

Elżbieta Jasińska\*, Edward Preweda\*

THE APPLICATION OF FACTORIAL ANALYSIS  
IN THE CLASSIFICATION OF THE ATTRIBUTES OF REAL ESTATES\*\*

---

## 1. Methods of selecting the factors

The base for differentiating factors comprises proper values determined based on the correlation matrix between the variables, i.e. variances differentiated by the subsequent factors. Based on this measure it is possible to establish how many factors are important from the practical point of view. The most commonly used method is Kaiser criterion. This criterion leaves only these factors that have their proper values bigger than 1. This means that if the factor does not differentiate at least as much as one original variable – this factor is rejected. The second method applied in the paper is a graphic talus test, where proper values are presented on a linear graph and then the place is defined in such a way that to the right from this place there is a mild fall of proper values. To the right from this point there is probably only a factorial talus.

The application of the Kaiser criterion makes that, in some cases, there are too many factors left, while the talus test usually leaves too few factors. Both methods give similar results when there is a high number of observations and small number of factors. In the paper both methods were applied. To verify if the differentiated number of factors was sufficient, the matrix of residual correlations was made. They enable us to find possible lacks in adjustment; it means to observe which correlation coefficients cannot be properly reconstructed by a given number of factors.

## 2. Examples with numbers

Based on transaction data, collected for land real estates planned for small apartment buildings in Krakow, the factorial analysis was made, considering the territorial division of the City into districts. For the considered real estates the obligatory and facultative attributes were established using the information from the notary acts.

---

\* AGH University of Science and Technology, Faculty of Mining Surveying and Environmental Engineering

\*\* The work was done within the Statute Grant in the Chair of the Terrain Information (AGH-UST)

There were six attributes:

- 1) Outline (the outline within the borders in a given district);
- 2) Time (counted in the number of months from the first transaction);
- 3) Limitations (loading with mortgage, services etc.);
- 4) Surface (the surface of the real estate counted in m<sup>2</sup>);
- 5) Shape (the shape of the site: favourable or unfavourable);
- 6) Sum of facilities (the sum of the following facilities: telephone, electricity, water supply, gas supply, sewerage and central heating).

Among the seventeen districts analysed – in this paper four districts were selected, because their results were the most characteristic.

## 2.1. Bronowice

In the analysis the defined earlier attributes able to affect the variability of prices of real estates were chosen. The results of the calculations were presented in Tables 1, 2 and Figure 1.

**Table 1.** Proper values for the Bronowice district

No.	Proper values	Total variance [%]	Cumulated proper values	Cumulated proper values [%]
1	2.21	36.9	2.21	36.9
2	1.24	20.6	3.45	57.5
3	0.93	15.5	4.38	73.0
4	0.81	13.5	5.19	86.5
5	0.42	7.0	5.61	93.5
6	0.39	6.5	6.00	100.0

**Table 2.** Factorial loads for the Bronowice district

Attribute	Factor 1	Factor 2
Outline	-0.34	-0.55
Time	-0.10	0.89
Limitations	-0.30	-0.50
Surface	0.57	0.30
Shape	0.87	-0.09
Facilities	0.83	0.08
Explained variance	1.99	1.46
Fraction	0.33	0.24

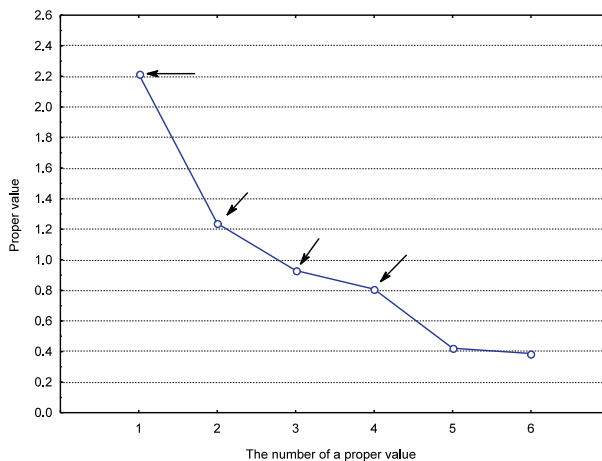


Fig. 1. The graph of the proper values for the Bronowice district

In the analysed district the Kaiser criterion indicates 2 factors, while the talus test does not comply with this criterion, indicating 4 factors. Based on detail analysis of the influence of the number of factors on the formation of the residual correlation matrix one can state that in this case only two factors are significant. They show that price-making attributes are shape, sum of facilities and time.

In Figure 2 two attributes making one factor were marked. Although the attributes "limitations" and "outline" also make a cluster, the low absolute value of factorial loads for these attributes allows the conclusion that they are not significant in examining the variability in prices of real estates. Moreover, these attributes are strongly correlated with each other.

Residual correlation matrix (Tab. 3) and the graph of factorial loads (Fig. 2) confirm the correct selection of the number of factors.

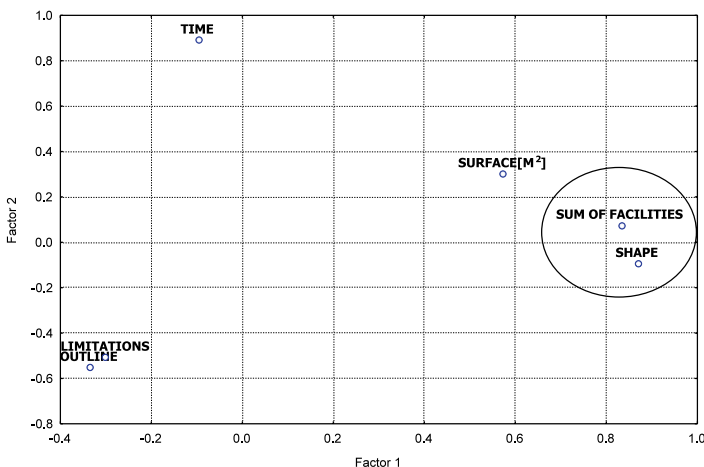


Fig. 2. The graph of factorial loads for the Bronowice district

**Table 3.** Residual correlation matrix for the Bronowice district

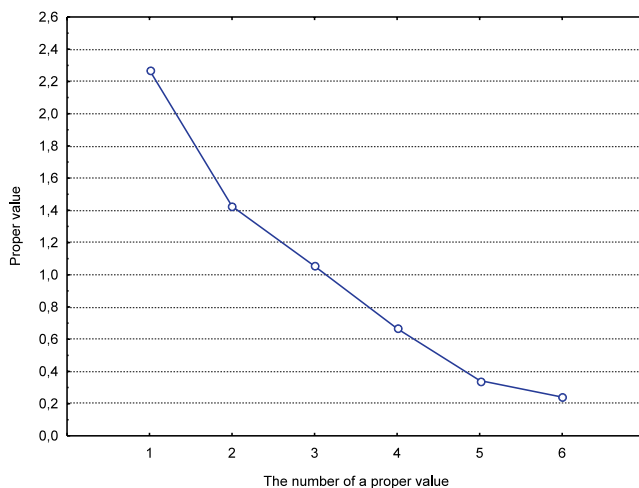
	Outline	Time	Limitations	Surface	Shape	Facilities
Outline	0.59	0.17	-0.20	0.19	-0.03	0.09
Time	0.17	0.19	0.18	0.06	0.08	0.03
Limitations	-0.20	0.18	0.65	0.24	0.07	-0.05
Surface	0.19	0.06	0.24	0.58	-0.08	-0.15
Shape	-0.03	0.08	0.07	-0.08	0.24	-0.17
Facilities	0.09	0.03	-0.05	-0.15	-0.17	0.30

## 2.2. Grębałów

In the case of the Grębałów district the studied attributes have different values, thus it would be groundless to reject any of them. As the result of the factorial analysis the results shown in Tables 4, 5 and Figure 3 were obtained

**Table 4.** Proper values for the Grębałów district

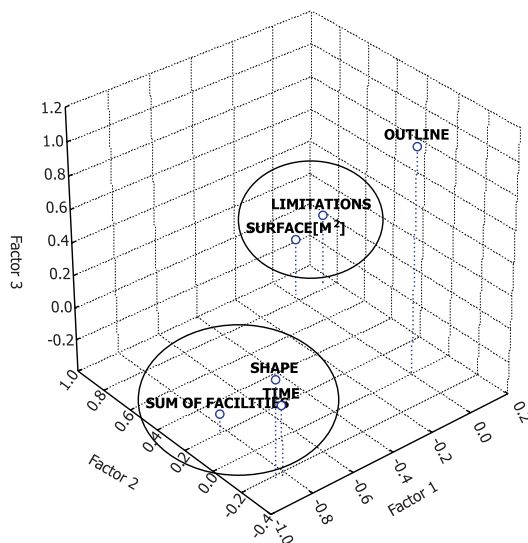
No.	Proper values	Total variance [%]	Cumulated proper values	Cumulated proper values [%]
1	2.27	37.85	2.27	37.85
2	1.43	23.76	3.70	61.60
3	1.05	17.56	4.75	79.16
4	0.67	11.10	5.42	90.26
5	0.34	5.72	5.76	95.98
6	0.24	4.02	6.00	100.00

**Fig. 3.** The graph of proper values for the Grębałów district

The graph of proper values does not allow establishing an unambiguous criterion for the selection of the number of factors. Figure 3 indicates that one should take six factors, but then there is no possibility to distinguish attributes influencing and not influencing the price. The Kaiser criterion (the stricter one) suggests the selection of three factors, which are presented in Table 5 and Figure 4.

**Table 5.** Factorial loads for the Grębałów district

Attribute	Factor 1	Factor 2	Factor 3
Outline	0.00	0.01	0.98
Time	-0.82	-0.22	0.03
Limitations	0.15	0.84	0.04
Surface	0.01	0.85	-0.04
Shape	-0.85	-0.21	0.20
Facilities	-0.84	0.19	-0.27
Explained variance	2.11	1.55	1.09
Fraction	0.35	0.26	0.18



**Fig. 4.** The tri-dimensional graph of factorial loads for the Grębałów district

In Figure 4 one can clearly see two groups of attributes able to influence the price. The first factor consists of the following attributes: time, shape and sum of facilities, second factor comprises limitations and surface. In this case the outline (factor 3) has also some importance. The attempt to limit the number of first two factors to much extent deteriorates the unexplained variance. In Table 6 the residual correlation matrix for the analysed district was presented.

**Table 6.** Residual correlation matrix for the Grębałów district

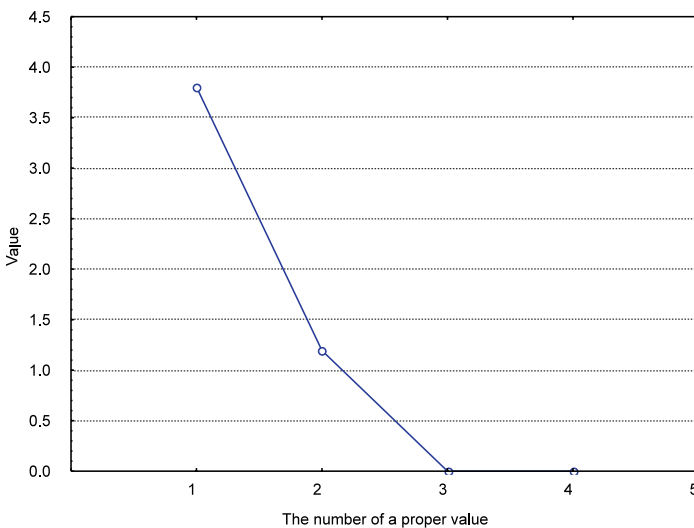
	Outline	Time	Limitations	Surface	Shape	Facilities
Outline	0.03	-0.02	-0.04	0.01	-0.06	0.06
Time	-0.02	0.28	-0.05	0.12	-0.14	-0.14
Limitations	-0.04	-0.05	0.27	-0.25	0.10	-0.01
Surface	0.01	0.12	-0.25	0.28	-0.08	-0.09
Shape	-0.06	-0.14	0.10	-0.08	0.20	-0.05
Facilities	0.06	-0.14	-0.01	-0.09	-0.05	0.19

### 2.3. Nowa Huta

Five attributes were taken for the analysis, because the attribute „limitations” has the constant value for all the real estates from the database. Based from the correlation matrix only two proper values can be determined, and based on them-two factorial loads, presented in Tables 7, 8 and Figures 5, 6.

**Table 7.** Proper values for the Nowa Huta district

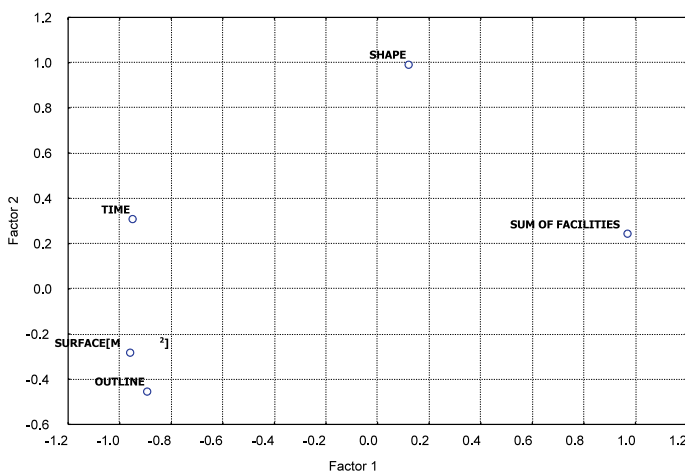
No.	Proper values	Total variance [%]	Cumulated proper values	Cumulated proper values [%]
1	3.81	76.11	3.81	76.11
2	1.19	23.89	5.00	100.00



**Fig. 5.** The graph of proper values for the Nowa Huta district

**Table 8.** Factorial loads for the Nowa Huta district

Attribute	Factor 1	Factor 2
Outline	-0.89	-0.45
Time	-0.95	0.31
Surface	-0.96	-0.28
Shape	0.12	0.99
Facilities	0.97	0.24
Explained variance	3.58	1.42
Fraction	0.72	0.28



**Fig. 6.** The graph of factorial loads for the Nowa Huta district

After selecting two factors a zero residual correlation matrix was obtained (Tab. 9). Thus it is groundless to reject any of the attributes. This criterion should be accepted, the removal of any factor significantly increases the unexplained variance (e.g. up to 27% – if attribute “shape” is rejected). Presented above calculations show that in this case all the analysed attributes are statistically significant.

**Table 9.** Residual correlation matrix for the Nowa Huta district

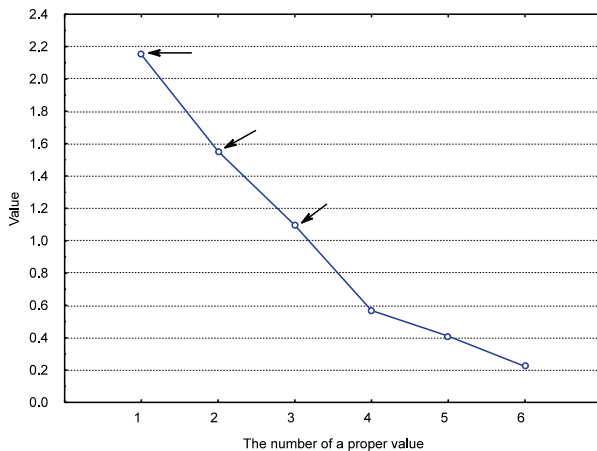
	Outline	Time	Surface	Shape	Facilities
Outline	0.00	0.00	0.00	0.00	0.00
Time	0.00	0.00	0.00	0.00	0.00
Surface	0.00	0.00	0.00	0.00	0.00
Shape	0.00	0.00	0.00	0.00	0.00
Facilities	0.00	0.00	0.00	0.00	0.00

## 2.4. Podgórze

The talus test (Fig. 7) shows the decreased slope of the graph after fourth proper value, while the Kaiser criterion suggests taking three factors (Tab. 10). After the analysis of factorial loads (Tab. 11), the Kaiser criterion was accepted as the correct one.

**Table 10.** Proper values for the Podgórze district

No.	Proper values	Total variance [%]	Cumulated proper values	Cumulated proper values [%]
1	2.15	35.84	2.15	35.84
2	1.55	25.88	3.70	61.72
3	1.10	18.25	4.80	79.98
4	0.57	9.48	5.37	89.46
5	0.41	6.84	5.78	96.29
6	0.22	3.71	6.00	100.00



**Fig. 7.** The graph of proper values for the Podgórze district

**Table 11.** Factorial loads for the Podgórze district

Attribute	Factor 1	Factor 2	Factor 3
Outline	0.01	0.07	<b>0.96</b>
Time	0.04	<b>0.88</b>	0.00
Limitations	<b>0.73</b>	0.26	-0.43
Surface	<b>-0.88</b>	0.02	-0.09
Shape	0.13	<b>-0.80</b>	-0.12
Facilities	<b>0.76</b>	-0.53	0.09
Explained variance	1.90	1.77	1.13
Fraction	0.32	0.29	0.19



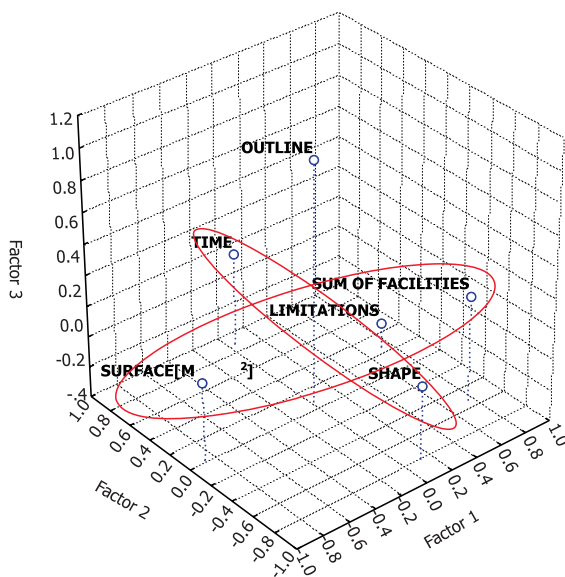


Fig. 8. The tri-dimensional graph of factorial loads for the Podgórze district

The presented in Figure 8 graph of factorial loads illustrates the principle of mutual orthogonality of the factors. The carried out analysis showed that all the attributes taken for the analysis influence the prices of real estates. From the presented in Table 12 residual correlation matrix one can state that maximal unexplained correlation is, due to this solution 5%.

Table 12. Residual correlation matrix for the Podgórze district

	Outline	Time	Limitations	Surface	Shape	Facilities
Outline	0.08	0.01	0.12	0.09	0.08	-0.03
Time	0.01	0.22	-0.03	0.03	0.23	0.02
Limitations	0.12	-0.03	0.22	0.17	0.06	-0.03
Surface	0.09	0.03	0.17	0.22	0.04	0.08
Shape	0.08	0.23	0.06	0.04	0.33	-0.08
Facilities	-0.03	0.02	-0.03	0.08	-0.08	0.14

### 3. Conclusions

Based on the set of results for the analysed districts, one can make the conclusions referring to the reasonability of the application and the course of the factorial analysis in the estimation of the values of real estates:

- Selecting the number of factors is a subjective decision and the proposed criteria make only hints. The results of carried out analyses indicate that only one criterion – Kaiser criterion should be applied.

- The analysis of the residual correlation matrix allows the verification of the correctness of the solution, i.e. accepting the proper number of factors.
- Final solution should be established based on several options carried out.
- In some cases, such as the district of Nowa Huta, the elimination of one factor significantly increases the explained variances.
- The application of factorial analysis allows the choice among many attributes of real estates, selecting the attributes actually influencing the market prices of real estates.

## References

- [1] Jasińska E.: *The application of factorial analysis in the evaluation of real estates*. A thesis of AGH University of Science and Technology, Cracow 2005 [in Polish]
- [2] [www.statsoft.pl](http://www.statsoft.pl) [Course Manual for the Statistica 6.0 pack]