир в них. Так, на основании риэлтерских данных, на рынке наиболее ценятся квартиры в современных кирпичных домах, наименее – в домах-«хрущевках». Поэтому для градаций номинального признака можно ввести отношение порядка, а значит перейти к порядковому признаку в бальном выражении: 1 – «хрущевка», 2 – «старый фонд», 3 – «современный панельный», 4 – «современный кирпичный».

Если порядковый признак имеет более двух градаций (как в рассмотренном примере), то при оцифровке существенным оказывается соотношение расстояний между соседними метками. Однако масштаб и точка отсчета выбираемой шкалы не имеют принципиального значения [8]. Так, две линейные регрессионные модели, в одной из которых оцифровка признака «состояние объекта» со значениями «удовлетворительное», «хорошее», «отличное» принята как 1, 2, 3 (соотношение между градациями $(3-2):(2-1) = 1:1$), а в другой – 0, 10, 20 (соотношение то же $(20-10):(10-0) = 10:10 = 1:1$) дадут эквивалентный результат. Однако он не совпадет с результатом модели с оцифровкой этого признака 1, 2, 4 (соотношение $(4-2):(2-1) = 2:1$). Выбор «правильного» соотношения между метками может осуществляться с помощью оптимизационных процедур (см. ниже) или экспертным путем.

Отметим, что еще до этапа оцифровки (присвоения числовых меток) эксперт-оценщик сталкивается с проблемой выбора градаций неколичественного признака (номинального или порядкового). Существенным при оцифровке неколичественных признаков для дальнейшего включения в линейную регрессионную модель оказываются:

- количество градаций признака (задает неявно соотношение между крайними градациями);
- порядок следования числовых меток градаций (для качественных признаков порядок следования градаций должен сохраняться);
- соотношение между градациями, то есть отношение расстояний (числа делений шкалы) между соседними градациями;

Выбор количества градаций

Неправильное разбиение на классы и выбор количества градаций могут привести к получению абсурдных результатов, не имеющих экономического смысла. Пример таких результатов изложен в [16], где рассмотрена плодотворная сама по себе идея использования объективной оптимизационной процедуры для оцифровки номинального признака, в данном случае - «этаж» при оценке стоимости жилых квартир. В качестве градаций признака автором работы был выбран формальный номер этажа, а после оцифровки получены следующие числовые метки (табл.1, рис.1).

Таблица 1. Числовые метки, соответствующие номерам этажей

Этаж	1	2	3	4	5	6	7	8	9	10
Балл	1	1.4	3.4	7	4.1	10.2	4.3	3.9	6.3	13.5

Учитывая, что задача решена для линейной регрессии, то есть зависимость результирующего признака «стоимость» от значений признака «этаж» предполагается монотонной, весьма странно выглядит, например, скачкообразная разница во влиянии 4, 5, 6 и 7 этажей. В самом деле, из опыта риэлтерской практики известно, что расположение квартиры на первом, а в некоторых случаях, и на последнем этаже является значимым фактором, заметно влияющим на цену квартиры; разница же в цене жилья для средних этажей незначительна.

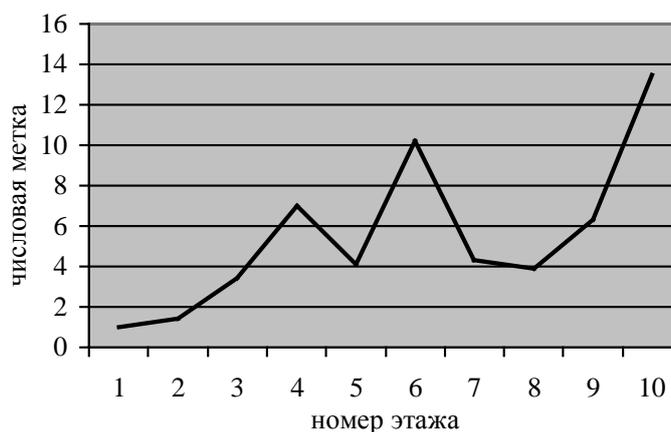


Рис.1. Оцифровка номинального признака «этаж» по [16]

Явная неадекватность полученной в [16] модели может быть вызвана несколькими причинами, и, прежде всего, неправильным выбором градаций признака:

- разбиение на классы (выбор количества градаций и порядка их следования) для признака производилось вне рассмотрения экономической гипотезы о характере его влияния на результирующий фактор (цену);
- выбраны лишние градации, в результате различия между числовыми метками определились не различиями в степени воздействия влияющего признака, а случайными колебаниями ценовых значений;
- оцифровка признака в многофакторной модели проведена без учета влияния других, возможно более значимых, признаков. В этом случае была сделана попытка объяснить колебания цен, вызванные влиянием неучтенных моделью факторов, за счет вариации лишь одного признака. Скачкообразные изменения числовых меток для средних этажей вызваны, скорее всего, именно неучтенным влиянием других факторов.

Как видно, наличие «лишних» градаций может приводить к усилению влияния случайных колебаний или других факторов даже при использовании объективных методов выбора числовых меток (оптимизационных процедур, бинарных переменных). Другими словами, допустимое число градаций должно согласовываться с фактической инструментальной погрешностью измерений свойства. При излишней детализации эксперт может допускать ошибки измерения значений влияющего признака, что с точки зрения теории регрессионных моделей крайне нежелательно, так как может привести к смещению и несостоятельности оценок [9]. Кроме того, излишняя детализация может потребовать неоправданно больших затрат на сбор рыночной информации. Отметим также, что при большом числе реально используемых градаций (классов, баллов) процедура оценивания значения порядкового признака в баллах приближается по содержанию к количественному оцениванию.

В практических задачах при разбиении на классы (как для номинальных, так и для измеренных в бальной шкале порядковых признаков) обычно рекомендуется использовать **3-7** градаций [17, 18]. Статистическая процедура, позволяющая вычислить необходимое число градаций в зависимости от диапазона допустимых количественных изменений признака и дисперсии ошибок ответов (для нашего случая – дисперсии ошибок экспертов при определении значений признака), приведена в [17]. Практическая рекомендация по выбору числа градаций неколичественного признака может быть сформулирована так: число градаций равно возможному числу классов рассматриваемых объектов. При определении же числа и границ классов значения рассматриваемого свойства в рамках одного класса должны быть однородными, а между классами – существенно различаться.

Например, при рассмотрении номинального признака «этаж» при оценке стоимости квартир в упомянутой работе [16] было неправильным в качестве градаций вводить физический номер этажа: 1,2,3,...10, поскольку из оценочной практики известно,

что различия в расположении на средних этажах не оказывают существенного влияния на стоимость квартир. В то же время, существенным недостатком квартиры, снижающим ее цену, является расположение на первом (без учета возможности перевода ее в нежилой фонд) и, в меньшей степени, – на последнем этаже. Исходя из этого номинальной переменной «этаж» можно сопоставить градации «первый этаж», «средние этажи», «последний этаж». Эти градации можно упорядочить в соответствии с предполагаемым увеличением цены квартир: 1 – «первый этаж», 2 – «последний этаж», 3 – «средние этажи», а признак «этаж» рассматривать далее как качественную переменную с тремя градациями.

Таким образом, уже *на первом шаге* процедуры оцифровки неколичественного признака экспертом-оценщиком *должна быть выдвинута экономическая гипотеза о характере* его влияния. После этого выбор градаций признака осуществляется с учетом следующих соображений:

- разбиение на классы (градации) должно производиться на основе выявления существенных различий, оказывающих заметное влияние на значение результирующей величины;
- количество градаций (степень детализации признака) должно быть согласовано с фактической погрешностью определения значений признака, определяемой, в основном, полнотой рыночных данных; рекомендуемое количество градаций – 3-7;
- упорядочение градаций должно производиться исходя из предполагаемой степени влияния признака на результирующую величину, а не по интенсивности проявления самого свойства объекта.

Процедуры оцифровки признаков

Существует несколько подходов к оцифровке признаков неколичественной природы.

Сведение к совокупности бинарных переменных [8, 9, 11, 19, 20], которые в эконометрической литературе чаще называются фиктивными, искусственными или структурными.

Этот метод достаточно универсален, поскольку подходит для оцифровки как номинальных, так и порядковых признаков. Кроме того, он объективен с точностью до количества градаций, поскольку значение градации фактически определяется вкладом фиктивной переменной, то есть самой регрессионной моделью.

В классической линейной регрессионной модели ищется зависимость в виде:

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_jx_j + \dots + a_kx_k. \quad (1)$$

Для учета неколичественного признака с m градациями требуется введение $m-1$ бинарной переменной. Для описания признака x_j с градациями $\{x_j^1, x_j^2, \dots, x_j^m\}$ вводятся бинарные переменные z_1, z_2, \dots, z_{m-1} . Для одного из значений признака, например, для x_j^1 , значения всех $z_q, q=1, 2, \dots, m-1$ полагаются равными нулю. Для остальных градаций:

$$x_j^{q+1} (q=1, 2, \dots, m-1) \text{ полагается } z_q=1, z_p=0, p \neq q.$$

Регрессионное уравнение (1) переписывается в виде (2):

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_{j-1}x_{j-1} + b_1z_1 + b_2z_2 + \dots + b_{m-1}z_{m-1} + a_{j+1}x_{j+1} + \dots + a_kx_k. \quad (2)$$

Например, номинальному признаку «тип жилого дома» с градациями «хрущевка», «современный кирпичный», «современный панельный», «старый фонд» можно сопоставить три бинарные переменные z_1, z_2 и z_3 . При этом $z_1=1$ для современных кирпичных домов, $z_2=1$ для панельных домов, $z_3=1$ для домов старого фонда; в остальных случаях переменные принимают значение 0. Тогда объекты в домах-«хрущевках» описываются тройками значений 0,0,0 ($z_1=0, z_2=0, z_3=0$); объектам в современных

кирпичных домах будут сопоставлены тройки 1,0,0; объектам в панельных домах – 0,1,0; объектам в домах старого фонда – 0,0,1.

В то время как число степеней свободы уравнения (1) с k влияющими признаками равно $n-k-1$, в уравнении (2) вместо одного из признаков используется $m-1$ переменная, и значит число степеней свободы этого уравнения уменьшится на $m-2$: $n-(k-1+m-1)-1 = n-k-m+1$. Отсюда, если признак имеет всего две градации, то число степеней свободы уравнения не изменится. Если же m велико, то переход к фиктивным переменным существенно уменьшает число степеней свободы регрессионной модели, что неприемлемо в условиях малой выборки (при небольших n), характерных для задач индивидуальной оценки. С другой стороны, при небольшом количестве градаций, значения фиктивных переменных часто оказываются сильно сопряженными [8], что также может ухудшить качество модели. Поэтому подход на основе использования совокупности бинарных (фиктивных) переменных хорош для задач массовой оценки (когда n – велико), в то время как его применение на практике для большинства случаев задач индивидуальной оценки затруднено из-за ограниченности объема рыночных данных и «дефицита» степеней свободы регрессионных моделей.

Если все же сформированная выборка объектов аналогов оказалась достаточно многочисленной для построения адекватной регрессионной модели с фиктивными переменными и нахождения оценок параметров $a_0, a_1, \dots, a_{j-1}, b_1, b_2, \dots, b_{m-1}, a_{j+1}, \dots, a_k$ уравнения (2), то в дальнейшем можно произвести оцифровку исходной переменной x_j и построить регрессионную модель вида (1) с большим числом степеней свободы, а значит, и с лучшими (более точными) статистическими оценками [19].

В самом деле, из вида фиктивных переменных z_q следует, что для объектов первого класса, у которых значение признака x_j совпадает с первой градацией x_j^1 , уравнение (2) примет вид

$$y = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_{j-1} x_{j-1} + a_{j+1} x_{j+1} + \dots + a_k x_k \quad (\text{так как все } z_q = 0);$$

для объектов со значением признака, равным второй градации x_j^2 , уравнение (2) переписывается как

$$y = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_{j-1} x_{j-1} + b_1 + a_{j+1} x_{j+1} + \dots + a_k x_k \quad (z_1 = 1, z_2 = z_3 = \dots = z_{m-1} = 0);$$

для q -того класса объектов

$$y = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_{j-1} x_{j-1} + b_{q-1} + a_{j+1} x_{j+1} + \dots + a_k x_k, \text{ и, наконец,}$$

для объектов m -того класса

$$y = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_{j-1} x_{j-1} + b_{m-1} + a_{j+1} x_{j+1} + \dots + a_k x_k.$$

Те же зависимости были бы получены, если бы переменная x_j вошла в регрессионное уравнение (1) с коэффициентом $a_j = 1$ и градациями $0, b_1, b_2, \dots, b_{q-1}, \dots, b_{m-1}$. Поэтому при построении регрессионной модели (1) в качестве числовых меток \tilde{x}_j^q градаций x_j^q признака x_j могут быть взяты оценки коэффициентов b_{q-1} при фиктивных переменных регрессионного уравнения (2):

$$\tilde{x}_j^1 = 0; \tilde{x}_j^2 = b_1; \dots; \tilde{x}_j^{m-1} = b_{m-1}. \quad (3)$$

В оцифровке (3) могут быть использованы и значения b_{q-1} , известные из моделей вида (2) массовой оценки, в предположении, что выявленные массовой оценкой закономерности ценообразования сохраняются и для рассматриваемого сегмента рынка недвижимости.

Если известны лишь нелинейные модели массовой оценки, то обычно в качестве меток рекомендуется брать известные средние удельные цены объектов каждого класса [11, 19], что аналогично подходу на основе оптимизации вида (4), рассмотренного ниже. Кроме того, соотношение коэффициентов для разных классов объектов в нелинейной модели может дать дополнительную информацию оценщику для экспертного назначения числовых меток или их начальных значений для оптимизационных процедур.

Для номинальных признаков использование совокупностей бинарных переменных является наиболее естественным способом оцифровки. Во избежании проблем «дефицита» степеней свободы регрессионных моделей при решении задач индивидуальной оценки следует, по возможности, уклоняться от использования большого числа номинальных признаков путем перевода их в порядковые (на основе экономических гипотез об отношениях порядка между классами, как это было рассмотрено выше). Для номинальной переменной возможно также сокращение числа учитываемых в регрессионной модели градаций (например, до двух) за счет соответствующего подбора объектов-аналогов (повышения однородности выборки).

Использование равномерного кодирования для неколичественных признаков, когда расстояние между числовыми метками соседних градаций одинаково. Например, «удовлетворительное», «хорошее», «отличное» состояние $\rightarrow 1, 2, 3$.

Такая кодировка весьма груба и может не отражать реальную степень отличия градаций фактора. Вместе с тем в задачах индивидуальной оценки, где рассматривается совокупность близких объектов, незначительно отличающихся друг от друга по своим характеристикам, даже такой весьма грубый подход может дать приемлемые результаты (см., например, сравнение моделей с равномерной и неравномерной оцифровкой в [11]).

Несколько сгладить недостатки, присущие равномерному кодированию, позволяет использование порядковой шкалы качественных оценок [1, 11] – табл.2. При этом, однако, задание «неравномерности» числовых меток полностью возлагается на эксперта, то есть весьма субъективно, а в ряде случаев – затруднительно, поскольку, как отмечалось ранее, численная оценка градаций факторов должна производиться не по степени выраженности свойства объекта, а по степени предполагаемого влияния этого признака на результирующую величину.

Таблица 2. Порядковая шкала качественных оценок

Градация оценки фактора	Значения порядковой шкалы
Наихудшее значение фактора	1 – 2
Незначительное преимущество	3 – 4
Значительное преимущество	5 – 6
Явное преимущество	7 – 8
Абсолютное преимущество	9

Альтернативой субъективному экспертному подходу является использование оптимизационных процедур [8, 16, 20-22] при оцифровке признаков. Оптимизационные методы оцифровки основаны на том, что числовые метки, присваиваемые градациям, должны быть «разумны» в рамках решаемой задачи. В частности, в рамках регрессионного анализа оптимизация основана на принципе максимизации зависимости между влияющей (x_j) и результирующей (y) переменными. Подход на основе оптимизационных процедур так же, как и подход на основе использования систем фиктивных переменных, объективен с точностью до количества градаций признака.

В качестве оптимизационных могут быть использованы следующие критерии, являющиеся взаимосвязанными:

- максимизация коэффициента сопряженности между x_j и y : $r_{y x_j} \rightarrow \max$;
- минимизация остаточной разности квадратов $\sum_{i=1}^n \varepsilon_i^2 \rightarrow \min$;
- максимизация коэффициента детерминации $R^2 \rightarrow \max$.

Перечисленные критерии сами по себе не накладывают никаких ограничений на порядок следования градаций признака, так что после оцифровки он может измениться. Для номинальных признаков и в случае, если порядковая переменная отражает лишь степень проявления некоторого качества объекта недвижимости безотносительно к его

влиянию на зависимый признак, изменение порядка следования градаций не критично. Однако если первоначальные метки градациям были назначены экспертом исходя из обоснованной экономической гипотезы влияния на результирующий признак, изменение их следования может свидетельствовать о неправильном выборе градаций признака или спецификации регрессионной модели.

В [16, 21] для оцифровки признаков предложено использовать оптимизационные процедуры *Поиск решения* MS Excel. Отмечая доступность и удобство данного математического аппарата, следует понимать и ограничения его применения. По умолчанию в MS Excel применяются алгоритмы нелинейной оптимизации (метод Ньютона, метод сопряженных градиентов), сходимость которых определяется, в частности, начальными условиями, то есть тем, как были оцифрованы градации признака перед запуском процедуры оптимизации. Кроме того, эти алгоритмы могут находить не главный, а лишь локальный экстремум (минимум, максимум), не представляя пользователю возможности различать эти ситуации.

Наилучших результатов при использовании инструмента *Поиск решения* MS Excel можно добиться, если в качестве начальных значений для оптимизационной процедуры (то есть в качестве начальной приближенной оцифровки) использовать значения, более-менее близкие к результирующим. Пример экономически обоснованного и весьма удачного выбора начальных значений приведен в [21], неудачного – в [16]. Применительно к решаемым задачам оценки недвижимости, в качестве начальной оцифровки перед применением нелинейной оптимизационной процедуры можно использовать неравномерную кодировку, задаваемую экспертным путем на основе содержательного анализа задачи оценки и имеющихся рыночных данных.

Вместе с тем, для линейной регрессионной модели может быть применен прозрачный метод оптимизации, заключающийся в том, что каждой градации x_j^q признака x_j ставится в соответствие среднее арифметическое наблюдаемых значений y_i зависимого признака по всем объектам, которые имеют то же значение градации $x_{ij}=x_j^q$. Пусть в исходной выборке данных, состоящей из n объектов, набралось n_q объектов, у которых значение рассматриваемого фактора совпало с градацией x_j^q . Тогда этой градации можно присвоить числовую метку \tilde{x}_j^q :

$$\tilde{x}_j^q = \frac{1}{n_q} \sum_{x_{ij}=x_j^q} y_i . \quad (4)$$

Такая перекодировка хорошо интерпретируема и максимизирует корреляцию y и x_j . Аналогом указанной процедуры является построение зависимости y только от совокупности фиктивных переменных, описывающих неколичественных признаков, без учета влияния остальных факторов [19]:

$$y = a_0 + b_1 z_1 + \dots + b_{m-1} z_{m-1},$$

а затем также использование значений коэффициентов в качестве числовых меток.

На сходной идее основано и использование в качестве числовых меток известных средних удельных цен для разных классов объектов [1, 11].

Заметим, что в многомерном случае такой подход может быть обоснован только для наиболее значимых факторов, влияние которых на значения y очевидно. То есть такая «прямая» оптимизационная процедура применима лишь в тех случаях, когда значения зависимой переменной явно отражают характер влияния градаций признака. В многофакторной модели такие случаи не так уж часты, поскольку значения y формируются в результате совокупного влияния многих факторов. Поэтому для второстепенных признаков, влияние которых на y прослеживается не столь явно, полученные по формуле (4) числовые метки могут противоречить экономическому смыслу.

В этом случае рекомендуется использовать метод *последовательного числового перекодирования* [8, 22].

Пусть построена регрессионная модель вида (1), в которую включено k_1 уже оцифрованных признаков (количественные, бинарные и уже оцифрованные неколичественные признаки). Тогда в качестве числовых меток \tilde{x}_j^q для градаций x_j^q нового неколичественного влияющего фактора, можно рассмотреть средние арифметические остатков ε_i , рассчитанные для тех объектов, у которых значения данного признака совпадают с градацией x_j^q :

$$\tilde{x}_j^q = \frac{1}{n_q} \sum_{x_{ij}=x_j^q} \varepsilon_i, \quad (5)$$

где $\varepsilon_i = y_i - \tilde{y}_i$ – разности между наблюдаемыми и модельными значениями результирующего признака, то есть та часть реальных рыночных цен, которую не удалось объяснить с помощью регрессионной модели с k_1 переменными.

Таким образом, в отличие от «прямой оптимизации» (4) при последовательном перекодировании (5) рассматривается влияние признака на еще необъясненную моделью часть наблюдаемых ценовых значений. Оцифровка (5) является оптимизационной – она минимизирует остаточную разность квадратов регрессии.

Вновь оцифрованный признак включается в модель (теперь с k_1+1 влияющим фактором), на основе которой может быть произведена оцифровка следующего неколичественного признака, и т.д.

В литературе не освещен вопрос о порядке, в котором следует оцифровывать признаки. Представляется логичным проводить оптимизацию значений сначала более значимых факторов, затем – менее. Высказать предположение о большей или меньшей значимости факторов до построения регрессионной модели эксперт может на основании известных результатов проведенных ранее массовых оценок, либо на основе содержательного анализа рыночных данных. Если будет выбран неверный порядок оцифровки признаков, то, скорее всего, получаемые числовые метки будут нарушать заданный порядок следования градаций признака. Поэтому при применении процедур оцифровки результаты вычислений могут служить для проверки высказанных экспертом экономических гипотез, и наоборот, *следует проверять полученные значения числовых меток на соответствие экономическому смыслу*.

Необходимо также отметить, что из выражений (4), (5) следует, что для более-менее надежного определения числовых меток значения градаций в выборке исходных данных должны повторяться. То есть для каждой градации q число $n_q > 1$ (желательно, $n_q \geq 3$ – число 3 выбрано как признак достаточной множественности).

Проиллюстрируем применение метода последовательного числового перекодирования на примере.

Пусть ставится задача оценки элитной квартиры на Невском проспекте. Такая квартира отличается, прежде всего, своим особым местоположением – на основной, символьной магистрали города, поэтому в качестве аналогов были выбраны квартиры, также расположенные на Невском проспекте. Все объекты находятся в домах исторической застройки Санкт-Петербурга, кроме того, все они расположены на средних этажах и не имеют явно выраженных дефектов. Поэтому такие характерные для оценки квартир влияющие факторы, как «местоположение» (удаленность от центра/метро), «тип здания», «этаж» принимаются равнозначными и не рассматриваются в регрессионной модели.

Состояние квартир-аналогов в зависимости от имеющихся улучшений оценено как удовлетворительное и хорошее. Кроме того, среди аналогов имеются квартиры, расположенные в домах с проведенным капитальным ремонтом, в них выполнены перепланировка и дизайнерские работы по оформлению интерьера. Состояние таких

квартир оценивалось как отличное. В качестве фактора, увеличивающего стоимость квартиры, отмечено наличие благоустроенного по европейским стандартам двора (по типу дворов Капеллы, «итальянских» двориков). Наличие неблагоустроенного двора, по мнению экспертов, не увеличивает ценовые значения квартир.

Кроме того, понятие элитной квартиры требует особого окружения. Исходя из этого, экспертами отмечена разница в ценовых характеристиках квартир, расположенных в различных частях дома. Так, большинство домов на Невском проспекте имеют небольшую фронтальную часть, расположение квартир в которой считается самым престижным. К ним приравнены квартиры, имеющие вид на памятники архитектуры в непосредственной близости от Невского проспекта. Дома на Невском имеют, как правило, вытянутую вглубь форму, зачастую с множеством внутренних проходных дворов-колодцев. Поэтому далее по степени престижности следуют квартиры, выходящие на прилегающие к Невскому улицы. Расположение квартиры внутри дома в удалении от фронтальной части с проходом через внутренние дворы и выходящие во внутреннюю территорию дома считается наименее благоприятным. Соответственно, экспертами был введен еще один влияющий фактор - местоположение внутри дома, названный «вид» с градациями «фронтальная часть», «улица» и «двор» (табл. 3).

Таблица 3. Исходные данные для оценки квартиры на Невском проспекте

№	Адрес	Площадь, кв.м.	Цена за 1 кв.м, \$,	Состояние здания	Двор	Вид
1	Невский пр. д 22	113.0	696.9	удовл	отсутств.	улица
2	Невский пр. д 22	150.0	600.0	хор	отсутств.	двор
3	Невский пр. д 51	89.0	741.6	удовл	отсутств.	улица
4	Невский пр. д 51	113.0	663.7	хор	сущ.	двор
5	Невский пр. д 64	138.5	752.7	удовл	отсутств.	фронт
6	Невский пр. д 84	70.0	685.7	удовл	евро	фронт
7	Невский пр. д 90	72.0	540.6	удовл	евро	двор
8	Невский пр. д 90	170.0	1614.7	отл	евро	улица
9	Невский пр. д 90	143.0	550.7	удовл	евро	двор
10	Невский пр. д 92	120.0	593.8	удовл	евро	двор
11	Невский пр. д 94	137.0	1615.0	отл	евро	улица
12	Невский пр. д 110	82.0	900.0	хор	отсутств.	улица
13	Невский пр. д 106	68.0	562.5	удовл	сущ.	двор
14	Невский пр. д.88	75.0	790.0	удовл	отсутств.	улица
объект	Невский пр. д.102	72.3		отл	отсутств.	двор

Присвоим значениям неколичественных признаков «состояние», «двор» и «вид», пользуясь процедурой равномерной оцифровки и высказанным гипотезам о влиянии признаков согласно таблице 4.

Таблица 4. Числовые метки при равномерной оцифровке

Наименование признака	Наименования градаций	Числовые метки
Состояние	удовлетворительное (удовл)	1
	хорошее (хор)	2
	отличное (отл)	3
Двор	двор отсутствует (отсутств)	0
	имеется неблагоустроенный двор (сущ)	0
	имеется благоустроенный двор (евро)	1
Вид	двор	1
	улица	2
	фронтальная часть (фронт)	3

Исходные данные после оцифровки неколичественных влияющих признаков X_2 , X_3 , X_4 представлены в табл.5. По этим данным построена регрессионная модель вида (1) и получена оценочная величина стоимости объекта $C_0=1208.51$. Статистические характеристики модели: коэффициент детерминации (правленный) $R^2=0.797$, СКО $s=160.55$, средняя ошибка аппроксимации $A=14.4\%$.

Таблица 5. Исходные данные после оцифровки

№	Адрес	Цена, \$ за 1 кв.м.	Площадь, кв.м.	Состояние здания	Двор	Вид
			X_1	X_2	X_3	X_4
1	Невский пр. д 22	696.9	113.0	1	0	2
2	Невский пр. д 22	600.0	150.0	2	0	1
3	Невский пр. д 51	741.6	89.0	1	0	2
4	Невский пр. д 51	663.7	113.0	2	0	1
5	Невский пр. д 64	752.7	138.5	1	0	3
6	Невский пр. д 84	685.7	70.0	1	1	3
7	Невский пр. д 90	540.6	72.0	1	1	1
8	Невский пр. д 90	1614.7	170.0	3	1	2
9	Невский пр. д 90	550.7	143.0	1	1	1
10	Невский пр. д 92	593.8	120.0	1	1	1
11	Невский пр. д 94	1615.0	137.0	3	1	2
12	Невский пр. д 110	900.0	82.0	2	0	2
13	Невский пр. д 106	562.5	68.0	1	0	1
14	Невский пр. д.88	790.0	75.0	1	0	2
объект	Невский пр. д.102		72.3	3	0	1

Как видно, при приемлемом, в целом, качестве модели, полученные оценки не слишком хороши (велико значение СКО s и средней ошибки аппроксимации A). Подобный результат неудивителен, так как числовые метки градациям признаков были присвоены субъективно с помощью равномерной оцифровки, в то время как эксперт-оценщик предполагает, что влияние градаций факторов «состояние», «вид» или «общая площадь» должно быть неравномерно. Например, очевидно, что квартиры, состояние которых оценено как отличное, будут стоить существенно дороже остальных. Кроме того, расположение квартиры в отдалении от фронтальной линии является существенным недостатком для рассматриваемого класса квартир и, по всей видимости, должно сильно отличаться от двух других градаций.

Постараемся повысить объективность оцифровки с помощью процедуры оптимального выбора меток на основе метода последовательного числового перекодирования. Построим двухфакторную модель, в которой учтены количественный признак «общая площадь» и бинарный признак «двор». На основе двухфакторной модели получены следующие модельные ценовые значения и вычислены величины отклонений (табл.6):

Таблица 6. Результаты модели с двумя влияющими факторами

№	1	2	3	4	5	6	7	8	9	10	11	12	13	14
\tilde{y}_i	754.2	914.0	650.5	754.2	864.4	723.2	731.8	1155.2	1038.5	939.2	1012.6	620.3	559.8	590.0
ε_i	-57.3	-314.0	91.1	-90.5	-111.7	-37.5	-191.2	459.5	-487.8	-345.4	602.4	279.7	2.7	199.0

Начнем с признака «состояние», поскольку, по мнению эксперта-оценщика, как и по результатам модели с равномерной оцифровкой, этот признак оказывает наиболее существенное влияние на ценообразование. Применим формулу (5) для расчета числовых меток для градаций признака «состояние»:

удовлетворительное	хорошее	отличное
937.1 / 9 = -104.1;	-124.8 / 3 = -41.6;	1061.9 / 2 = 530.9.

Порядок числовых меток соответствует порядку следования градаций признака.

Попробуем применить формулу (5) для расчета числовых меток признака «вид»:

двор	улица	фронтальная часть
-1426.2 / 6 = -237.7	1575.3 / 6 = 262.6	-149.1 / 2 = -74.6

Как видно, для этого признака порядок числовых меток не соответствует порядку следования градаций, поэтому первым будем оцифровывать признак «состояние».

Построим модель с тремя влияющими факторами: «общая площадь», «двор» и «состояние» с оптимизированными числовыми метками. Статистика R^2 для этой модели достигает значения 0.924.

Таблица 7. Результаты модели с тремя влияющими факторами

№	1	2	3	4	5	6	7	8	9	10	11	12	13	14
\tilde{Y}_i	663.6	728.4	689.9	768.9	635.7	630.3	628.1	1590.1	550.4	575.5	1626.2	802.8	712.9	705.2
ε_i	33.3	-128.4	51.7	-105.2	117	55.5	-87.4	24.6	0.3	18.2	-11.2	97.2	-150.4	84.8

Если рассмотреть модель с четырьмя факторами, добавив признак «вид» с неоптимизированной равномерной оцифровкой, получим: $R^2=0.969$, $A=6.2\%$, $C_0=1634.32$. Посмотрим, внесет ли какие-то улучшения оптимизация оцифровки признака «вид».

На основе результатов модели с тремя влияющими факторами (табл. 7), вычислим числовые метки для признака «вид»:

двор	улица	фронтальная часть
-452.8 / 6 = -75.5	280.3 / 6 = 46.7	172.5 / 2 = 86.2

Как видно, числовые метки не нарушают порядка следования градаций признака «вид» и соответствуют выдвинутой экспертом экономической гипотезе о характере влияния признака на результирующую величину стоимости.

Построим окончательно четырехфакторную модель, используя оптимизированные числовые метки для признака «вид». В финале имеем: оценочная величина стоимости объекта $C_0=1540.86$, $R^2=0.982$, $s=48.26$, $A=4.5\%$. Изменение оценочного значения стоимости объекта оценки по сравнению с неоптимизированной моделью составило 27.5%, остаточное СКО и средняя ошибка аппроксимации уменьшились в 3 (!) раза по сравнению с первоначальными.

Как видно, применение оптимизационного метода последовательного числового перекодирования для оцифровки неколичественных признаков позволило существенно улучшить качество регрессионной модели.

Поскольку для рассматриваемой выборки сильное влияние состояния объектов на ценовые значения очевидно, для признака «состояние» можно было бы применить «прямую» оцифровку согласно (4). Такая оцифровка приведет к получению схожих результатов (относительные изменения основных показателей C_0 и s не превышают 1.8%):

- признак «состояние» – «прямая» оцифровка по (4):

удовлетворительное	хорошее	отличное
657.2	721.2	1614.8

- признак «вид» – оцифровка на основе метода последовательного числового перекодирования по соотношению (5):

двор	улица	фронтальная часть
-70.1	43.2;	80.9

- оценочное значение стоимости объекта $C_0= 1563.12$, $R^2=0.980$, $s=50.73$, $A=4.7\%$.

Для рассмотренного примера оценки квартир на Невском проспекте были проведены и другие процедуры оптимизации числовых меток признаков «состояние» и «вид»:

- Построена линейная регрессионная модель с количественной переменной x_1 «общая площадь», бинарной переменной x_2 «двор» и двумя совокупностями бинарных переменных z_1 и z_2 – для признака «состояние», v_1 и v_2 – для признака «вид» (напомним, что для описания m градаций требуется совокупность $m-1$ бинарных переменных):

$$y = a_0 + a_1 x_1 + a_2 x_2 + b_1 z_1 + b_2 z_2 + c_1 v_1 + c_2 v_2.$$

Для этой регрессионной модели с шестью переменными получены следующие оценки:
 $b_1=102.5$, $b_2=873.9$, $c_1=199.6$, $c_2=166.9$; $C_0=1434.02$, $R^2=0.985$, $s=43.98$, $A=3.6\%$.

После определения числовых меток по формуле (3) для признаков «состояние» и «вид» статистические оценки модели (R^2 , s) несколько улучшились (табл.8), что объясняется увеличением степеней свободы регрессионной модели при переходе от шести к четырем переменным.

- Выполнена процедура одновременной нелинейной оптимизации *Поиск решения* MS Excel для обоих признаков. Первоначально в среде MS Excel построена четырехфакторная линейная регрессионная модель и выведена статистика по ней с помощью стандартной функции ЛИНЕЙН. Для признаков «состояние» и «вид» при этом применена равномерная оцифровка градаций 1, 2, 3. Далее запущена процедура *Поиск решения*. В качестве изменяемых данных использованы ячейки, содержащие первоначальные значения числовых меток признаков «состояние» и «вид», в качестве оптимизационного критерия выбран минимум значения СКО.

- Выполнен *Поиск решения* для тех же данных и с теми же начальными значениями, но с максимизацией R^2 в качестве критерия оптимизации.

- Проведена последовательная двухэтапная оптимизация в MS Excel с помощью процедуры *Поиск решения*. Сначала построена линейная трехфакторная регрессионная модель (с переменными «общая площадь», «двор» и «состояние») и подобраны метки признака «состояние». Затем построена модель с четырьмя влияющими признаками с использованием определенных на предыдущем шаге числовых меток для признака «состояние» и вычислены метки для признака «вид». В качестве оптимизационного критерия на обоих этапах выбран минимум СКО.

- Проведена последовательная аналогичная двухэтапная оптимизация по максимуму R^2 .

Результаты моделирования, полученные с помощью различных процедур оптимизации сведены в табл.8.

Как видно, результаты применения различных оптимизационных процедур для выбора числовых меток могут различаться (в частности, размах оценок стоимости объекта составил 7%). Интересно, что выбор в качестве критерия оптимизации минимума СКО или максимума R^2 не меняет результатов оптимизации; одновременное применение процедуры *Поиск решения* для обоих признаков аналогично оцифровке с использованием совокупностей бинарных переменных, а поэтапное применение той же процедуры *Поиск решения* дает отличающиеся результаты. Конечно, полноценное сравнение этих процедур возможно только после многократного их тестирования на различных примерах. Тем не менее основной вывод сделать можно – к процессу назначения числовых меток недопустимо подходить как к чисто математической процедуре.

В самом деле, при выборе процедуры оцифровки, казалось бы, в первую очередь следует ориентироваться на достижение наилучших значений интегральных оценок качества регрессионной модели в целом – СКО, коэффициента детерминации, ошибки аппроксимации, критерия Фишера. С этих позиций следовало бы отдать предпочтение оцифровкам 3-5. Однако следует иметь в виду, что применение любой из описанных выше оптимизационных процедур приводит к «подгонке» исходных данных под выбранную экспертом модель, в то время как смыслом экономического моделирования является,

скорее, нахождение модели, наиболее адекватно отображающей реально существующие рыночные данные. В частности, оптимизационная процедура, в которой не ведется контроль соответствия содержательной стороне задачи, стремится «объяснить» с помощью варьируемых ею факторов *все* воздействия на результат, вызванные в том числе и не учитываемыми моделью факторами. Поэтому следует соблюдать определенную осторожность при выборе оптимизационной процедуры, помня, что решаемая задача носит экономический, а не абстрактно-математический характер.

В рассматриваемом примере все примененные процедуры оптимизации, кроме последовательной числовой перекодировки 1 и основанной на ней 2, нарушают порядок следования градаций признака «вид». Поэтому, с точки зрения авторов, именно процедурам присвоения числовых меток 1 и 2 следовало бы отдать предпочтение как не нарушающим экономического смысла решаемой задачи.

Таблица 8. Сравнение оптимизационных процедур оцифровки

Процедура получения оцифровки	1	2	3	4	5	6	7
Характеристики оцифровки и регрессионной модели	Последовательное перекодирование	«состояние» – прямое кодирование, «вид» - последовательное перекодирование	Оцифровка на основе бинарных переменных	Поиск решения Excel, критерий – min СКО	Поиск решения Excel, критерий – max R ²	Двухэтапный Поиск решения, min СКО	Двухэтапный Поиск решения, max R ²
Числовые метки признака «состояние»	-104.1 -41.6 530.9	657.2 721.2 1614.8	0 102.5 873.9	1.5 1.6 2.9	1.2 1.5 3.3	1.2 1.3 3.5	1.3 1.4 3.3
соотношение расстояний между метками	1: 9.16	1: 13.96	1: 7.53			1: 35.52	
порядок градаций сохранен	да	да	да	да	да	да	да
Числовые метки признака «вид»	-75.5 46.7 86.2	-70.1 43.2 80.9	0 199.6 166.9	0.7 2.8 2.5	0.7 2.8 2.5	0.3 3.1 2.6	0.5 3.0 2.5
соотношение расстояний между метками	1: 0.32	1: 0.32	1: -0.16			1: -0.19	
порядок градаций сохранен	да	да	нет	нет	нет	нет	нет
Оценка стоимости объекта C_0	1540.86	1563.12	1434.02			1489.33	
коэффициент R^2	0.982	0.983	0.988			0.980	
остаточная разность СКО s	48.26	50.73	38.79			50.60	
ошибка аппроксимации A	4.5%	4.7%	3.6%			4.2%	

Отметим, что на малых выборках, где сильна роль случайных колебаний, скорее всего, не всегда можно получить осмысленную с экономической точки зрения оптимизацию числовых меток неколичественного признака. Негативным фактором может стать парная сопряженность или мультиколлинеарность признаков. В этом случае для некоторых признаков, возможно, имеет смысл оставлять равномерную оцифровку или прибегнуть к другим методам оцифровки.

ВЫВОДЫ

1. Наибольшее прикладное значение в задачах индивидуальной оценки недвижимости имеет случай сочетания разнотипных (количественных и неколичественных) влияющих признаков, значения которых измеряются в разных

шкалах. В рассматриваемых задачах наиболее предпочтительным является приведение всех признаков к количественному типу или совместное использование количественных, квазиколичественных и небольшой совокупностей бинарных признаков.

2. Количество градаций и порядок их следования для каждого неколичественного признака должны быть согласованы с экономической гипотезой о характере влияния признака на результирующий показатель, а также с достижимой погрешностью измерения его значений, обусловленной, в основном, полнотой рыночных данных. Рекомендуемое для рассматриваемого класса задач количество градаций – 3-6.
3. Оцифровка (присвоение числовых меток) неколичественных признаков может и должна проводиться с применением оптимизационных процедур, обладающих объективными критериями и позволяющих существенно повысить точностные показатели регрессионных моделей. При этом, как и при выборе градаций, результаты оцифровки должны проверяться на соответствие экономическому характеру описываемых зависимостей.

ЛИТЕРАТУРА

1. Грибовский С.В., Сивец С.А., Левыкина И.А. Новые возможности сравнительного подхода при решении старых проблем.//Вопросы оценки. – №4, 2002. – С.22-29.
2. Анисимова И.Н. Отчет по НИР «Применение регрессионных методов в задачах индивидуальной оценки объектов недвижимости при сравнительном подходе» – СПб: СПбГИЭУ, 2003 – 133 с.
3. Грибовский С.В., Баринов Н.П., Анисимова И.Н. О повышении достоверности оценки рыночной стоимости методом сравнительного анализа // Вопросы оценки, №1, 2002, С.2-10.
4. Анисимова И.Н., Баринов Н.П., Грибовский С.В. О требованиях к количеству сопоставимых объектов при оценке недвижимости сравнительным подходом.//Вопросы оценки. - №1, 2003. – С.2-7.
5. Енюков И.С. Методы, алгоритмы, программы многомерного статистического анализа: Пакет ППСА.–М.: Финансы и статистика, 1986. – 232с.
6. Орлов А.И. Эконометрика. – М.: Изд-во «Экзамен», 2002. – 576 с.
7. Клигер С.А., Косолапов М.С., Толстова Ю.И. Шкалирование при сборе и анализе социологической информации. – М.:Наука, 1978. – 112 с.
8. Котюков В.И. Многофакторные кусочно-линейные модели. – М.: Финансы и статистика, 1984. – 216 с.
9. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс: Учеб. – М.: Дело, 2001 – 400с.
10. Катышев П.К., Я.Р.Магнус, Пересецкий А.А. Сборник задач к начальному курсу эконометрики. – М.: Дело, 2002. – 208 с.
11. Сивец С.А., Левыкина И.А. Эконометрическое моделирование в оценке недвижимости. – Запорожье: «Полиграф», 2003. – 220 с.
12. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Исследование зависимостей./Спр.изд. – М.:Финансы и статистика, 1985. – 471 с.
13. Красильников В.В. Статистика объектов нечисловой природы. – Набережные Челны, 2001.
14. Миркин Б.Г. Анализ качественных признаков и структур. – М.:Статистика, 1980. – 319 с.

15. Никифоров А.М. Обобщение модели классического регрессионного анализа на неколичественные данные. – Л.: ЛНИВЦ АН СССР, 1981.
16. Андреев Д.М. Оптимизационная модель назначения балльных оценок значениям ценообразующих факторов.//Вопросы оценки. – №3, 2002. – С.15-19.
17. Орлов А.И. Устойчивость в социально-экономических моделях. – М.:Наука, 1979. – 296 с.
18. Тюрин Ю.Н. О математических задачах в экспертных оценках.//Экспертные оценки. – М.:АН СССР, 1979. – С.7-16.
19. Отчет «Разработка методики определения уровня арендной платы за нежилые помещения в Санкт-Петербурге». – СПб.: Администрация Санкт-Петербурга, КУГИ, ГУИОН, 1997. – 62 с.
20. Эконометрика: учебник / Под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2001. – 344 с.
21. Ковалев А.П., Курова Е.В. Массовая оценка оборудования: методика и модели. Часть первая.//Вопросы оценки. – №1, 2003. – С.14-19.
22. Котюков В.И. Некоторые нестандартные статистические модели прогнозирования в эконометрии. – Новосибирск: Новосиб. институт железнодорожн. транспорта, 1977. – 15 с.

Опубликовано: Вопросы оценки, №2, 2004.- с.2-15