

Yayın Geliş Tarihi (Submitted): 24/12/2020

Yayın Kabul Tarihi (Accepted): 02/03/2021

Makele Türü (Paper Type): Araştırma Makalesi – Research Paper

Please Cite As/Atıf için:

Demirel O. (2021), A factor analysis application on income sources of university students in European countries, *Nicel Bilimler Dergisi*, 3(1), 16-42. doi: 10.51541/nicel.846104

A FACTOR ANALYSIS APPLICATION ON INCOME SOURCES OF UNIVERSITY STUDENTS IN EUROPEAN COUNTRIES

Oğuzhan Demirel¹

ABSTRACT

This statistical study has been carried out to examine and interpret the income types that make up the income sources of students. For this purpose, a data set has created using 14 European countries that included all the data for 9 variables. These countries have been chosen as they did not contain missing data for each variable. Initially, factor analysis, which is a statistical method suitable for the purpose of the study, has been applied to a data set. Then, principal component analysis (PCA), one of the exploratory factor analysis (EFA) methods, has been preferred for this analysis. According to the rotated component matrix results, a result consisting of 3 factors has been obtained for 8 variables. These factors consisting of variables that make up the income sources of university students; has been named “income that students can save”, “students' main sources of income” and “special incomes of students”.

Keywords: Student Income Sources, Factor Analysis, Principal Component Analysis

¹Sorumlu yazar, İstatistik Bölümü, Fen Fakültesi, Hacettepe Üniversitesi, Ankara, Türkiye, ORCID ID: <https://orcid.org/0000-0003-4352-6531>

AVRUPA ÜLKELERİNDEKİ ÜNİVERSİTE ÖĞRENCİLERİNİN GELİR KAYNAKLARI ÜZERİNE BİR FAKTÖR ANALİZİ UYGULAMASI

ÖZET

Bu istatistiksel çalışma, öğrencilerin gelir kaynaklarını oluşturan gelir türlerini incelemek ve yorumlamak amacıyla yapılmıştır. Bu amaçla, 14 Avrupa ülkesi kullanılarak 9 değişken için tüm verileri içeren bir veri seti oluşturulmuştur. Bu ülkeler, her bir değişken için eksik veri içermedikleri için seçilmiştir. İlk olarak, araştırmanın amacına uygun istatistiksel bir yöntem olan faktör analizi veri setine uygulanmıştır. Daha sonra, açıklayıcı faktör analizi (AFA) yöntemlerinden biri olan temel bileşen analizi (TBA), bu analiz için tercih edilmiştir. Döndürülmüş bileşen matrisi sonuçlarına göre 8 değişken için 3 faktörden oluşan bir sonuç elde edilmiştir. Üniversite öğrencilerinin gelir kaynaklarını oluşturan değişkenlerden oluşan bu faktörler; “öğrencilerin biriktirebileceği gelir”, “öğrencilerin ana gelir kaynakları” ve “öğrencilerin özel gelirleri” olarak adlandırılmıştır.

Anahtar Kelimeler: Öğrenci Gelir Kaynakları, Faktör Analizi, Temel Bileşenler Analizi

1. INTRODUCTION

One of the most important periods in human life is the period of university education. This is the period when students are freest in their lives, develop socially, and their thoughts begin to become stereotyped in their minds. The university period, which is the best period of student life, is undoubtedly shaped by the income of the students. In this period, student incomes may have many sources. These sources can be national state student support, family, spouse, private scholarship, and work. The size and type of income obtained by students can change the purpose of using the resource. At this point, the income sources of university students have been examined in this study. The main purpose of this study is; to classify the monthly income sources of university students, to group the sources that are related to each other, and to interpret them statistically. Due

to the most appropriate statistical method for the purpose of the study is factor analysis, it was decided to perform all these applications with factor analysis.

2. LITERATURE REVIEW

Since no studies have been conducted on the factors that make up the monthly income of university students, the literature has been examined in two sections: studies related to factor analysis and studies related to students.

2.1. Studies Related to Factor Analysis

Muthén (1991) have analyzed mathematics success data from the Second International Mathematics Study (SIMS) in which the U.S.A. students are measured at the starting and end of eighth grade in this article. The article aims to address some important analysis questions in the SIMS data and indicate the potential of multilevel factor analysis methodology. Issues related to between- and within-class dissociation of success variance and the change of this dissociation over the course of the eighth grade have been studied. Initially, random effects ANOVA has been considered for each success score. The effects of unreliability on variance dissociation have been shown with the help of a multilevel factor analysis model. It has been found that the strong elements of following in eighth-grade math classes result in between-class variation in the success scores which is about as wide as the within-class student variation.

Martínez-García et al. (2007) have conducted this study to evaluate whether lung function, dyspnea, and extension of the disease are distinct factors in the influence of bronchiectasis upon patients using factor analysis. Patients with bronchiectasis diagnosed by high-resolution computed tomography (HRCT) and airflow obstruction determined by FEV1/FVCo 70% have included. Data have been collected relating to clinical history, three dissimilar clinical ratings of dyspnea, which are Medical Research Council (MRC), Borg Scale, and Basal Dyspnea Index, the extent of bronchiectasis, and functional variables. Four factors have been found that explained 84.1% of the total data variance. Factor 1 (45.6% of the data variance) has contained the three measurements of dyspnea. Factor 2 (16% of the data variance) has comprised airflow obstruction parameters (FEV1,

FEV1/FVC, and PEF). Factor 3 (13.8% of the data variance) has contained RV/TLC and RV (lung hyperinflation). Factor 4 (8.6% of the data variance) has included bronchiectasis extent. Dyspnea has more nearly correlated with lung hyperinflation than with airflow obstruction parameters. As a result, lung hyperinflation, airflow obstruction, dyspnea, and the lung extent of the bronchiectasis are four independent factors in the influence of bronchiectasis upon patients.

Dickinson and Adelson (2016) have applied multitrait-multimethod (MTMM) confirmatory factor analysis (CFA) within a structural equation modeling (SEM) framework to student accomplishment data to demonstrate experimentally how commonly used measures of student accomplishment may bounce different information about student performance. Using student population-level data from a sole state, this study provided a strong demonstration of the similarities and differences among three commonly used performance measures: American college testing (ACT) scores, state test scores, and grade point average (GPA). The results indicate that state test scores and ACT scores measured a similar achievement construct, while student grades reflected less of the achievement construct and a higher level of method effects. Besides, GPA method factor loadings have found higher than state assessment factor loadings, showing that there is less non-trait commonality between GPA measures and ACT measures than between the state assessment and ACT.

Feng et al. (2017), have aimed to validate the existence of underlying symptom sub-domains of the 22-item Sino-Nasal Outcome Test (SNOT-22) using confirmatory factor analysis (CFA) and to develop a sub-domain model that applies and researchers can use to define chronic rhinosinusitis (CRS) symptomatology in this study. A total of 800 patients with CRS have been added into this cross-sectional study. Their SNOT-22 responses have analyzed using exploratory factor analysis (EFA) to determine the number of symptom sub-domains. A CFA has made to develop a validated measurement model for the underlying SNOT-22 sub-domains along with diverse tests of validity and goodness of fit. EFA has demonstrated 4 distinct factors reflecting: sleep, nasal, otologic/facial pain, and emotional symptoms, independent of geographic locale. The use of the 4-sub-domain structure (sleep symptoms, nasal symptoms, otologic/facial pain

symptoms, emotional function) for SNOT-22 has been validated as the most appropriate to reckon SNOT-22 sub-domain scores for patients from dissimilar geographic regions using CFA.

Takenaka et al. (2021) have aimed to explore the risk factors for surgery-related complications in patients undergoing primary cervical spine surgery for degenerative diseases. 5,015 patients with degenerative cervical diseases who underwent primary cervical spine surgery from 2012 to 2018 were examined. As subcategories, the presence of cervical kyphosis $\geq 10^\circ$, the presence of ossification of the posterior longitudinal ligament (OPLL) with a canal-occupying ratio $\geq 50\%$, and foraminotomy were selected. Afterwards, multivariate logistic regression analysis was performed. As a result of the analysis performed, the significant risk factors ($p < 0.050$) for ULP were OPLL (odds ratio (OR) 1.88, 95% confidence interval (CI) 1.29 to 2.75), foraminotomy (OR 5.38, 95% CI 3.28 to 8.82), old age (per ten years, OR 1.18, 95% CI 1.03 to 1.36), anterior spinal fusion (OR 2.85, 95% CI 1.53 to 5.34), and the number of operated levels (OR 1.25, 95% CI 1.11 to 1.40). OPLL was also a risk factor for neurological deficit except ULP (OR 5.84, 95% CI 2.80 to 12.8), dural tear (OR 1.94, 95% CI 1.11 to 3.39), and dural leakage (OR 3.15, 95% CI 1.48 to 6.68). Among OPLL patients, dural tear and dural leakage were frequently observed in those with a canal-occupying ratio $\geq 50\%$. Cervical rheumatoid arthritis (RA) was a risk factor for SSI (OR 10.1, 95% CI 2.66 to 38.4).

Correddu et al. (2021) have analyzed the milk fatty acid profile of a sample of 993 Sarda breed ewes with PCA and MFA to compare the ability of these 2 multivariate statistical techniques in investigating the possible existence of latent substructures, and in studying the influence of physiological and environmental effects on the new extracted variables. Individual scores of PCA and MFA were analyzed with a mixed model that included the fixed effects of parity, days in milking, lambing month, number of lambs born, the altitude of flock location, and the random effect of flock nested within altitude. Both techniques detected the same number of latent variables (9) explaining 80% of the total variance. Due to PCA structures were rough for interpreting, only 4 principal components were associated with a clear meaning. Especially, principal component 1 was easy for interpreting and agreed with the interpretation of the first factor, with both being associated with the fatty acid of mammary origin. Key pathways of the milk fatty acid metabolism

were identified as mammary gland de novo synthesis, ruminal biohydrogenation, desaturation performed by stearoyl-coenzyme A desaturase enzyme, and rumen microbial activity, confirming previous findings in sheep and other species. In general, the new extracted variables were mainly influenced by physiological factors as days in milk, parity, and lambing month; the number of lambs born had no impact on the new variables, and altitude influenced only one principal component and factor. Both techniques were able to summarize a larger amount of the original variance into a reduced number of variables. Furthermore, factor analysis approved its ability to identify latent common factors clearly related to FA metabolic pathways.

2.2. Studies Related to Students

Christie, Munro and Rettig (2001) have aimed to discover the ways in which the current financial regime for supporting students influences on the choices they make while studying for their first degree in this article. They have focused especially on the financial choices students make (or feel forced to make) in relation to work, debt, and economizing. The results have indicated that the degree of discretion that students have is crucially related to the financial support they get from their parents. However, even where parents are generous, most students have sought an additional source of income to increase their autonomy in spending decisions. Parental behaviors have been found to be significant determinants of the ordering of drawing on other income. There have been obtained to be a financially undefended group of students whose fragile financial position largely results from their parents being unable to offer much financial support. This group in particular has found their time at university featured by considerable amounts of paid work and growing debt.

O'Farrell and Morrison (2003) have analyzed school ties and their related structure (for example, school engagement) using factor analysis. This study was conducted on a specific population of upper elementary-aged students, primarily Latino/a, from disadvantaged communities in the central California region. A factor analysis was conducted using selected survey items from various school bonding and related measures, yielding five factors. The orthogonal rotation would mathematically force the separation of factors. Therefore, the oblique rotation method was used for determining factors. Factor 1 consists items of a pattern of items that address

relationships that take place at school. Factor 2 consists items that tap the degree to which students feel accepted and cared about at school, and how competent they sense in school subjects. Factor 3 refers to students' feelings of belonging to a specific school as well as their sense of acceptance by adults who represent the institution of school. Factor 4 represents an attitude model generally related to academic tasks such as doing school work and academic issues. Factor 5 addresses students' expectations of their future academic and professional endeavors, perceptions of the importance of the school for personal and parents, as well as the level of supervision of the school and homework tasks. These results and factors indicate that school engagement consists of distinct dimensions and that students may have differing levels of engagement among the five dimensions. Gwosc and Schwarzenberger (2009) have presented an experimental analysis of the public funding system for higher education in Germany and comparison with five other European countries. A large number of discrete student support items in Germany has made it an exception and worked up the system obscure. The allocations of public expense to German institutions have under average. This situation has shown an indication of underfunding of teaching at higher education institutions. As far as the composition of support to households has concerned, Germany has provided the largest share of support for students' parents and the lowest share of cash help for students.

Karataş and Fer (2011) have made this study to determine the availability and credibility of the evaluation scale developed by the researcher stand on the principles of Stufflebeam's Context, Input, Process, Products (CIPP) Evaluation Model (1988) within the context of the evaluation of English curriculum of Yildiz Technical University. Whereas the scale preparation, by taking benefit of the theoretical principles of CIPP Evaluation Model the rude scale composing 65 items have created. Finally, the scale has performed on a large group of students (n=415) from dissimilar faculties of Yildiz Technical University. The results of the factor analysis showed that the CIPP Evaluation Model Scale had four factors occurring of context, input, process, and product and consisted of 46 items.

Karadeniz (2012) have obtained validity and reliability results of a scale for determining student satisfaction in blended learning environments where e-learning is supported by face-to-face

learning in this study. The structural equation model of the scale is built on three main factors, which are perceived usability of learner interface, perceived usefulness of e-content and course structure, and instructor support. Analyzes were made over 760 survey data. Confirmatory factor analysis was used for content validity and construct validity in the validity of the scale, and Cronbach Alpha coefficients were used for reliability. According to the results of the study, a valid and reliable scale for determining student satisfaction in blended learning environments was obtained.

Hoxby and Turner (2013) have used a randomized controlled trial to appraise interventions that provide students with semi-customized information on the application process and colleges' net costs in this study. These interventions also provide students with no-paperwork application fee waivers. The ECO Comprehensive (ECO-C) intervention costs about \$6 per student, and this situation has caused high-achieving, low-income students to apply and be accepted for more colleges, especially those with high graduation rates and generous instructional resources. Also, their freshman grades have been found as good as the control students', despite the fact that the control students attend less selective colleges and so rival with peers whose incoming preparation is substantially inferior. Subsequently, benefit-to-cost ratios for the ECO-C intervention are extremely high, even under the most conservative assumptions.

Pengpid et al. (2015) have determined the aim of this study to estimates of the prevalence and social correlates of physical inactivity among university students in 23 low-, middle-, and high-income countries. The International Physical Activity Questionnaire (IPAQ) was usaged to collect data from 17,928 undergraduate university students (mean age 20.8, SD = 2.8) from 24 universities in 23 countries. As a result, in multivariate logistic regression, older age (22-30 years), studying in a low- or lower-middle-income country, skipping breakfast, and deficit of social support have associated with physical inactivity. In men, being overweight or obese, being underweight, not refraining fat and cholesterol, not having vigorous depression symptoms, low beliefs in the health benefits of physical activity, low personal control and knowledge of exercise-heart link, and in women, not trying to eat fibre, low personal mastery and medium personal control were

additionally have associated with physical inactivity. Besides, four in each ten students are physically inactive.

Pengpid and Peltzer (2018) have conducted this study to estimate the prevalence of physical inactivity among university students in 23 low, middle, and high-income countries and their social relations. In a cross-sectional study using anonymous questionnaires, 15,122 university students have evaluated. These students have assessed using the International Physical Activity Questionnaire (short version) and sociodemographic, health status, health behavior, and anthropometric measurements. The result has been found only partial benefits of VPA regarding the well-being of university students. In multivariate logistic regression, advanced age (22-30 years), education, skipping breakfast, and lack of social support have associated with physical inactivity in a low- or middle-income country. Being underweight, overweight or obese in men, not avoiding fat and cholesterol, lack of severe depression symptoms, low belief in the health benefits of physical activity, low self-control and knowledge of exercise-heart connection, and not trying in women fiber diet, low self-control, and moderate personal control has also associated with physical inactivity. Four out of ten students have been physically inactive and require strategic intervention from relevant professionals in higher education institutions.

3. MATERIAL AND METHODS

This study has been carried out to examine the factors that make up the monthly income of university students. For this purpose, the composition data of the monthly income of the students in the selected European countries have been taken from the Euro student site (Eurostudent, 2017). Since the current data belong to 2017, the data for 2017 has been used. A data set has been created by selecting countries that use their own currency in Europe. These selected countries have been Albania (AL), Czech Republic (CZ), Denmark (DK), Georgia (GE), Croatia (HR), Hungary (HU), Switzerland (CH), Iceland (IS), Norway (NO), Poland (PL), Romania (RO), Serbia (RS), Sweden (SE) and Turkey (TR). PCA, which is one of the methods of exploratory factor analysis (EFA), has been applied.

3.1. Factor Analysis

Factor analysis is a statistical method has based on Charles Spearman’s article ‘General Intelligence’ Objectively Determined and Measured published in 1904 (Spearman, 1904). Factor analysis has helped to identify representative variables to be used in other analyzes among a large number of variables. Factors obtained from the original data set are used in a wide range of dependent methods such as regression, correlation and discriminant analysis, and internally dependent cluster analysis (Hair, Anderson, Tatham, and Black, 1998). Factor analysis is examined by the EFA and the confirmatory factor analysis (CFA) methods, which are quite different from each other. CFA has allows the researchers to test the hypothesis that a relationship among the variables and their underlying hidden construct(s) exists. The researchers have uses knowledge of the theory, empirical study, or both, postulates the relationship model a priori, and then tests the hypothesis statistically. On the other hand, the number of constructs and the underlying factor structure are identified with exploratory factor analysis (EFA) (Suhr, 2006). In the context of exploratory factor analysis preferred in this study, two methods commonly are used to extract factors are PCA and principal axis factoring (PAF) (Russell, 2002). For these techniques, the purpose of the PCA method is to extract the maximum variance for each component. Moreover, the aim of PAF is to produce a new correlation matrix in which the factors are perpendicular to each other and reveal the latent structure (Şencan, 2005; Tabachnick and Fidell, 2013). Besides that the suitability of factor analysis should be tested. For this purpose, Kaiser-Mayer-Olkin (KMO) sample suitability test is used. The KMO formula, r_{ij} is being basic correlation coefficient between i.variable and j.variable. a_{ij} is being the partial correlation coefficient between i.variable and j.variable, formulated as follows (Turanlı, Taşpınar, and Bozkır, 2014):

$$KMO = \frac{\sum_{i \neq j} \sum r_{ij}^2}{\sum_{i \neq j} \sum r_{ij}^2 + \sum_{i \neq j} \sum a_{ij}^2} \quad (1)$$

KMO value varies from 0 to 1. KMO values between 0.8 to 1.0 specify the sampling is adequate. KMO values between 0.7 to 0.79 are middle class and values between 0.6 to 0.69 are poor. KMO values less than 0.6 show the sampling is not adequate and the remedial action should be taken. If the value is less than 0.5, the results of the factor analysis undoubtedly won’t be very

appropriate for the analysis of the data. If the sample size is < 300 the average communality of the retained items has to be tested. An average value > 0.6 is acceptable for sample size < 100 , an average value between 0.5 and 0.6 is acceptable for sample sizes between 100 and 200 (Shrestha, 2021).

3.1.1. The PCA

PCA is a multivariate statistical method, which is usually, used in the elimination of the structure of dependence among variables and the degradation of the number of variables in other words the dimension reduction or the data reduction. This analysis is also used as a data preparation technique for other analyses (Tatlıdil, 2002). PCA is a mathematical process that enables the researchers to decrease the number of correlated variables into a lesser number of components, linearly independent of each other, which represents a percentage of the total covariance (Hair, Black, Babin, Anderson, and Tatham, 2006; Meyers, Gamst, and Guarino, 2006). In PCA, the communalities for the measures are set at 1.0 value. This method supposes that all of the variance in a measure is potentially explicable by the components. It should be noted that because the communalities of the measures are set at 1.0 value, a PCA takes out the factors based on the correlations among the measures (i.e., a correlation matrix is analyzed) (Russell, 2002).

3.1.2. Rotated Component Matrix: Varimax Method

Varimax is the most widely used rotation criterion and, like most other such criteria, it inclines to drive at least some of the loadings in each component towards zero. This is not the only probable type of simplicity. A component whose loadings are all roughly equal is easy to interpret but may be avoided by most standard rotation criteria (Jolliffe, Trendafilov and Uddin, 2003). This method is generally used when the items in the component matrix are concentrated on the factor or when there is the factor that is not explained by any item.

A rotated component matrix is found as: Let us suppose that X has been a $(n \times p)$ matrix whose i_{th} row includes the values of the p original variables for the i_{th} (of n) observations. It has assumed, for ease, that each column of X has been centered to have mean zero. Let $Z = XA$, where A has been a matrix $(p \times q)$ whose columns are vectors a_k of loadings associated

with a subset of q components that have been rotated. The i_{th} row of Z consists of the values of the q selected components for the i_{th} observation.

Rotating the q principal components have been achieved by post-multiplying A by a matrix ($q \times q$) T to get the rotated loadings $B = AT$. Values of the rotated components have been elements of the matrix.

$$F = XB = XAT = ZT \quad (2)$$

For orthogonal rotation, T has been an orthogonal matrix. It should have been noted that rotation is carried out with respect to the loadings rather than the observations (Ian T, 1995, 30). For the normalization constraint of equation (3), it has been shown in the appendix that

$$B'B = T'LT \quad (3)$$

and the covariance matrix for F has been

$$\text{var}(F) = T'L^2T \quad (4)$$

where L has been a ($q \times q$) diagonal matrix whose elements have been the eigenvalues corresponding to the q choose components. Neither $B'B$ nor $\text{var}(F)$ has been a diagonal matrix, so the rotated loadings have not been orthogonal, and also the rotated components have not been uncorrelated. For the normalization of equation (5),

$$\begin{aligned} B'B &= I_q \\ \text{var}(F) &= T'LT \end{aligned} \quad (5)$$

where I_q has been the identity matrix ($q \times q$). Thence, the rotated loadings have been orthogonal but the rotated components have been correlated. It should have been noted that the optimal rotation matrix T ought generally to be dissimilar for different normalization constraints. For notational simplicity, T has been used to denote all such rotation matrices, however, the reader should have avoided falsely equating apparently identical expressions for different normalizations. For the normalization of equation (6),

$$B'B = T'L^{-1}T \quad (6)$$

$$var(F) = I_q$$

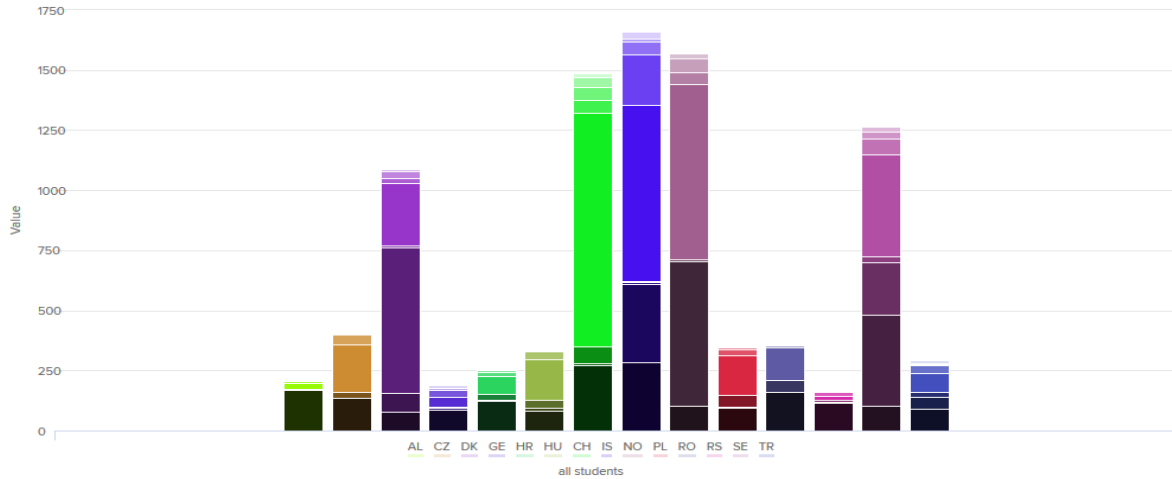
Hereby, the rotated loadings have not been orthogonal but the rotated components have been uncorrelated (Ian T, 1995, 31).

4. ANALYZES

All variables consist of monthly average data on the basis of the Euro exchange rate. Analyzes on this data set have been made in the IBM SPSS Statistics program. Abbreviations for variables in the data set are shown in Table 1. The distribution of the variables used by country is given as in Figure 1. While the students with the most income are in IS, the students with the least income are in RS. The meanings of the coloring in Figure 1 are given in Figure 2.

Table 1. Abbreviations of Variables

Variables	Abbreviations
Family/partner (excl. transfers in-kind)	I-FP
National public student support (repayable)	I-NPSSR
National public student support (non-repayable)	I-NPSSN
Non-country sources	I-NCS
Self-earned income (current paid job)	I-SEIC
Self-earned income (previous paid job)	I-SEIP
Other public sources	I-OPUS
Other private source (non-repayable)	I-OPRN
Other private source (repayable)	I-OPRR



Source: Eurostudent (2017).

Figure 1. Composition of Students' Total Monthly Income

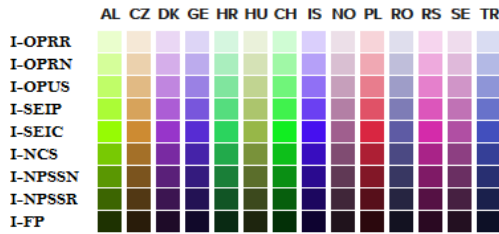


Figure 2. Composition of Students' Total Monthly Income

Initially, descriptive statistics of 9 variables used in the study have given in Table 2. Statistics of mean, median, maximum, and minimum values have given information about the distribution of variables. If an example has given for the I-NPSSR variable; respectively, the maximum value, the minimum value, the median, and the mean have found as 604.500, 0.900, 26.100, and 70.107. According to these values, the median and mean values have been closer to the minimum value, and the graph of the I-NPSSR variable has skewed to the right. Similar comments can be made for other variables in Table 2.

Table 2. Descriptive Statistics

	I-FP	I-NPSSR	I-NPSSN	I-NCS	I-OPRN	I-OPRR	I-OPUS	I-SEIC	I-SEIP
Mean	138.392	70.107	83.023	2.476	10.253	6.069	17.653	276.246	40.676
Median	116.100	26.100	3.700	1.5000	6.700	2.400	6.100	166.900	27.400
Maximum	284.100	604.500	598.500	10.600	42.800	27.800	56.400	971.200	209.200
Minimum	79.900	0.900	0.000	0.000	2.300	0.400	0.000	15.900	1.000

Correlation refers to the coefficient of the relationship between the two variables. It is between -1 and +1 and it is not possible to go beyond these limits. If the correlation coefficient is; If it is equal to 1, there is a 100% positive relationship between these 2 variables, if it is equal to -1, there is a 100% negative relationship. Generally, 50% and below are defined as weak relationships (Spearman, 1904). A correlation matrix explains the relationship between the m variables in which the cross elements in the matrix are equal. The square obtained from the variance-covariance matrix is a symmetrical mxm-sized matrix. Both of these matrices contain similar information, but since the correlation matrix is simpler and more explanatory, it makes it easier to associate variables in the matrix (Horn and Johnson, 1985). Also, this correlation coefficient should be statistically significant. Statistical significance is possible only if the p-value of the correlation coefficient is less than 0.05. In other words, the H_{1a} hypothesis must be rejected.

Correlation coefficients and p-values of variables in this study have given in Table 3. Coefficients with p-values less than 0.05 in the table have been statistically significant. There is a positive relationship of 73.92% between, for instance, I-NPSSN and I-OPUS variables. The p-value of this positive relationship has found to be 0.0039 ($p < 0.05$). The positive relationship of 73.92% found has been statistically significant at the 95% confidence level. Similar interpretations can be made for other correlation coefficients in Table 3.

H_{1a} : The correlation coefficient is not statistically significant.

H_{1b} : The correlation coefficient is statistically significant.

Table 3. Correlations

Variables	I-FP	I-NPSSR	I-NPSSN	I-NCS	I-OPRN	I-OPRR	I-OPUS	I-SEIC	I-SEIP
I-FP	1.0000								

I-NPSSR	-0.224	1.0000							
	0.4619	-----							
I-NPSSN	0.1377	-0.0582	1.0000						
	0.6535	0.8500	-----						
I-NCS	-0.0233	0.7371	0.3694	1.0000					
	0.9397	0.0040*	0.2141	-----					
I-OPRN	0.5854	0.0375	0.3247	0.0094	1.0000				
	0.0355*	0.9031	0.2790	0.9755	-----				
I-OPRR	0.7086	-0.0794	0.3665	0.3656	0.5112	1.0000			
	0.0067*	0.7964	0.2180	0.2192	0.0741	-----			
I-OPUS	0.5809	0.1451	0.7392	0.3937	0.7920	0.6602	1.0000		
	0.0373*	0.6362	0.0039*	0.1831	0.0012*	0.0140*	-----		
I-SEIC	0.6689	0.0268	0.6001	0.2204	0.8932	0.6562	0.9533	1.0000	
	0.0124*	0.9306	0.0301*	0.4692	0.0001*	0.0148*	0.0001*	-----	
I-SEIP	0.6498	-0.1268	0.5262	0.4282	0.3243	0.8835	0.6430	0.6177	1.0000
	0.0162*	0.6796	0.0647	0.1443	0.2797	0.0001*	0.0177*	0.0245*	-----

Note: Those marked with “*” are statistically significant at 95% confidence level.

In factor analysis, the KMO test (Kaiser, 1970) was applied to test the adequacy of the sample size. The KMO test and Bartlett's test are suitable tests to measure sampling adequacy. If the p-value of this test is less than 0.05, there is no sphericity among variables. This means that the variables are correlated very adequate to provide a convenient basis for factor analysis (Barrett and Morgan Jr, 2005). The value found as a result of the KMO test should be greater than 0.50.

According to the result in Table 4, the sample adequacy value was found to be 0.610. So it is appropriate to use this sample.

Besides that the variables used in the factor analysis must provide the normality assumption. For this purpose, Bartlett's test (Bartlett, 1950) has been applied. Since the p-value of Bartlett's test has given in Table 4 is less than 0.05 level, the H_{2a} hypothesis has been rejected and the normality assumption is provided ($p < 0.05$). The correlation matrix has been not a unit matrix and there is no sphericity. Thus, based on the results, it is appropriate to proceed with factor analysis to examine factors that affect the composition of students' total monthly income.

H_{2a} : The correlation matrix is the unit matrix. There is a sphericity.

H_{2b} : The correlation matrix is not the unit matrix. There isn't a sphericity.

Table 4. The KMO and Bartlett's Tests

Tests	Results	
KMO Measure of Sampling Adequacy	0.610	
Bartlett's Test of Sphericity	Approx. Chi-Square	95.416
	Df	28
	p-value	0.001

Explanation rates of the important parts of the variables have given in Table 5. The variable with the highest disclosure rate is the strongest variable in factor analysis. These disclosure rates have given in the extraction column. According to the results in Table 5, the variable I-SEIC has been the strongest variable. 98.1% of the I-SEIC variable has been explained by the key components that matter. However, since the Extraction value of the I-NPSSR variable is $0.456 < 0.50$, it is not suitable to use. Therefore, the variable I-NPSSR has been omitted from factor analysis.

Table 5. Communalities

Variables	Initial	Extraction
I-FP	1.000	0.726
I-NPSSR	1.000	0.456*
I-NPSSN	1.000	0.865
I-NCS	1.000	0.975
I-SEIC	1.000	0.981
I-SEIP	1.000	0.967
I-OPUS	1.000	0.964
I-OPRN	1.000	0.936
I-OPRR	1.000	0.868

The extraction values of the 8 variables remaining after removing the I-NPSSR variable have given in Table 6. The variable with the highest disclosure rate is the strongest variable in factor analysis. These disclosure rates have given in the extraction column. According to the results in

Table 6. Communalities

Variables	Initial	Extraction
I-FP	1.000	0.759
I-NPSSN	1.000	0.919
I-NCS	1.000	0.975
I-SEIC	1.000	0.968
I-SEIP	1.000	0.958
I-OPUS	1.000	0.916
I-OPRN	1.000	0.953
I-OPRR	1.000	0.904

Table 5, the variable I-NCS has been the strongest variable. 97.5% of the I-NCS variable is explained by the key components that matter.

According to the total variance explained results in Table 7, the eigenvalue of the 1st factor is 4.490, the eigenvalue of the 2nd factor is 1.810 and the eigenvalue of the 3rd factor is 1.052. The variance rates explained by these factors within the total variance have been 56.129%, 22.620%, and 13.154%, respectively. Factors with an eigenvalue greater than 1.000 determine the number of basic components. Figure 3 shows that the eigenvalues for these 3 factors are greater than 1. In this factor analysis, the dimension has been determined as 3 and these dimensions explain 91.902% of the total variance.

Table 7. Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of	
	Total	% of Variance	Cumulative %	Total	% of Variance
1	4.490	56.129	56.129	4.490	56.129
2	1.810	22.620	78.749	1.810	22.620
3	1.052	13.154	91.902	1.052	13.154
4	0.360	4.499	96.401		
5	0.183	2.284	98.685		
6	0.053	0.664	99.349		
7	0.041	0.516	99.865		
8	0.011	0.135	100.000		

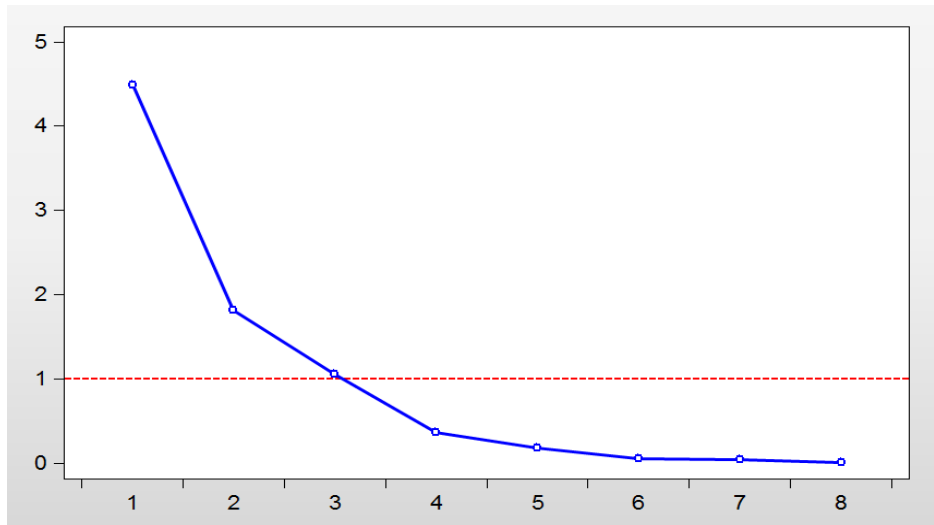


Figure 3. Scree Plot (Eigenvalues)

The component matrix for 3 factors has given in Table 8, where Factor 3 has not to contain any items. For this reason, rotation must be carried out on the factor dimension, so that the rotated component matrix has obtained. This rotation will be done by the Varimax method, where varimax rotation is chosen because it is easier to analyze in theory.

Table 8. Component Matrix

Variables	Component		
	1	2	3
I-FP	0.741	-0.316	-0.108
I-NPSSN	0.031	0.876	0.262
I-NCS	0.355	0.876	-0.203
I-SEIC	0.938	-0.068	0.297
I-SEIP	0.810	0.014	-0.538
I-OPUS	0.919	0.111	0.223
I-OPRN	0.783	-0.170	0.538
I-OPRR	0.855	-0.024	-0.380

The rotated component matrix utilizing Varimax rotation has given in Table 9. Varimax with Kaiser normalization has been chosen as the rotation method. After performing the varimax rotation method, Factor 1 has comprised of three items with factor loadings ranging from 0.815 to 0.968. Factor 2 has comprised of three items with factor loadings ranging from 0.638 to 0.948. Factor 3 has comprised of two items with factor loadings of 0.919 and 0.926. Factor equations created by using Table 9 have given below.

$$F_1 = 0.894*I\text{-SEIC} + 0.815*I\text{-OPUS} + 0.968*I\text{-OPRN} \quad (7)$$

$$F_2 = 0.638*I\text{-FP} + 0.948*I\text{-SEIP} + 0.873*I\text{-OPRR}$$

$$F_3 = 0.926*I\text{-NPSSN} + 0.919*I\text{-NCS}$$

Table 9. Rotated Component Matrix

Variables	Component		
	1	2	3
I-FP	0.534	0.638	-0.259
I-NPSSN	0.094	-0.228	0.926
I-NCS	-0.002	0.362	0.919
I-SEIC	0.894	0.405	0.073
I-SEIP	0.226	0.948	0.096
I-OPUS	0.815	0.438	0.244
I-OPRN	0.968	0.123	-0.035
I-OPRR	0.370	0.873	0.072

5. CONCLUSION

This statistical study has been carried out to examine the income types that make up the monthly income of students in selected countries, by using factor analysis. Firstly, descriptive

statistics of variables are given. Thus, the general structure of the variables has been understood. Then, the association ratios between the variables have been interpreted with a correlation matrix. The KMO and Bartlett's Tests have been applied to determine the appropriateness of factor analysis and it has been determined that factor analysis is suitable to examine the factors affecting the total monthly income composition of the students.

It has been observed that the I-NPSSR variable has not sufficiently explained by calculating the extraction values. In other words, since the extraction value of this variable is less than 0.5, this variable has been removed from the factor analysis. Then, the extraction values have been recalculated and there was no extraction problem with the variables has been determined. Owing to I-NPSSR variable could not be explained sufficiently, 8 variables have used in factor analysis.

According to the total variance explained results, 3 factors have created for 8 variables. The total variance explanation ratio of these 3 factors is 91.902%. Finally, the rotated component matrix with the varimax method has been applied in order to determine the significant variables that make up the factors. I-SEIC, I-OPUS, and I-OPRN variables for factor 1 have consisted of factor loads of 0.894, 0.815, and 0.968, respectively. This factor can be named as " income that students can save" in terms of the variables it contains. I-FP, I-SEIP, and I-OPRR variables for factor 2 have consisted of factor loads of 0.638, 0.948, and 0.873, respectively. This factor can be named as "students' main sources of income" in terms of the variables it contains. Lastly, I-NPSSN and I-NCS variables for factor 3 have consisted of factor loads of 0.926 and 0.919, respectively. This factor can be named as "special incomes of students" in terms of the variables it includes. As a result, factor analysis has been applied for 8 variables consisting of composition data of students' monthly income, and a 3-factor result has been obtained. These factors have helped to group and interpret the income sources of university students statistically.

ETHICAL DECLARATION

In the writing process of the study titled “A Factor Analysis Application On Income Sources of University Students in European Countries”, there were followed the scientific, ethical

and the citation rules; was not made any falsification on the collected data and this study was not sent to any other academic media for evaluation.

REFERENCES

- Bartlett, M.C. (1950), Tests of Significance in Factor Analysis, *British Journal of Psychology*, 3(2), 77-85.
- Christie, H., Munro, M. and Rettig, H. (2001), Making Ends Meet: Student Incomes and Debt, *Studies in Higher Education*, 26(3), 363-383.
- Correddu, F., Cesarani, A., Dimauro, C., Gaspa, G and Macciotta, NPP. (2021), Principal Component and Multivariate Factor Analysis of Detailed Sheep Milk Fatty Acid Profile, *J Dairy Sci.*, 2021 Jan 27, S0022-0302(21)00083-7.
- Dickinson, E.R. and Adelson, J.L. (2016), Choosing Among Multiple Achievement Measures: Applying Multitrait–Multimethod Confirmatory Factor Analysis to State Assessment, ACT, and Student GPA, *Data Journal of Advanced Academics*, 27(1), 4-22.
- Eurostudent (2017), *Composition of Students' Total Monthly Income (Excl. Transfers in Kind)*, <http://database.eurostudent.eu/>, Accessed: 27.06.2020.
- Feng, A.L., Wesely, N.C., Hoehle, L.P., Phillips, K.M., Yamasaki, A., Campbell, A.P., Gregorio, L.L., Killeen, T.E., Caradonna, D.S., Meier, J.C., Gray, S.T. and Sedaghat, A.R. (2017), A Validated Model for the 22-item Sino-Nasal Outcome Test Subdomain Structure in Chronic Rhinosinusitis, *Int Forum Allergy Rhinol*, 7(12), 1140-1148.
- Gwosc, C. and Schwarzenberger, A. (2009), Public/Private Cost-Sharing in Higher Education: an In-Depth Look at the German System Using a Comparative Study, *Journal of Higher Education Policy and Management*, 31(3), 239-249.
- Hair, J.F., Anderson, R.E., Tatham, R.L. and Black, W.C. (1998), *Multivariate Data Analysis*, Prentice Hall, New Jersey.

- Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E. and Tatham, R.L. (2006), *Multivariate Data Analysis (6th Edition)*. Upper Saddle River: Pearson Prentice Hall, New Jersey.
- Horn, R.A. and Johnson, C.R. (1985), *Topics in Matrix Analysis*, Cambridge University Press, Cambridge.
- Hoxby, C. and Turner, S. (2013), Expanding College Opportunities for High-Achieving, Low Income Students, *Stanford Institute For Economic Policy Research Discussion Paper No. 12-014*, 1-37.
- Ian T.J. (1995), Rotation of Principal Components: Choice of Normalization Constraints, *Journal of Applied Statistics*, 22(1), 29-35.
- Jolliffe, I., Trendafilov, N. and Uddin, M. (2003), A Modified Principal Component Technique Based on the LASSO. *Journal of Computational and Graphical Statistics*, 12(3), 531-547.
- Kaiser, H.F. (1970), A Second-Generation Little Jiffy, *Psychometrika*, 35(4), 401-415.
- Karadeniz, Ş. (2012), Development of Student Satisfaction on Blended Learning Questionnaire, *Electronic Journal of Social Sciences*, 11(40), 161-172.
- Karataş, H. and Fer, S. (2011), CIPP Evaluation Model Scale: Development, Reliability and Validity, *Procedia Social and Behavioral Sciences*, 15, 592-599.
- Leech, N.L., Barrett, K.C. and Morgan, G.A. (2005), *SPSS for Intermediate Statistics: Use and Interpretation*, Psychology Press, London.
- Martínez-García M.A., Perpiñá-Tordera M., Soler-Cataluña J.J., Román-Sánchez P., Lloris-Bayo A. and González-Molina A. (2007), Dissociation of Lung Function, Dyspnea Ratings and Pulmonary Extension in Bronchiectasi, *Respir Med.*, 101(11), 2248-2253.
- Meyers, L.S., Gamst, G. and Guarino, A.J. (2006), *Applied Multivariate Research: Design and Interpretation*, Sage, California.
- Muthén, B. (1991), Multilevel Factor Analysis of Class and Student Achievement Components, *Journal of Educational Measurement*, 28(4), 338-354.

- O'Farrell, S.L. and Morrison, G.M. (2003), A Factor Analysis Exploring School Bonding and Related Constructs Among Upper Elementary Students, *The California School Psychologist*, 8(1), 53-72.
- Pengpid, S., Peltzer, K., Kassean, H.K., Tsala Tsala, J.P., Sychareun, V. and Müller-Riemenschneider, F. (2015), Physical Inactivity and Associated Factors Among University Students in 23 Low-, Middle- and High-Income Countries, *International Journal of Public Health*, 60(5), 539-549.
- Pengpid, S. and Peltzer, K. (2018), Vigorous Physical Activity, Perceived Stress, Sleep and Mental Health Among University Students From 23 Low- and Middle-Income Countries, *International Journal of Adolescent Medicine and Health*, 32(2), 1-7.
- Russell, D.W. (2002), In Search of Underlying Dimensions: The Use (and Abuse) of Factor Analysis in Personality and Social Psychology Bulletin, *Personality and Social Psychology Bulletin*, 28(12), 1629-1646.
- Shrestha, N. (2021), Factor Analysis as a Tool for Survey Analysis, *American Journal of Applied Mathematics and Statistics*, 9(1), 4-11.
- Spearman, C. (1904), General Intelligence Objectively Determined and Measured, *The American Journal of Psychology*, 15(2), 201-293.
- Stufflebeam, D.L. and Webster, W.J. (1988), *Evaluation as an Administrative Function*. In N. Boyan (Ed.), *Handbook of Research on Educational Administration* (pp. 569-601), White Plains, New York.
- Suhr, D.D. (2006), *Exploratory or Confirmatory Factor Analysis?* Presented: San Francisco, CA, SAS Users Group International Conference (SUGI31).
- Şencan, H. (2005), *Sosyal ve Davranışsal Ölçümlerde Güvenilirlik ve Geçerlilik (Birinci Baskı)*, Seçkin Yayınları, Ankara.
- Tabachnick, B.G. and Fidell, L.S. (2013), *Using Multivariate Statistics (6th Edition)*. Pearson Education Limited, London.

Takenaka, S., Kashii, M., Iwasaki, M., Makino, T., Sakai, Y. and Kaito, T. (2021), Risk Factor Analysis of Surgery-Related Complications in Primary Cervical Spine Surgery for Degenerative Diseases Using a Surgeon-Maintained Database, *Bone Joint J.*, 103-B(1), 157-163.

Tatlıdil, H. (2002), *Uygulamalı Çok Değişkenli İstatistiksel Analiz*, Akademi Matbaası, Ankara.

Turanlı, M., Taşpınar-Cengiz, D. and Bozkır, Ö. (2014), Faktör Analizi ile Üniversiteye Giriş Sınavlarındaki Başarı Durumuna Göre İllerin Sıralanması, *Istanbul University Econometrics and Statistics e-Journal*, 0(17), 45-68.

Appendix

Table 10. Dataset 1/2

Countries	Family/Partner (Excl. Transfers in Kind)	National Public Student Support (Repayable)	National Public Student Support (Non-repayable)	Non-country Sources
AL	170.3	3.1	0.9	0.2
CZ	134.1	0	26.3	1.7
DK	79.9	75.8	604.5	10.6
GE	85.2	0	8.1	3.7
HR	124.1	1.9	26.1	1.5
HU	81.3	12.7	32.2	2.1
CH	269.9	8.6	73.1	0
IS	284.1	323.5	8.2	7
NO	103.8	598.5	8	3.4
PL	95.8	3	47.6	0.2
RO	162	0	47.2	1.2
RS	116.1	3.7	9.7	0.2
TR	92.5	48.5	19.5	0.4

Table 11. Dataset 2/2

Countries	Self-earned Income (Current Paid Job)	Self-earned Income (Previous Paid Job)	Other Public Sources	Other Private Source (Non- repayable)	Other Private Source (Repayable)
AL	23.7	6.2	1.6	4.9	0.4
CZ	197.3	39	0	6.7	0.9
DK	259	22.2	26.9	9.5	4
GE	44.1	27.4	2.3	7.1	10.4
HR	72.1	17.5	6.1	2.4	1.1
HU	168	32.4	5.2	5.6	0.9
CH	971.2	52.7	53.4	42.8	14.8
IS	731.4	209.2	53.7	14.4	27.8
NO	725.5	52.9	56.4	20.9	4.8
PL	166.9	22.3	8.8	4.5	2.1
RO	136.4	1	8	2.3	2.4
RS	15.9	15.2	3.7	5	1
TR	79.7	30.8	3.4	7.2	8.3

Source: Eurostudent (2017),