

For citation: Pashkovskaya O. V., Brening D. V. Attribute features application in specification of regression model of apartments cost. *Siberian Journal of Science and Technology*. 2019, Vol. 20, No. 4, P. 443–450. Doi: 10.31772/2587-6066-2019-20-4-443-450

Для цитирования: Пашковская О. В., Бренинг Д. В. Использование атрибутивных признаков при спецификации регрессионной модели стоимости квартир // Сибирский журнал науки и технологий. 2019. Т. 20, № 4. С. 443–450. Doi: 10.31772/2587-6066-2019-20-4-443-450

ATTRIBUTE FEATURES APPLICATION IN SPECIFICATION OF REGRESSION MODEL OF APARTMENTS COST

O. V. Pashkovskaya*, D. V. Brening

Reshetnev Siberian State University of Science and Technology
31, Krasnoyarsky Rabochy Av., Krasnoyarsk, 660037, Russian Federation

*E-mail: pashkovskaya@sibsau.ru

In the study of statistical data with a pronounced endogenous variable, it is necessary to identify factors (explanatory variables) that have a strong impact on the result. In this, factors can be both quantitative and attributive. To assess influence of numerical features, regression analysis methods can be used. Influence of attributive features is not taken into account. However, often these are they who make the decisive contribution to variation of the result. It is necessary to develop methods for analyzing influence of attributive features and accounting for these features in regression models.

On the example of sets of apartments proposed for sale in the city of Krasnoyarsk, a new method is used to assess influence of attributive features on the quantitative using ranking them in accordance with their influence on the endogenous variable. Method of fictitious variables is used to analyze the attribute features. Each attribute with m values is assigned $(m-1)$ dummy variables and a regression model is constructed. Influence of exogenous variables can be expressed using standardized regression coefficients. In this case, influence of attributes can be estimated by cumulative correlation coefficient calculated on the basis of a regression model with fictitious variables.

For further research, set is proposed to rank, assigning each element a "rank" – value of a standardized coefficient which reflects closeness of the relationship with the endogenous variable. Thus, all features have a numerical value. A standardized regression model is constructed.

Proposed approach can be used in the analysis of statistical aggregates, units of which are characterized by quantitative and attributive features.

Keywords: *attributive features (qualitative variables), ranking of variables, correlation, standardized multiple regression model.*

ИСПОЛЬЗОВАНИЕ АТРИБУТИВНЫХ ПРИЗНАКОВ ПРИ СПЕЦИФИКАЦИИ РЕГРЕССИОННОЙ МОДЕЛИ СТОИМОСТИ КВАРТИР

О. В. Пашковская*, Д. В. Бренинг

Сибирский государственный университет науки и технологий имени академика М. Ф. Решетнева
Российская Федерация, 660037, г. Красноярск, просп. им. газ. «Красноярский рабочий», 31

*E-mail: pashkovskaya@sibsau.ru

При изучении статистических совокупностей с явно выраженной эндогенной переменной, необходимо выявить факторные признаки, которые оказывают сильное влияние на результат. При этом факторные признаки могут быть, как количественными, так и атрибутивными. Для оценки влияния числовых признаков можно использовать методы регрессионного анализа. Влияние атрибутивных признаков не учитывается. Однако часто именно они вносят решающий вклад в вариацию результата. Необходимо разработать методы анализа влияния атрибутивных признаков и учета этих признаков в регрессионных моделях.

На примере совокупности квартир, предлагаемых к продаже в г. Красноярске, применен новый метод оценки влияния атрибутивных признаков на количественные с использованием ранжирования их в соответствии с их влиянием на эндогенную переменную. Для анализа атрибутивного признака используется метод фиктивных переменных. Каждому атрибуту, имеющему t значений, ставится в соответствие $t-1$ фиктивная переменная, строится регрессионная модель. Силу влияния экзогенных переменных можно выразить с помощью стандартизованных коэффициентов регрессии. В этом случае влияние атрибутов можно оценить

совокупным коэффициентом корреляции, вычисленным на основе регрессионной модели с фиктивными переменными.

Для дальнейшего исследования совокупность предлагается ранжировать, присваивая каждому элементу «ранг» – значение стандартизированного коэффициента, который отражает тесноту связи с эндогенной переменной. Таким образом, все признаки имеют числовое значение. Построены стандартизированные регрессионные модели.

Предлагаемый подход можно использовать при анализе статистических совокупностей, единицы которых характеризуются количественными и атрибутивными признаками.

Ключевые слова: атрибутивные признаки, ранжирование переменных, корреляция, стандартизированная модель множественной регрессии.

Introduction. To assess influence of factor variables on a endogenous variable, various econometric tools are used [1–14]. For example, in the work, published earlier [1], cost of an apartment in aggregate of apartments offered for sale in the housing market in the city of Krasnoyarsk in 2017, a regression model is being constructed, in which quantitative explanatory variables (total area, residential area and others) are used. As a result, some characteristics were obtained, according to which the degree of influence of any factor [1–14] was estimated. When studying the aggregate of apartments, offered for sale on the primary housing market in the city of Krasnoyarsk in 2017 [15], it was revealed that a strong influence on the price of an apartment, in addition to quantitative, such characteristics as total area of the apartment, qualitative variables: district of the city, in which the house is located and material from which this house is built, also exert influence.

If the statistical analysis of numerical features does not cause any difficulties, but when dealing with qualitative variable, various difficulties arise: meaning of attribute is represented by the word - attribute, data needs ranking, substituting each meaning by rank. If, when ranking, values are replaced simply by serial numbers, then the numerical value will distort influence on the resultant attribute. For example, when ranking on the basis of “district”, apartments in Akademgorodok are given rank 12, and apartments in Pokrovka are ranked 6. It turns out that apartments in the same region are twice “bigger” than in the other. The question arises, why exactly twice, and is it really “bigger” or is it just an unconfirmed guess of the researcher?

For the studied aggregate [5], it is proposed to apply a new method for assessing influence of qualitative variable on the endogenous variable. We distinguish two attribute signs: wall materials and region in which the house is built. We will develop this method for a set of one-room apartments, and then test it on a set of 2, 3, 4-room apartments.

Method of fictitious variables. When studying influence of attributive features, method of fictitious variables will be used.

Let us evaluate influence of the district (z), in which the house is built, on formation of the price of apartment (y) (structure of the studied phenomenon is shown in fig. 1).

To account for the qualitative variable (z), we introduce twelve fictitious variables $z_1, z_2, \dots, z_{12}$ into the regression model. Based on available sample of one-room apartments (volume of 765 elements), we calculate esti-

mates of parameters of regression equation. Regression equation will be built in a standardized form:

$$\hat{t}_y = -0.01t_{z_1} + 0.25t_{z_2} - 0.08t_{z_3} - 0.01t_{z_4} - 0.05t_{z_5} - \\ - 0.30t_{z_6} - 0.18t_{z_7} + 0.04t_{z_8} - 0.14t_{z_9} - 0.44t_{z_{10}} - \\ - 0.04t_{z_{11}} + 0.09t_{z_{12}} + 0.60t_x \quad (1)$$

where t_y is the price of apartment (standardized variable

corresponding to the value of the apartment), $t_y = \frac{y - \bar{y}}{s_y}$,

$t_{z_1}, t_{z_2}, \dots, t_{z_{12}}$ standardized fictitious variables, corresponding to attributive feature “district”, t_x – total area (standardized variable); for evaluating each parameter, the observed value of t-statistics is given below.

In equation (1) adjusted coefficient of determination was 82.6 %, what indicates high quality and statistical significance of constructed equation.

To assess influence of qualitative variable (z) – “district”, basing on the constructed correlation matrix, we calculate total correlation coefficient:

$$r_{y|z_1, z_2, \dots, z_{12}} = \beta_1 r_{yz_1} + \dots + \beta_{12} r_{yz_{12}} = \\ = (-0.012) \cdot (-0.029) + \dots + 0.092 \cdot 0.268 = 0.404.$$

Share of influence of quantitative feature in the explained variation was 51.2 %, and of qualitative variable – 48.7 %.

Next, we evaluate influence of material of the house walls (u) on formation of apartment price (y) (structure of the phenomenon is shown in fig. 2).

To account for the qualitative variable (u), we introduce two fictitious variables u_1, u_2 into the regression model. Let's build a standardized regression model:

$$\hat{t}_y = -0.11t_{u_1} + 0.05t_{u_2} + 0.71t_x \quad (2)$$

For evaluation influence of attribute (u) – “wall material”, based on constructed correlation matrix

$$Q = \begin{pmatrix} 1 & -0.088 & 0.174 & 0.709 \\ -0.088 & 1 & -0.198 & 0.045 \\ 0.174 & -0.198 & 1 & 0.145 \\ 0.709 & 0.045 & 0.145 & 1 \end{pmatrix} \quad \text{let's calculate}$$

total correlation coefficient:

$$r_{y|u_1, u_2} = \beta_1 r_{yu_1} + \beta_2 r_{yu_2} = \\ = (-0.11) \cdot (-0.088) + (0.05) \cdot (0.174) = 0.018.$$

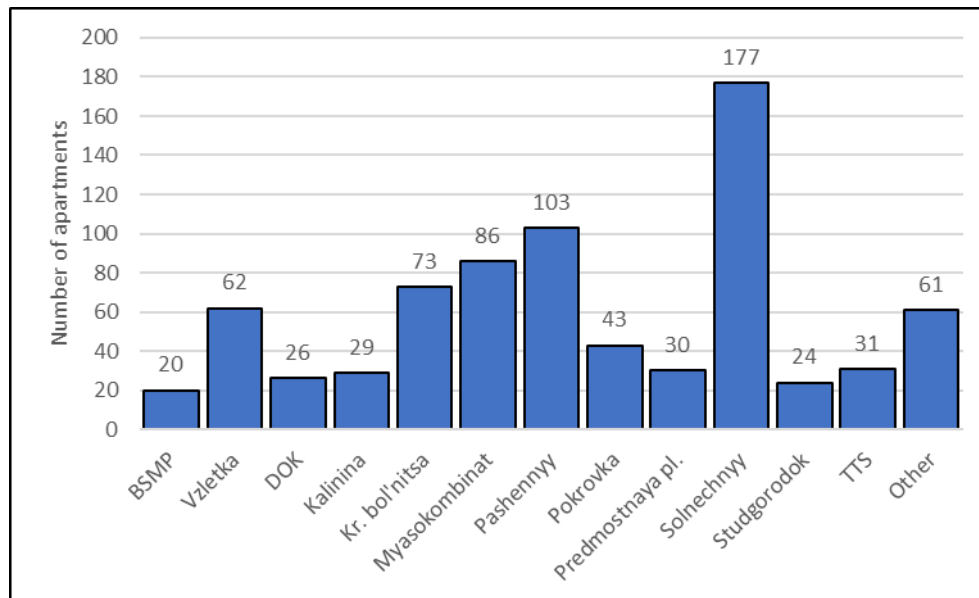


Fig. 1. Distribution of offered for sale one-room apartments by district of the city Krasnoyarsk, 2017 year

Рис. 1. Распределение предлагаемых к продаже однокомнатных квартир по районам, г. Красноярск, 2017 г.

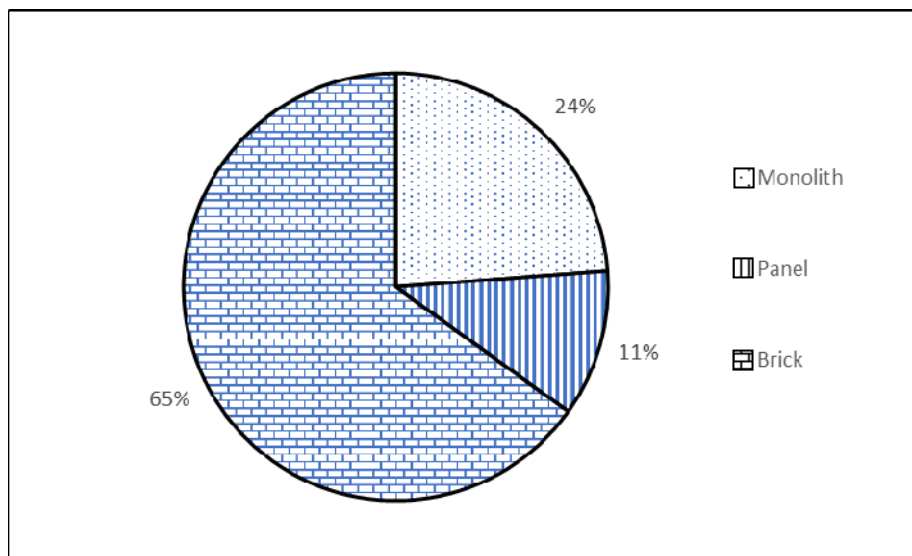


Fig. 2. Distribution of offered for sale one-room apartments by wall material, Krasnoyarsk, 2017 year

Рис. 2. Распределение предлагаемых к продаже однокомнатных квартир по материалу стен, г. Красноярск, 2017 г.

Ranking of aggregate. In the initial population, we replace values of attribute features with corresponding standardized regression coefficients, thus assigning each element a “rank”.

Now, in aggregate under consideration, all the features are quantitative. We get the following regression equation:

$$\hat{t}_y = \frac{0.63}{39.92} t_x + \frac{0.09}{5.93} t_u + \frac{0.55}{34.95} t_z. \quad (3)$$

Equation (3) by 81.6 % explains change in the price of apartments (adjusted coefficient of determination is 0.815) and with probability of 99 % is statistically significant, standard mistake is $S = 0.43$.

Approbation. In analogy, we analyze aggregate of two-room apartments (547 units of observation). In this aggregate, average apartment area is 55.13 square meters. m, average price of 1 square. m – 48.41 thousand rubles,

average cost of an apartment – 2668.93 thousand rubles. Structure of aggregation is shown in fig. 3–4.

In this population, there is a very strong influence of qualitative variable – part of influence of the qualitative variable in the overall determination coefficient is – 49.12 %. The resulting equation is:

$$\hat{t}_y = 0.58 t_x - 0.02 t_u + 0.57 t_z.$$

$\frac{30.47}{0.58}$
 $\frac{-1.24}{-0.02}$
 $\frac{29.31}{0.57}$

Statistically significant with probability of 99 % (adjusted coefficient of determination is equal to 0.818).

Next, we analyze sample of three-room apartments (200 units of observation). In this aggregate, average apartment area is 76.62 square meters, average price of 1 square. m – 51.05 thousand rubles, average cost of an apartment is – 3910.94 thousand rubles. The structure of aggregate is shown in fig. 5, 6.

Replacing attribute feature values with standardized coefficients

№	Cost of apartment, thousand rubles	Area of apartment, apt. m	Wall material		District	
			Attributes	Rank	Attributes	Rank
1	1863	48.27	Monolith	–0.11	Myasokombinat	–0.302
2	1262	31.77	Monolith	–0.11	Myasokombinat	–0.302
3	1010	26.01	Monolith	–0.11	Myasokombinat	–0.302
4	1010	26.01	Monolith	–0.11	Myasokombinat	–0.302
5	1646	42.20	Monolith	–0.11	Myasokombinat	–0.302
6	1646	42.20	Monolith	–0.11	Myasokombinat	–0.302
7	1545	42.70	Monolith	–0.11	Solnechnyy	–0.435
...	...	...	...	...	...	...
687	1677	36.11	Brick	0.00	Solnechnyy	–0.435
688	1531	37.91	Brick	0.00	Myasokombinat	–0.302
689	2315	42.09	Panel	0.05	Innokentyevskiy	0.000
...	...	...	...	...	...	...
763	3162	51.00	Brick	0.00		0.242
764	1100	17.60	Brick	0.00	BSMP	–0.012
765	3162	51.00	Brick	0.00	Vzletka	0.242

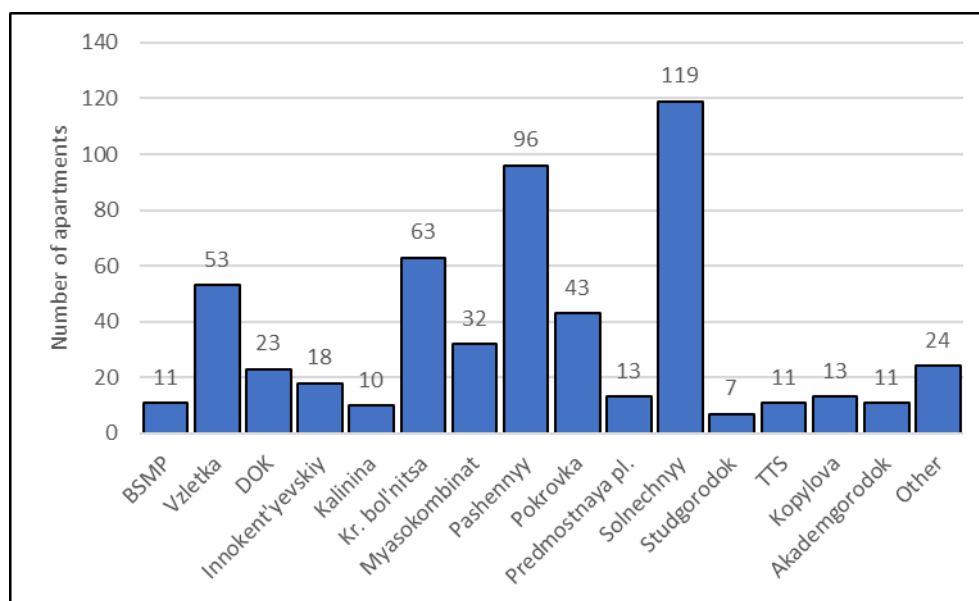


Fig. 3. Distribution of offered for sale two-room apartments by district of the city Krasnoyarsk, 2017 year

Рис. 3. Распределение предлагаемых к продаже двухкомнатных квартир по районам, г. Красноярск, 2017 г.

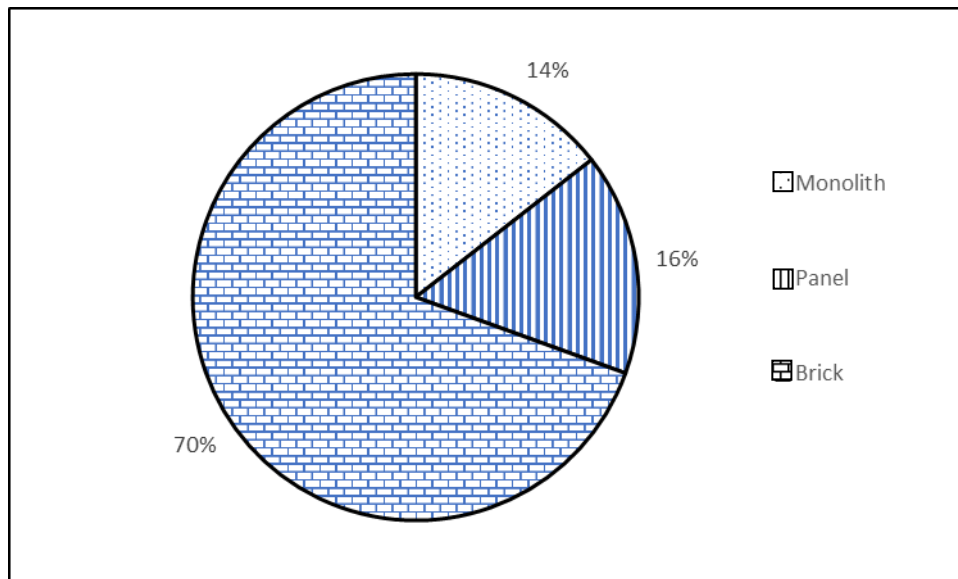


Fig. 4. Distribution of offered for sale two-room apartments by wall material, Krasnoyarsk, 2017 year

Рис. 4. Распределение предлагаемых к продаже двухкомнатных квартир по материалу стен, г. Красноярск, 2017 г.

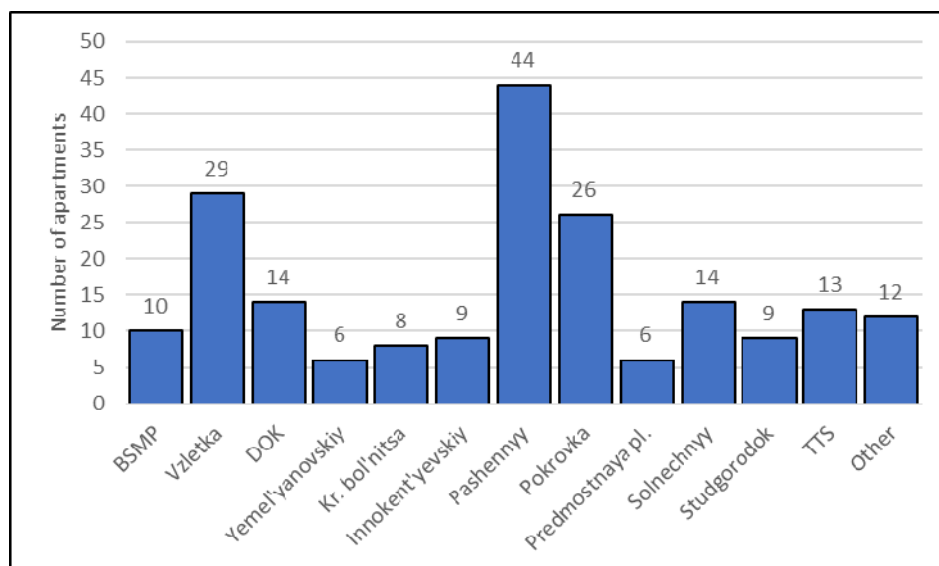


Fig. 5. Distribution of offered for sale three-room apartments by district of the city Krasnoyarsk, 2017 year

Рис. 5. Распределение предлагаемых к продаже трехкомнатных квартир по районам, г. Красноярск, 2017 г.

Based on the available quantitative characteristics (y – is price of apartment, thousand rubles; x_1 – is the total area, sq. m; x_2 – is living area, sq. m; x_3 – is the kitchen area, sq. m) we construct a regression model in a standardized form:

$$\hat{t}_y = 0.78t_{x_1} + 0.06t_{x_2} + 0.05t_{x_3}.$$

12.19 1.28 0.86

For these models: adjusted coefficient of determination is 71.4 % and regression equation is statistically significant with a probability of 99 %. Low values

of t -statistics indicate that estimates of regression parameters are not statistically significant, and high values of correlation coefficients between factors ($r_{x_1x_2} = 0.549$; $r_{x_1x_3} = 0.648$) signal about presence of multicollinearity in the model. Let's try to improve quality of the model by adding attributive features and excluding factors (x_2 – living area, sq. m; x_3 – kitchen area, sq. m) closely associated with the sign x_1 – total area, sq. m

Dealing with qualitative variables wall material (u) and district (z) and quantitative feature total area (x) and

using method, presented in this article earlier, we obtain standardized equation:

$$\hat{t}_y = 0.68t_x + 0.01t_u + 0.42t_z.$$

$\frac{24.31}{0.39}$
 $\frac{14.31}{14.31}$

Note, that correlation matrix does not show close relationship between factors: $r_{xu} = 0.272$; $r_{xz} = 0.379$; $r_{zu} = 0.390$, therefore, there is no need to eliminate multicollinearity.

Share of influence of attribute feature in the overall coefficient of determination is 33.54 %. Such equation is

statistically significant with probability of 99 % and adjusted R -square is 0.868.

Finally, we analyze aggregate of four-room apartments (83 units of observation).

In this aggregate, average apartment area is 107.82 square meters. m, average price of 1 square. m – 52.79 thousand rubles, average cost of an apartment – 5692.53 thousand rubles. The structure of population is shown in fig. 7, 8. Aggregate is heterogeneous (coefficient of variation is 41.2 %), apartments in brick houses predominate.

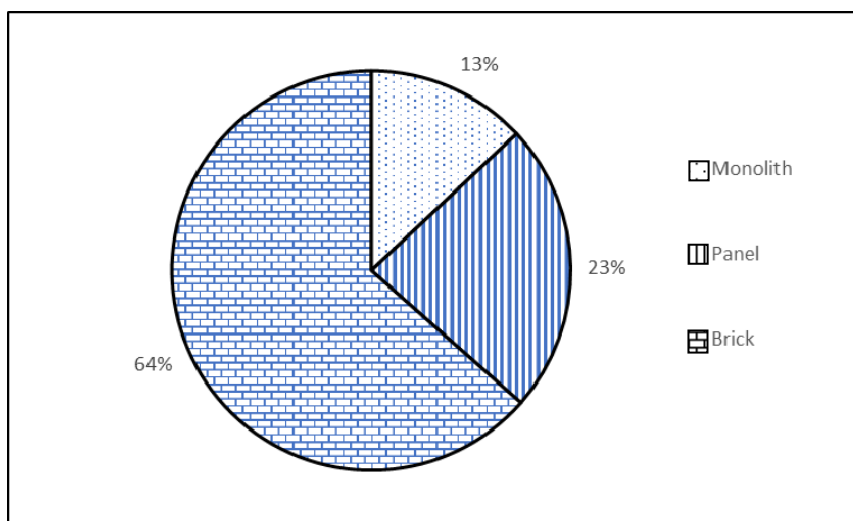


Fig. 6. Distribution of offered for sale three-room apartments by wall material, Krasnoyarsk, 2017 year

Рис. 6. Распределение предлагаемых к продаже трехкомнатных квартир по материалу стен, г. Красноярск, 2017 г.

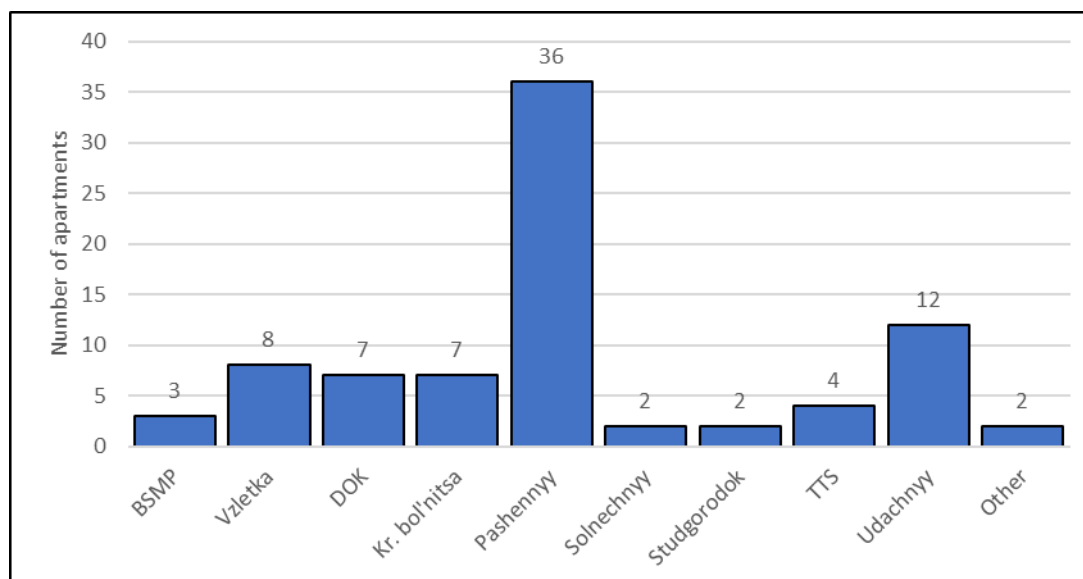


Fig. 7. Distribution of offered for sale four-room apartments by district of the city Krasnoyarsk, 2017 year

Рис. 7. Распределение предлагаемых к продаже четырехкомнатных квартир по районам, г. Красноярск, 2017 г.

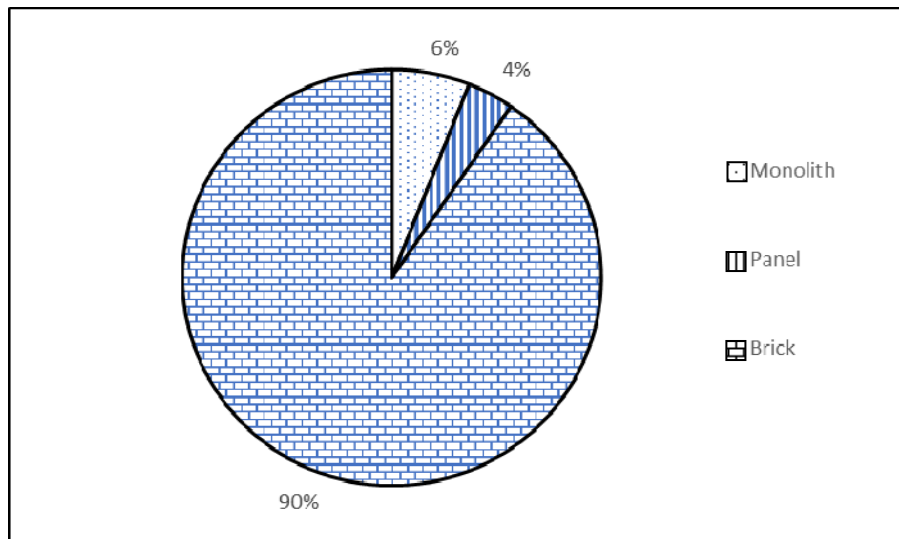


Fig. 8. Distribution of offered for sale four-room apartments by wall material, Krasnoyarsk, 2017 year

Рис. 8. Распределение предлагаемых к продаже четырехкомнатных квартир по материалу стен, г. Красноярск, 2017 г.

In a model containing only quantitative attributes:

$$\hat{t}_y = \frac{1.11}{18.11} t_{x_1} - \frac{0.17}{-2.74} t_{x_2},$$

adjusted coefficient of determination is 0.930 and regression equation is statistically significant with a probability of 99 %.

Dealing with attributive features, we obtain equation

$$\hat{t}_y = \frac{0.96}{29.13} t_x + \frac{0.04}{1.38} t_u + \frac{0.001}{0.02} t_z.$$

Conclusion. Thus, the proposed method allows to analyze statistical data, elements of which are characterized by both numerical and attribute features (qualitative variable). This significantly improves quality of models and takes into account large amount of factors. This method can be easily applied to arbitrary aggregates.

References

1. Brening D. V., Pashkovskaya O. V., Senashov S. I., Savostyanova I. L. [Mathematical modeling of the cost of secondary housing in the city of krasnoyarsk in 2017]. *Sibirskiy zhurnal nauki i tekhnologii*. 2018, Vol. 19, No. 4, P. 581–588 (In Russ.).
2. Senashov S. I., Yuferova N. Yu., Vaytekunene E. L. *Ekonometricheskoe modelirovanie stoimosti zhil'ya v Krasnoyarske* [Econometric modelling of residential property cost in Krasnoyarsk]. Krasnoyarsk, SibGAU Publ., 2015, 178 p. (In Russ.).
3. Pashkovskaya O. V., Senashov S. I., Savostyanova I. L., Yuferova N. Yu. [Modeling the cost of housing on the secondary real estate market in the city of Krasnoyarsk in 2016]. *Sibirskiy zhurnal nauki i tekhnologii*. 2017, No. 4 (18), P. 788–796 (In Russ.).
4. Senashov S. I. et al. [Actual modeling of real estate in Krasnoyarsk]. *Vestnik SibGAU*. 2013, No. 2 (48), P. 86–91 (In Russ.).
5. Senashov S. I., Yuferova N. Yu., Groshak E. V. *Modelirovaniye stoimosti zhil'ya v g. Krasnoyarske* [Simulation of the cost of residential property in the city of Krasnoyarsk]. Krasnoyarsk, SibGTU Publ., 2007, 204 p.
6. Senashov S. I., Juferova N. Yu., Surnina E. V. [Information system of valuation of apartments on the secondary market as a tool for investment management]. *Vestnik SibGAU*. 2009, No. 4 (25), P. 219–223 (In Russ.).
7. Senashov S. I. et al. [Cost estimation of information system of apartments at secondary housing markets as a management investment tool]. *Vestnik SibGAU*. 2009, No. 5, P. 154–157 (In Russ.).
8. Denisenko I. K., Senashov S. I., Juferova N. Yu. [Modeling the cost of housing in Krasnoyarsk] *Vestnik Sib. gos. tehnologich. un-ta*. 2000, No. 1, P. 185–190 (In Russ.).
9. Senashov S. I., Savostyanova I. L., Yuzaeva A. G., Savchenko L. M. [Simulation of the cost of residential properties in Krasnoyarsk for 2013–2014]. *Vestnik SibGAU*. 2016, Vol. 17, No. 3, P. 830–836 (In Russ.).
10. Senashov S. I., Surnina E. V., Puzanova G. A., Filyushina E. V. [Methods of valuation of land]. *Reshetnevskie chteniya*. 2009, Vol. 2, No. 13, P. 549–650 (In Russ.).
11. Surnina E. V., Senashov S. I. [The use of GIS technology for estimation of cost of apartments]. *Aktualnye problemy aviatsii i kosmonavтики*. 2010, Vol. 1, No. 6, P. 444–445 (In Russ.).
12. Senashov S. I., Savostyanova L. I. [Econometric modeling of the cost of housing]. *Lesnoy i khimicheskij kompleksy – problemy i resheniya*. 2016, Vol. 2, P. 235–236 (In Russ.).
13. Savchenko L. M., Yusaeva A. G., Senashov S. I. [Modeling the cost of housing in the city of Krasnoyarsk in 2013]. *Prospekt Svobodnyy-2016*. 2016, P. 67–70 (In Russ.).

14. Rusakova A. A., Starodubtsev A. A., Senashov S. I. [Modeling the cost of housing in the Central district of Krasnoyarsk]. *Aktualnye problemy aviatsii i kosmonavтики*. 2017, Vol. 2, P. 398–401 (In Russ.).

15. Brening D. V., Savostyanova I. L., Kvartiry Krasnoyarska 2017. Rynok pervichnogo zhil'ya [Flats of Krasnoyarsk 2017. Primary real estate market]. Svidetel'stvo o gosudarstvennoy registratsii bazy dannykh № 20186621065.

Библиографические ссылки

1. Математическое моделирование стоимости вторичного жилья в городе Красноярске в 2017 г. / Д. В. Бренинг, О. В. Пашковская, С. И. Сенашов, И. Л. Савостьянова // Сибирский журнал науки и технологий. 2018. № 4 (19). С. 581–588.

2. Сенашов С. И., Юферова Н. Ю., Вайтекунене Е. Л. Эконометрическое моделирование стоимости жилья в Красноярске. Красноярск : Сиб. гос. аэрокосмич. ун-т., 2015. 178 с.

3. Моделирование стоимости жилья на вторичном рынке недвижимости в г. Красноярске в 2016 г. / О. В. Пашковская, С. И. Сенашов, И. Л. Савостьянова, Н. Ю. Юферова // Сибирский журнал науки и технологий. 2017. № 4 (18). С. 788–796.

4. Актуальное моделирование недвижимости в Красноярске / С. И. Сенашов и др. // Вестник СибГАУ. 2013. № 2 (48). С. 86–91.

5. Сенашов С. И., Юферова Н. Ю., Грошак Е. В. Моделирование стоимости жилья в г. Красноярске : монография / СибГТУ. Красноярск, 2007. 204 с.

6. Сенашов С. И., Юферова Н. Ю., Сурнина Е. В. Информационная система оценки стоимости квартир на вторичном рынке жилья как инструмент управления инвестициями // Вестник СибГАУ. 2009. № 4(25). С. 219–223.

7. Cost estimation of information system of apartments at secondary housing markets as a management

investment tool / С. И. Сенашов и др. // Вестник СибГАУ. 2009. № 5(26). Р. 154–157.

8. Денисенко И. К., Сенашов С. И., Юферова Н. Ю. Моделирование стоимости жилья в Красноярске // Вестник Сиб. гос. технологич. ун-та. 2000. № 1. С. 185–190.

9. Simulation of the cost of residential properties in Krasnoyarsk for 2013–2014 / С. И. Сенашов и др. // Вестник СибГАУ. 2016. Т. 17, № 3. С. 830–836.

10. Методы оценки земельных участков / С. И. Сенашов и др. // Решетневские чтения. 2009. Т. 2, № 13. С. 549–650.

11. Сурнина Е. В., Сенашов С. И. Использование ГИС-технологий для оценки стоимости квартир // Актуальные проблемы авиации и космонавтики. 2010. Т. 1, № 6. С. 444–445.

12. Сенашов С. И., Савостьянова И. Л. Эконометрическое моделирование стоимости жилья // Лесной и химический комплексы – проблемы и решения : материалы Всерос. науч.-практ. конф. Красноярск, 2016. Т. 2. С. 235–236.

13. Савченко Л. М., Юзаева А. Г., Сенашов С. И. Моделирование стоимости жилья в г. Красноярске за 2013 год // Проспект Свободный-2016 : сб. материалов Междунар. конф. студентов, аспирантов и молодых ученых. 2016. С. 67–70.

14. Русакова А. А., Стародубцев А. А., Сенашов С. И. Моделирование стоимости жилья Центрального района города Красноярска // Актуальные проблемы авиации и космонавтики : сб. материалов III Междунар. науч.-практ. конф. 2017. Т. 2. С. 398–401.

15. Квартиры Красноярска 2017. Рынок первичного жилья: свид. 20186621065 / Д. В. Бренинг, И. Л. Савостьянова и др. ; заявитель и правообладатель ФГБОУ ВО СибГУ им. М.Ф. Решетнева (RU) №20186621065. Дата регистр. 12.07.2018. Реестр баз данных. 1 с.

© Pashkovskaya O. V., Brening D. V., 2019

Pashkovskaya Olga Vladimirovna – Cand. Sc., Docent, Department of Information Economic Systems; Reshetnev Siberian State University of Science and Technology. E-mail: pashkovskaya@sibsau.ru.

Brening Darya Vladimirovna – student; Reshetnev Siberian State University of Science and Technology. E-mail: brening98@gmail.com.

Пашковская Ольга Владимировна – кандидат физико-математических наук, доцент кафедры информационных экономических систем, Сибирский государственный университет науки и технологий имени академика М. Ф. Решетнева. E-mail: pashkovskaya@sibsau.ru.

Бренинг Дарья Владимировна – студент группы БПЭ 16-01, Сибирский государственный университет науки и технологий имени академика М. Ф. Решетнева. E-mail: brening98@gmail.com.