

TYOLOGY OF FOREIGN STUDENTS INTERESTED IN STUDYING AT CZECH UNIVERSITIES

Marta Žambochová

Introduction

The main motivation of our research was to analyze alternative ways of funding tertiary education. Primarily, we focused on the tuition fees payable by foreign students. The main research questions were.

- Who (or what) influences the student in deciding whether to travel?
- What is expected from a student studying abroad?
- What the student feels could be an obstacle to studying abroad?
- What is the basic amount for monthly living expenses while studying abroad, which the student is willing to accept?

We have created a classification of foreign candidates for studying at Czech universities, based on specific information. We have described the main characteristics of each category, their financial capacity and motivation, but also their mutual differences. We used some selected tests involving hypotheses, cluster analysis and differing methods for creating decision trees in the course of our examination.

1 Methodology – Rresearch Methods

1.1 Cluster Analysis

In carrying out the classification of students interested in studying abroad, we first used a cluster analysis, which is one of the classification methods of „unsupervised learning“. Cluster analysis [1], [2] deals with the similarities of data objects. It divides sets of objects into several previously unspecified groups (clusters) so that objects within individual clusters were

the most similar and objects from different clusters were the least.

Cluster analysis can be performed by way of application of many different methods. The various methods differ both in terms of determining the similarity of objects (similarity measure) as well as the application of a cluster analysis.

The selection of the method of cluster analysis depends first and foremost on whether we have direct source data or aggregated data (eg, frequency table, or matrix of similarities). If we have source data, the choice of method depends on the type of variables (nominal, ordinal, quantitative variables).

In our case we had the source data and we worked with a combination of different types of variables and thereby we limited the selection of method. Otherwise we would have had to modify the data transfer of each of the nominal or ordinal variables on quantitative variables, more precisely, into several binary variables (acquiring only the values of 0 and 1).

The processing of our data is based on a relatively large number of objects. That is the reason why the hierarchical methods are not suitable. *K*-means algorithm is designed for clustering objects described by quantitative variables, which was not applicable in our case. To use this method, we had to preprocess the data by means of binarization, ie the transfer of each categorical variable to the number of binary variables (variable acquiring only the values 0 and 1). The best method for processing data was a two-step method.

The principles of the TwoStep cluster analysis are given in [2]. This method uses an algorithm Birch (Balanced Iterative Reducing and Clustering using Hierarchies), which is

described in more detail in [8], or [9]. The TwoStep method was first included in the 11.5 version of the statistical system SPSS.

1.2 Decision Trees

Decision trees belong to a group of classification methods known as „supervised learning“, where the decision rules for the classification of objects into classes are based on the creation of a learning (training) set. They are a very widespread group of trees, and are used in data models. Decision trees are structures that recursively divide the examined data according to certain decision-making criteria. The root of the tree represents the entire population file. The internal nodes of the tree represent a subset of the population file. The tree leaves can be read as values of a dependent variable. They employ two types of decision trees, classification trees (each sheet is a class assignment) and regression trees (each leaf is assigned constants – to estimate the value of the dependent variable).

A decision tree is generated recursively as dividing the space of predictors values (explanatory variables) on the basis of search questions (splitting rules) that best divides the space into subsets of examined data.

The process of division stops when the criterion to stop emerges (the stopping rule). The next step is to prune the tree. It is necessary to determine the „correct“ size of the tree (trees too small do not capture all the regularities in the data, very large trees include a description and some properties of random data). The generation of the sub-trees arises from the tree algorithm and compares the quality of the generalization of these sub-trees (as well as describes the data).

The procedure may be such that the first decision trees are created on the training data and then their quality is subsequently verified by means of the test data. Another way is a form of cross-validation in order to create the tree and its sub-trees to apply all the data. Once the data is divided into several disjoints of approximately equal parts, it then gradually forms and a part of the data is continually removed from the file. The quality of the tree and its sub-trees is verified using the resulting data sets. The selection of a subtree is such that the lowest estimate of the actual error becomes apparent. If there are multiple

subtrees with comparable estimates of actual errors, it selects the smallest one.

Many algorithms were developed to create decision trees. Most of these are used by CART, ID3, C4.5, AID, CHAID, and QUEST, see eg [4], [7] or [10]. For the work we used three types, the algorithms are implemented in the SPSS statistical system, and CART, CHAID, and QUEST.

1.3 Tests of hypotheses

The Kruskal Wallis (or ordinal) test is a non-parametric test and is suitable for different choices of layouts from the normal distribution. The null hypothesis is as follows, „all samples have the same layout“. The test works by ranking the individual elements. See [3]. The Median test is used for testing whether two groups differ in their median value. It is a nonparametric test. Chi-square test for independence assesses whether paired observations on two variables, expressed in a contingency table, are independent of each other.

2 Data processing

2.1 The data structure

The research sample consisted of 1093 students from 6 countries, namely: Slovakia, Greece, Ukraine, Russia, Belarus and China. Most respondents were from China. Chinese students accounted for almost half of the monitored file. The smallest sample data was gathered in the Russian Federation.

The average age of the students involved was 20.87 years. Of these respondents, 62 % were women and 38 % men. Nearly half the respondents were students majoring in a social specialization, as opposed to only one tenth of the respondents being students of humanities or artistic disciplines. Technical and natural sciences were represented by approximately 20 %.

The data also revealed that half of the respondents are students at the bachelor level, as well as students planning to undertake a higher level of education abroad. A quarter of the respondents are secondary school students, as well as students planning to study abroad for bachelor degree studies. However, it is not true that each of the students wishing to undertake degree studies abroad are directly traceable to the current degree.

Students, where neither parent had been an undergraduate, dominated in the sample data. Employing the Chi-square test, we verified that the division into categories according to the level of the parents' education diverges in different countries. In most of the countries surveyed, the structure was consistent with the structure of the total population ie, the largest proportion of students who do not have one of the parents university educated (this category is most pronounced in China – 80.9 % and 70.9 % of the Slovak Republic) and less numerous categories of students whose parents had both been undergraduates (the least frequent in this category were the Slovak Republic – 6.9 % and China – 7.3 %). The Russian Federation and the Ukraine were significantly different in this respect The Russian Federation had the most numerous group (38.2 %) of students where both parents had been undergraduates. The most numerous group (with 42.1 %) of students with one university educated parent was in the Ukraine.

2.2 Classification of students

On the basis of variables related to

- Considering study abroad (yes – perhaps even – no).
- The length of study period (month – semester – year – longer).
- The level of study programme (Bachelor – Master – Doctoral).
- Type of study programme (social – natural – technical – artistic sciences).
- Learning the language (Czech – English/ /Czech – English).
- The length of the preparatory language course (1/2 year – year – or longer).
- Desired outcome of the study programme (level of the degree, the pan-European Bc. – Mgr./ Ing. – Ph.D.; level of the dual degree).

Here a cluster analysis was performed, especially the procedure of SPSS TwoStep Analysis, which created four clusters. We have dealt with this issue already in [6].

Description of the Characteristics of People from Different Clusters

1st cluster (348 respondents)

- Ambivalent about whether to study abroad,
- If they would study abroad, only for 1 semester or 1 year,

- They would rather study at the the Bachelor level,
 - Most of the students are in social science disciplines ,
 - They would be interested in studying in Czech or in the combination Czech + English,
 - They would like to have completed a course in the language within 6 months,
 - 1500 Euros for a 6 month course does not seem advantageous to them (maybe they would consider it),
 - They would interested in an outcome with a European -wide certificate validity – level Bc. or Mgr.
- 2nd cluster (382 respondents)
- Ambivalent about whether to study abroad, ñlf they would study abroad, then for 1 year,
 - They would rather study at the Master's level,
 - Most are students in social sciences , technical and artistic disciplines ,
 - They would be interested in studying in English,
 - They did not think about the tuition fee of 1500 Euros for a six-month course,
 - They would interested in a degree with a European – wide certificate validity – level Mgr. or Bc.
- 3th cluster (27 respondents)
- Not interested in studying abroad,
 - Other questions mostly remained unanswered.
- 4th cluster (336 respondents)
- They have thought about studying abroad, but not at the present moment ,
 - If they would study abroad, it would be for longer than 1 year,
 - They would rather study at the Doctoral level,
 - Most of the students are from the natural and technical sciences disciplines,
 - They would be interested in studying in a combination of Czech + English,
 - They would like to have a completed language course duration of 1 year,
 - The tuition fee of EUR 3000 per year for the course they find acceptable (maybe would consider it),
 - They would interested in a degree with a European – wide certificate validity – all levels.

Slovakia dominated in the 1st cluster in terms of states, Greece in the 2nd cluster and

China in the 4th cluster. Nobody dominated in the 3th cluster.

In addition, we employed the Chi-square test for independence in which we examined which of the following questions regarding the current situation were raised and this correlates with the distribution of the generated clusters.

- Sex (M – F)
- Age
- Degree studies (secondary – Bc. – Mgr./Ing. – Ph.D.)
- Field studies (social sciences – natural sciences – technical – artistic)
- Accomodation during the study period(at home – near the school)
- The financial level of the family (below average – average – above average)
- The payment of tuition fees (no – yes/no problems – yes/minor problems – yes/high load)
- Study results (below average – average – above average)
- Employment opportunities (no problems – problems)
- Tourist stay abroad (no – yes)
- Study abroad (no – yes)
- Another stay abroad (no – yes)
- Considering studying abroad (not – perhaps ever – yes)
- Number of university-educated parents (none – one – two)
- Degree of influence on decision making (1 ... max – 5 ... min)
 - family
 - examples in the surroundings
 - teachers at school
 - promotional materials
- The rating of possible expectations (1 ... max – 5 ... min)
 - knowledge of the country
 - to learn a language
 - studies at a prestigious school
 - gaining practical experience
 - a degree with EU-wide validity
 - better labor market
 - settlement in the country after studying
- Degree of concern (1 ... max – 5 ... min)
 - separation from family
 - separation from friends
 - security situation in the country
 - financial costs associated with studying abroad
 - language barriers
 - different religious sentiments

The null hypotheses of these tests were independent of given variables and the variable meaning belongs to a given cluster. We tested at the 5% significance level.

Only a few variables released the p-value of 0.05. This means that these variables are not affected by membership of individual clusters. The group of these variables includes both concerns about the security situation in the country (p-value = 0.586), concern about the financial costs associated with the stay (p-value = 0.575), fear of different religious beliefs (p-value = 0.206), fear of language barriers (p-value = 0.168) and expectations of obtaining a diploma with a Europe-wide validity (p-value = 0.278), the expectation of settling there after the study programme (p-value = 0.219), expectations for better labor market outcomes (p-value = 0.133).

For other variables, the p-value went up to 0.038, even though for most variables the p-value is almost zero, indicating rejection of the null hypothesis of independence. Because these variables are correlated with the distribution of the clusters, therefore we can say what the prevailing value of the variable for the cluster is.

In the next step of the classification process, we designed decision trees. We used three algorithms implemented in the statistical system SPSS (CART, CHAID, and QUEST).

We chose the newly created variable „belonging to the cluster“ for the dependent variable and we chose the variables belonging to individual survey questions, which in previous Chi-square tests for independence demonstrated dependence, as the dependent variable. The Values risk estimates published in all of the produced trees were similar (in the range 0.282 to 0.325. It is not ideal, but at an adequate level. The worst hit was the QUEST algorithm, primarily CART.

According to various algorithms, members of the clusters in terms of answers to selected questions can be categorised as follows (3rd in the cluster or one type of tree not exempted – will be featured in the list).

QUEST Algorithm

1st cluster

- Students of social science. Disciplines.
- 2nd cluster
- Secondary school students of humanistic or artistic directions and university students currently paying tuition.

Fig. 1: Classification Tree Generated by CART

Source: own

4th cluster

- Students of natural and technical science disciplines currently paying tuition.

CART Algorithm (see Fig. 1)

1st cluster

- Students of social science disciplines coming from families with below-average incomes.

2nd cluster

- Students coming from families with average to above-average incomes, exhibiting an average to below-average academic performance, who currently pay tuition fees, whereby this doesn't cause additional problems.

4th cluster

- Students of natural, technical and artistic disciplines, who don't currently pay tuition fees, or if were the case it would cause them great difficulty. These students expect to have practical experience abroad.

CHAID Algorithm

1st cluster

- Students of social science disciplines, expressing a strong fear of separation from the family (answers 1 and 2 on a five-point scale) and coming from families with below-average incomes.

2nd cluster

- Students of Bachelor degree studies with average to below-average academic performance, without much fear of being separated from the family (3 to 5 answers on the five-point scale).

4th cluster

- Students of natural and technical science disciplines exhibiting a strong fear of being separated from the family (answers 1 and 2 on a five-point scale).

This classification furnished us with interesting results. The generated clusters are defined/analysed very well (see the above description of the clusters). Furthermore, it appears that in each of the resulting clusters, belonging to one (or several) countries of origin of the student is predominant. It pays to invest in segments with a sufficient number of students from the same country. The **fourth cluster of predominantly Chinese students appears to be very promising from this point of view.**

The various algorithms for building classification trees also provided us with some

good and reasonably interpretable results. **The CHAID algorithm showed how important the fact of student separation from the family is. This implies an important argument for the creation of a marketing strategy proposing a „foster family.“ The CART algorithm again outlined the fourth segment and is a cluster of low-income students who are interested in practical work experience. If the program is designed and offered as a „paid internship“ it will not necessitate the need to spend a lot of money on marketing activities necessary for the attraction of such students.**

A university of regional significance could also be very successful in attracting foreign students belonging to another cluster, ie, students without financial problems, with weaker academic performance, who are likely to accept less prestigious universities and appreciate all the other activities geared to their personal development (using their own personal funds).

In the next phase, we performed an analysis of the jurisdiction of the respondent to the individual clusters and the amount of tuition that the student is willing to pay.

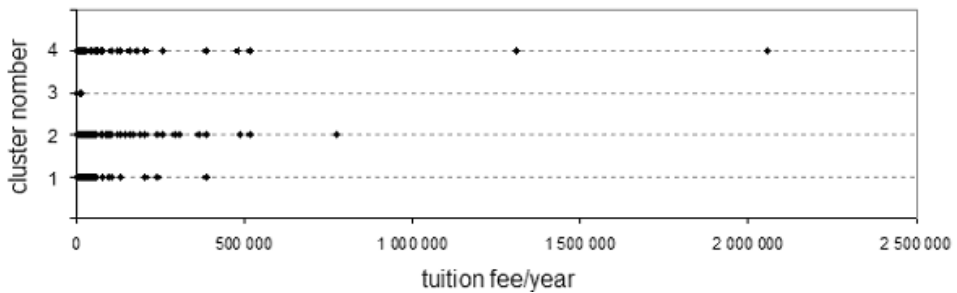
First, we performed the Kruskal-Wallis test to determine whether the value of the tuition fees within the jurisdiction of the various groups was comparable to a cluster or not (i.e., independence or dependence of observed variables).

The null hypothesis was independence and we worked at the 5% significance level. The resulting p-value was almost zero (more precisely 1.45×10^{-14}). This means the 5% level rejecting the null hypothesis. For the students assigned to individual clusters, there are statistically significant differences in the amount of annual tuition fees that students are willing to pay. Figure 2 shows the nonzero values of the annual tuition fees (CZK) within individual clusters.

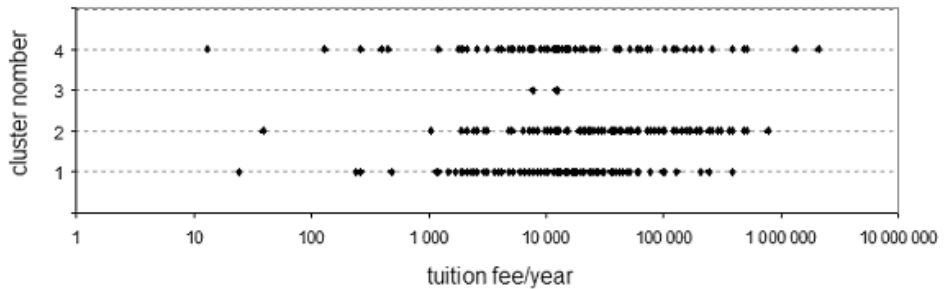
In a group dominated by the lower value of the annual tuition fees, there are also very significant outliers (more than 2 million CZK). Therefore, the graph in Fig. 2 is not perfectly assessable. We display identical data in a logarithmic scale, which allows a different assessment of this data, in Fig. 3.

The collected data are summarized in Tab. 1. In each cluster is shown here:

- number of students assigned to the cluster, n
- number of students in the cluster, who

Fig. 2: The Annual Tuition Fees According to the Clusters (Common Scale)

Source: own

Fig. 3: The Annual Tuition Fees According to the Clusters (Logarithmic Scale)

Source: own

fulfilled the nonzero tuition fees and are willing to pay,

- proportion of students with a completed non-zero tuition fees of all students in a given cluster,
- average tuition fees (CZK), which are paid by the students and the cluster,
- average tuition fees (CZK) calculated from all non-zero values in a given cluster.

Tab. 1: Comparison of Clusters According to the Amount of Annual Tuition

cluster	number of respondents	mean (in CZK)	number of non-zero values	proportion of non-zero values	mean of non-zero values (in CZK)
1	348	16,555.49	282	0.8103	20,488.88
2	382	54,603.64	269	0.7042	77,947.21
3	27	1,203.07	3	0.1111	10,827.67
4	336	51,363.53	262	0.7798	65,870.78

Source: own

Both graphs (Fig. 2 and Fig. 3) and Tab. 1 show that in terms of annual tuition, that the student is willing to pay, the worst cluster number is 3. In this cluster there is only 11.11 % of non-zero values, i.e. 88.89 % of these students are not willing to pay any tuition fees. At the same time, both observed average amounts of annual tuition fees are the lowest of all clusters. This cluster is assigned to students, who are not thinking about studying abroad, and thus they are uninteresting from our point of view, hence the poor results of this cluster are not necessarily negative.

Cluster number 1, which are assigned to students from families with below-average incomes who are planning a short-term bachelor's degree in social sciences and are willing to study in Czech, in Czech or in combination with English, contains the highest proportion of students who are willing to pay any tuition fees (81.03 %). Unfortunately the average amount of tuition, that these students are willing to pay is quite low (20,489 CZK, 16,555 CZK respectively).

Cluster number 2, in which students are from families with above-average income, with average or below-average academic performance who plan on studying abroad in the medium term follow-up of social, artistic or technical sciences, contains a high proportion of students who complete a non-zero amount of any fees (70.42 %). Both of the average level of fees that the student is willing to pay is the highest of all clusters (77,947 CZK, 54,604 CZK respectively).

Cluster number 4, in which students are planning a long-term foreign doctoral study

programme in the natural sciences or engineering disciplines who would be willing to pay for a year-long czech language course and continue to study in the combined Czech + English language, and who after completing the study programme abroad, would then like to have some practical experience in the field. This contains a very high proportion of non-zero values of fees (77.98 %). This means that most students in this cluster are willing to pay tuition. Both of the average values are also high in comparison with other clusters (65,871 CZK, 51,364 CZK respectively).

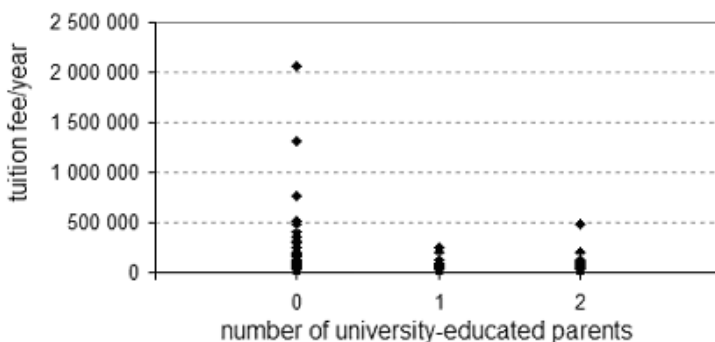
It can be gleaned from the above that the most advantageous group to focus on must be the fourth cluster of students, because they are willing to pay a high amount of long-term tuition fees. The total value that these students will pay is therefore the highest. The problem is the need for accreditation of the doctoral studies.

Very conveniently, it also appears to focus on the master degree students, who, although only in the medium-term plan to stay abroad, nevertheless are willing to pay the highest average annual amount of any annual fees.

2.3 Monitoring the Relationship between Variables and Parents' Education Level of Fees

Fig. 4 and 5 are showing the value of any annual fees (CZK) depending on the number of university-educated parents for an ordinary student, respectively, in a logarithmic scale.

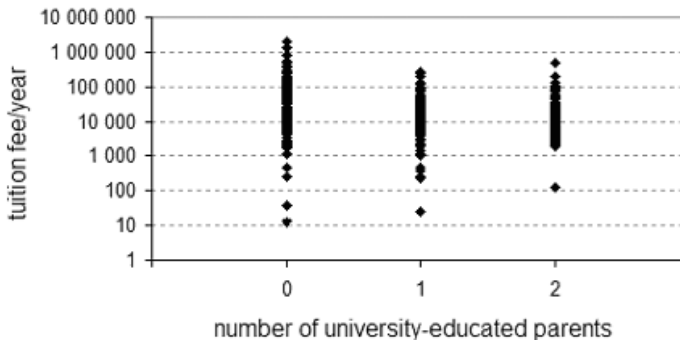
Fig. 4: The Annual Tuition Fees According to the Number of University-Educated Parents (Common Scale)



Source: own

Fig. 5:

The Annual Tuition Fees According to the Number of University-Educated Parents (Logarithmic Scale)



Source: own

The collected data are summarized in Tab. 2. They are presented in the following information:

- Students with a certain number of university-educated parents,
- Number of students assigned to the cluster,
- Number of students in the cluster, who fulfill the nonzero tuition fees that are willing to pay,
- Proportion of students with a completed non-zero tuition fees of all students in a given cluster,
- Average tuition fees (CZK), which are paid by students and the cluster,
- Average tuition fees (CZK) calculated from all non-zero values in a given cluster.

Tab. 2:

Comparison of the Number of University-Educated Parents and the Subsequent Amount of Annual Tuition

number of university parents	number of respondents	mean (in CZK)	number of non-zero values	proportion of non-zero values	mean of non-zero values (in CZK)
0	698	52,871.14	496	0.7106	69,606.57
1	226	23,212.84	170	0.7522	30,995.97
2	158	24,969.02	135	0.8544	29,223.00

Source: own

We performed a Kruskal-Wallis test to determine whether the values of the level of fees within each class given the number of university-educated parents of the student are or are not comparable (ie, independence or dependence of observed variables).

The null hypothesis was independence and we worked at the 5% significance level. The resulting p-value was very small (more precisely, 0.001347). This means that the 5% level rejects the null hypothesis, i.e. **the amount of annual tuition fees that students are willing to pay depends on the number of university-educated parents of the student.**

From Tab. 2, the highest average annual amount for tuition students are willing to pay who do not even have one college-educated parent can be seen (69,607 CZK, 52,871 CZK respectively). The percentage of students who are willing to pay at least some tuition in this category is also quite high (71.06 %).

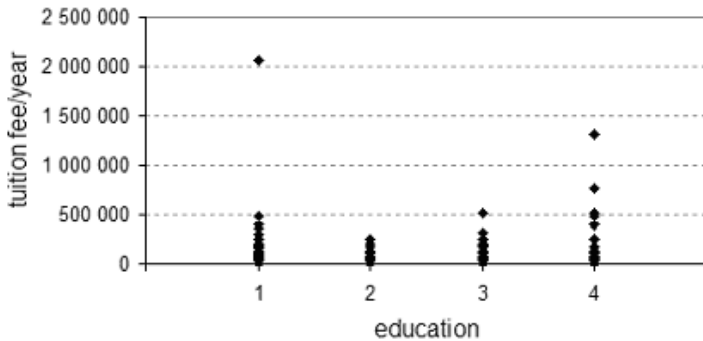
The highest proportion of students who are willing to pay at least some tuition is highest for students whose parents are both university graduates (85.44 %).

2.4 Monitoring the Relationship between Variables and Higher Education Student Tuition

We performed a Kruskal-Wallis test to determine whether the values of the level of

fees within each class given the current level of education of the parents of the student are, or are not comparable (i.e., independence or dependence of observed variables).

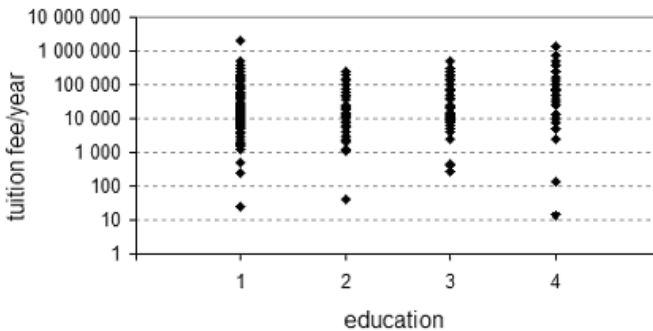
Fig. 6: The Annual Tuition Fees According to the Level of the Student's Education (Common Scale)



Source: own

Note: 1 ... secondary, 2 ... Bc., 3 ... Mgr./Ing., 4 ... Ph.D.

Fig. 7: The Annual Tuition According to the Level of Education of the Student (Logarithmic Scale)



Source: own

Note: 1 ... secondary, 2 ... Bc., 3 ... Mgr./Ing., 4 ... Ph.D.

The null hypothesis of independence was two and we worked at the 5% significance level. The resulting p-value was almost zero (more precisely 5.9656×10^{-8}). This means that at the 5% level, is the rejection of the null hypothesis of independence, ie the amount of annual tuition fees that students are willing to pay depends on the level of education of the student.

Fig. 6 and 7 show the amount of any annual fees (CZK) according to the ordinary student undergoing education, respectively, in logarithmic scale.

The collected data are summarized in Tab. 3. There are presented as follows here:

- Educational level of the student,
- Number of students assigned to the cluster,
- Number of students in the cluster, who

fulfilled nonzero tuition fees that are willing to pay,

- Proportion of students with a completed non-zero tuition fees of all students in a given cluster,
- Average tuition fees (CZK), which are paid by students and the cluster,
- Average tuition fees (CZK) calculated from all non-zero values in a given cluster.

Tab. 3 shows that the highest average annual amount students are willing to pay for tuition is by those students who have a doctoral education (118,314Kč, 81,209 CZK respectively). The percentage of students who are willing to pay at least some tuition in this category is high (66.38 %).

The highest proportion of students who are willing to pay at least some tuition is highest for students having a bachelor's degree (81.68 %).

Tab. 3: Comparison of Students' Educational Level and the Amount of Annual Tuition

education	number of respondents	mean (in CZK)	number of non-zero values	proportion of non-zero values	mean of non-zero values (in CZK)
secondary	223	25,433.35	134	0.6009	42,325.52
bachelor	535	34,601.45	437	0.8168	42,540.10
follow-up	202	48,421.04	161	0.7970	60,751.81
doctoral116	81,209.42	77	0.6638	118,314.16	

Source: own

Tab. 3 shows that the highest average annual amount for tuition students are willing to pay are those students who have a doctoral education (118,314 CZK, resp. 81,209 CZK). Percentage of students who are willing to pay at least some tuition in this category is high (66.38 %). The highest proportion of students who are willing to pay at least some tuition is highest for students having a bachelor's degree (81.68 %).

In the following part of the research we focused on monitoring the acceptable level of fees depending on other factors (Who or what affects student when deciding whether to travel. What is expected of students studying abroad? What could amount to an obstacle for students who would like to study abroad?) We used the Kruskal-Wallis and MedianTest. The results of all tests are summarized in Tab. 4. We have dealt with this issue already in [5] and [11].

In all observed cases, if the p-value is less than 0.05, we rejected the null hypothesis of independence. From the resulting p-values, it is thus obvious **that an acceptable level of fees depends on the influence that travel, family and the teachers at the school, then on the**

assumption that study abroad will improve the knowledge of foreign languages, the possibility of gaining practical experience in the foreign country and of acquiring a diploma with European-wide validity. A certain amount of tuition is acceptable but is negatively impacted by the fear of separation from friends and the language barrier.

We observed those factors which influence the division of students the most. We created a classification tree for the dependent variable „acceptable level of fees“ in CZK. We selected all observed factors as explanatory variables. We employ two methods to create a tree (CHAID and CART). The quality of the result was superior when applying the CHAID. The tree structure shows **that the most acceptable level of tuition fees is distinguished from the viewpoint of obtaining practical experience after graduation abroad and then from the viewpoint of improvement in a foreign language.** These are the two most important factors in the conducting of marketing campaigns aimed at attracting foreign students.

Tab. 4: The Result of Kruskal-Wallis and Median Tests to Monitor the Level of Fees Depending on Various Factors

question	Kruskal-Wallis test		Median Test	
	CHI	p-value	CHI	p-value
15-family	9.830607606	0.0434	11.00494301	0.0265
15-example	6.883056454	0.1422	10.19532294	0.0373
15-staff	15.6785413	0.0035	19.84890014	0.0005
15-media	13.54819165	0.0089	6.469089732	0.1667
16-culture	9.376714799	0.0523	13.07074678	0.0109
16-langue	15.36958979	0.004	21.65624135	0.0002
16-prestige	5.483692695	0.2412	7.01786513	0.1349
16-practice	17.96345757	0.0013	13.33873788	0.0097
16-degree	9.93869166	0.0415	11.29288298	0.0234
16-job	3.199492824	0.5250	5.712854282	0.2216
16-live	2.409758715	0.6609	3.191380815	0.5263
17-family	2.229029992	0.6937	1.301200215	0.8611
17-friends	10.97936615	0.0268	12.41120044	0.0145
17-country	4.874616897	0.3004	2.782469038	0.5948
17-costs	4.600302626	0.3308	1.96693061	0.7418
17-language	11.711417	0.0196	11.01600837	0.0263

Source: own

Conclusions

This article demonstrated what significance for the student, separation from the family assumes. This necessitates the creation of a marketing strategy involving a „foster family“. Furthermore, a group of students with low incomes who are interested in practical work experience abroad indicated a possible marketing perspective. If the program is designed and offered as a „paid internship“ there will not need to be an extensive and expensive marketing strategy necessary for the attraction of such potential students.

It appears to be very advantageous to focus on master degree students, who, although only planning a medium-term stay abroad, nevertheless are willing to pay the highest average annual amount of any annual fees. Doctoral students are willing to pay a high amount for long-term tuition, thus the total value that these students will pay will be the highest. The problem is the need for accreditation of such doctoral studies.

A university of regional significance could also be very successful in attracting foreign students without financial problems, displaying weaker academic performance, who are likely to accept less prestigious universities and appreciate all the other activities geared to their personal development (using their personal funds).

Acknowledgment: This research was financed by an internal grant agency within the J.E. Purkyne University in Usti nad Labem, the Czech Republic.

References

- [1] KLÍMEK, P. Shlukovací metody v data miningu. *E+M Ekonomie a Management*. 2008. Vol. 11, Iss. 2, pp. 120–127. ISSN 1212-3609.
- [2] ŘEZANKOVÁ, H., HÚSEK, D., SNÁŠEL, V. *Cluster analysis of data*. Praha: Professional Publishing, 2009. 220 p. ISBN 978-80-86946-81-8.
- [3] ŘEZANKOVÁ, H. *Analýza dat z dotazníkových šetření*. 2. vyd. Praha: Professional Publishing, 2010. 217 p. ISBN 978-80-7431-019-5.

- [4] TIMOFEEV R. *Classification and Regression Trees (CART) Theory and Applications*. Berlin, 2004. A Master Thesis. Humboldt University, CASE-Center of Applied Statistics and Economics.
- [5] TIŠLEROVÁ, K., ŽAMBOCHOVÁ, M. Study Abroad: are Expectations of Chinese Students Different? In CHEN, D. (ed.). *2nd International Conference on Education and Management Technology, IPCSIT. August 19–21 2011*. Vol.13. Shanghai (China), Singapore: IACSIT Press, 2011. pp. 82–86. ISSN 2010-4626.
- [6] TIŠLEROVÁ, K., ŽAMBOCHOVÁ, M. Marketing Segmentation of Students Willing to Study Abroad Based on Cluster Analysis. *Engineering and Technology. World Academy of Science*. 2011, vol. 77, pp. 468–472. ISSN 2070-3740.
- [7] WILKINSON, L. *Tree Structured Data Analysis: AID, CHAID and CART* [online]. Sun Valley: ID, Sawtooth/SYSTAT Joint Software Conference, 1992. [cit. 2011-09-07]. 10 p. (PDF). Available from: <[http://bus.utk.edu/stat/DataMining/Tree%20Structured%20Data%20Analysis%20\(SPSS\).pdf](http://bus.utk.edu/stat/DataMining/Tree%20Structured%20Data%20Analysis%20(SPSS).pdf)>.
- [8] ZHANG, T., RAMAKRISHNAN, R., LIVNY, M. BIRCH: An Efficient Data Clustering Method for Very Large Databases. *ACM SIGMOD Record*. 1996, Vol. 25, Iss. 2, pp. 103–114. Available also form: <<http://www.cs.sfu.ca/CourseCentral/459/han/papers/zhang96.pdf>>. ISSN 0163-5808.
- [9] ZHANG, T., RAMAKRISHNAN, R., LIVNY, M. BIRCH: A New Data Clustering Algorithms and Its Applications. *Journal of Data Mining and Knowledge Discovery*. 1997, Vol. 1, Iss. 2, pp. 141–182. ISSN 1384-5810.
- [10] ŽAMBOCHOVÁ, M. Jak na rozhodovací stormy. *Informační Bulletin*. 2008, Vol. 19, Iss. 3, pp. 1–12, ISSN 1210-8022.
- [11] ŽAMBOCHOVÁ, M., TIŠLEROVÁ, K. Potential of Indirect Financing of Higher Education Institutions in Terms of Global Economic Development. In KOCOREK, A. (ed.). *Proceedings of the 10th International Conference Liberec Economic Forum 2011*. Liberec: Technical university of Liberec, 2011, pp. 604–613. ISBN 978-80-7372-755-0.

RNDr. Marta Žambochová, Ph.D.

University of JE Purkyne
Faculty of Social and Economic Studies
Department of Mathematics and Informatics
marta.zambochova@ujep.cz

Doručeno redakci: 7. 11. 2011

Recenzováno: 28. 12. 2011, 20. 1. 2012

Schváleno k publikování: 5. 4. 2012

TYOLOGY OF FOREIGN STUDENTS INTERESTED IN STUDYING AT CZECH UNIVERSITIES

Marta Žambochová

This paper undertakes and investigates some marketing segmentations of those students wishing to study abroad and the economic potential of such resulting segmentation within the individual groups. Three algorithms were employed as classification trees – CART, CHAID and QUEST for its evaluation from the group of supervised learning methods. From the group of unsupervised learning methods the two–step cluster analysis was chosen. This paper analyzes the results obtained using these methods. It aims to create classes of respondents harbouring similar opinions with respect to tuition fees to be paid by foreign students. We have used various tests for the hypotheses such the as Chi-square Test of Independence and the Kruskal-Wallis Test.

The research sample consisted of more than one thousand students from six countries. A „perfect“ type of person as candidate was profiled in terms of willingness and the financial means to pursue their studies in the Republic invested and this survey formed the basis of the profiling. Our research shows that an acceptable level of fees depends on several factors such as: the amount of travel involved, family and teachers at the school, then on the assumption that the study stay abroad would improve knowledge of foreign languages, practical work experience offered in the foreign country and the subsequent acquisition of a diploma with Europe-wide validity. There exists an acceptable level of tuition and this in turn reveals that the most acceptable level of tuition fees may be dependent upon the viewpoint of potential practical work experience abroad and then from the viewpoint of improvement in a foreign language. The amount of annual tuition fees that students are willing to pay depends too on the number of university-educated parents the students possess.

Key Words: Classification, Cluster Analysis, Decision Trees, Marketing Segmentation, Economic Potential.

JEL Classification: C12, C38, H52, I25, M30.