IRSTI 27.41.15

G. Assanbayeva[1], Sh. Kadyrov[2]
[1,2]Suleyman Demirel University, Kaskelen, Kazakhstan

PRINCIPAL COMPONENT ANALYSIS AND A MULTILINGUAL CONSTRUCT TO DETERMINE THE UNDERGRADUATE MAJOR SELECTION FACTORS

Abstract. In this article, we review mathematics behind well-known Principal Component Analysis from Linear Algebra implemented in various applied fields. As an application, we develop a construct to measure factors that affect college students in their major selection. This is a multilingual construct given in three languages, namely Kazakh, Russian, and English. To this end, we prepare a survey consisting of 27 Likert scale items in three languages and it is conducted among 314 undergraduate students in Kazakhstan. For dimensionality reduction, Principal Component Analysis is carried in python programming language which resulted in 9 major scales with only 22 elements. The overall reliability of the test is calculated to be 0,856. The nine scales are the effect of Uniform National Testing, state grant affect, personal interest affect, skills affect, occupation salary affect, teacher affect, external affect, university cost affect, parent's affect.

Keywords: Principal Component Analysis, Factor Analysis, Varimax rotation, Reliability, Major selection, Construct.

***

Аңдатпа. Бұл мақалада біз әртүрлі қолданбалы салаларда енгізілген сызықтық алгебрадан белгілі негізгі компонентті талдаудың артындағы математиканы қарастырамыз. Бағдарлама ретінде біз университет студенттеріне негізгі мамандық таңдау кезінде әсер ететін факторларды өлшейтін сауалнама жасаймыз. Бұл үш тілде, атап айтқанда қазақ, орыс және ағылшын тілдерінде берілген көптілді сауалнама. Осы мақсатта авторлар үш тілде 27 Likert шкаласынан тұратын сауалнама дайындады және ол Қазақстандағы 314 студенттер арасында өткізілді. Өлшемділікті төмендету үшін негізгі компоненттік талдау Python арқылы есептелінді, нәтижесінде 22 негізгі элементтерден тұратын 9 ірі компоненттер алынды. Тесттің жалпы сенімділігі 0,856 құрайды. Тоғыз шкалалар: ұлттық тестілеу нәтижесі әсері, мемлекеттік грант нәтижесі,жеке қызығушылық әсері,өз қабілетінің әсері, мамандықтың жалақысы әсері, мұғалімнің әсері, сыртқы әсер, университеттің құнының әсері,ата-ананың әсері.

Түйін сөздер: Негізгі компоненттік әдіс, факторлық әдіс, варимакс айналымы, сенімділік, мамандық таңдау, құрастыру.

***

Аннотация. В этой статье мы рассматриваем математику, лежащую в основе хорошо известного анализа главных компонентов из линейной алгебры, реализуемой в различных прикладных областях. В качестве приложения мы разрабатываем конструкцию для измерения факторов, влияющих на студентов университета в их главном выборе. Это многоязычная конструкция, представленная на трех языках, а именно на казахском, русском и английском. С этой целью авторы подготовили опрос, состоящий из 27 пунктов шкалы Лайкерта на трех языках, и он был проведен среди 314 студентов бакалавриата Казахстана. Для уменьшения размерностибыл проведен анализ главных компонент в python, который привел к 9 основным масштабам с только 22 элементами. Общая достоверность испытания, по расчетам, составляет 0,856. Девять шкал: влияние единого национального тестирования, влияние личного интереса, влияние государственного гранта, влияние заработной платы по профессии, влияние навыков, влияние преподавателя, влияние внешних факторов, влияние стоимости университета, влияние родителей.

Ключевые слова: анализ главных компонент, факторный анализ, Варимакс - вращение, достоверность, выбор специализации, построение.

1. Introduction

Linear algebra is a branch of mathematics that deals with system of linear equations, vector spaces, linear maps and their properties. Matrices are one of the building blocks of linear algebra. A numerical data consisting of m cases and n variable entries for each case can be thought of as $m \times n$ matrix. This representation enables us to carry various manipulations available to us from linear algebra and interpret the results. When n is large, it often becomes difficult to derive meaningful conclusions from the data. Principal Component Analysis (PCA) is one of the widely used techniques from linear algebra that helps with dimensionality reduction and makes it possible to extract hidden features of the data (Sanguansat, 2012). Even though this is a century old method invented by K. Pearson (Pearson, 1901), in its original form and in improved versions it is still being used nowadays for handling various large datasets. Some of the research areas where PCA is used include signal processing (Turan, et al., 2018), genetics (Li, et al., 2019), quantitative finance (Avellaneda, et al., 2010), neuroscience (Subasi, et al., 2010), and questionnaire development (Brown, 2010).

Our goal in this article is to review mathematics behind this powerful tool and show how it can be applied in developing a construct that measures factors influencing students' major selection. There are various questionnaires are used in the literature to analyse factors related to major selection. Factors such as Interest in major, Peer pressure, Family pressure, Academic ability, Major's reputation, Job availability, Job salary, Major's prestige, Public sector job, Private sector job were analyzed in (Aldosary, et al., 1996) from 447 students of King Fahd university and Job availability, Salary, social status and prestige were found to be the main affecting factors in that order. Another study was carried with 111 participants to investigate college students' academic major declaration (Galotti, 1999). An exploratory factor analysis was carried by Sarwar et al, (2015) to analyse the variables affecting the specialization selection of 300 business graduates in Lahore resulting 6 main factors: academic factors, social capital factors, future prospect factors, human capital factors, market demand factors and finally job prospect factors. This 31-item construct is calculated to have high reliability of 0.845. Another study was carried (Fizer, 2013) at the University of Tennessee, Martin to determine the variables that influence agriculture students' choices in deciding their career path. The findings show that the main variable (22%) is family influence followed by a factor "a career that is personally rewarding" (21%).

In the next section, we provide the methodology used to develop our construct. More specifically, we will brief on the participants and the questionnaire conducted, and review the background information needed to understand PCA methodology. The section 3 contains the application of PCA to extract main factors via dimensionality reduction. The paper ends with discussion and conclusion section where we interpret our findings and highlight the possible future research directions.

2. Methodology

1.1. Participants and questionnaire

The main purpose of this study is an attempt to reduce the number of factors and define main aspects of the resulted construct. The survey is prepared by using various sources like (Singh Swapnika), (Sarwar, et al., 2015) and adapted to the context of Kazakhstan. An online survey questionnaire is consisting 27 questions. The survey is prepared languages,namely Kazakh, Russian, and English and send out to students from 16 universities within the country and received 314 students participants. Students took as a sample through non probability convenience sampling technique.First part of questionnaire is directed to collect demographic data and items related to major selection were

in the second part of the survey. Table 1 provides demographic information on participants. The number of respondents in the Kazakh language is 109,in Russian 93 and in English 112.

Table 1

Language	**Age group**	**Gender**	**University GPA**
Kazakh	16-18y	Male	3.5-4.0
34.7%	18.5%	45.9%	43.3%
Russian	19-21y	Female	2.5-3.4
29.6%	42.7%	54.1%	46.2%
English	22-24y		1.5-2.4
35.7%	25.1%		9.55%
	24-more		1.0-1.4
	13.7%		0,95%

Most respondents are between 19-21 years old students. They took 42.7% (134 students) from total. However, 24 years or older than 24 years are 13.7% from 314 students. Students of the engineering speciality took part in the survey by 21.3% (67students) and it is the highest frequency of respondents. Then comes students majoring in pedagogy and mathematics with 20% (63 students). The minimum size of the surveyed participants are attended by journalists and fine and applied art. They took only 1% (3 students from each).

Figure 1. Major distribution of participants

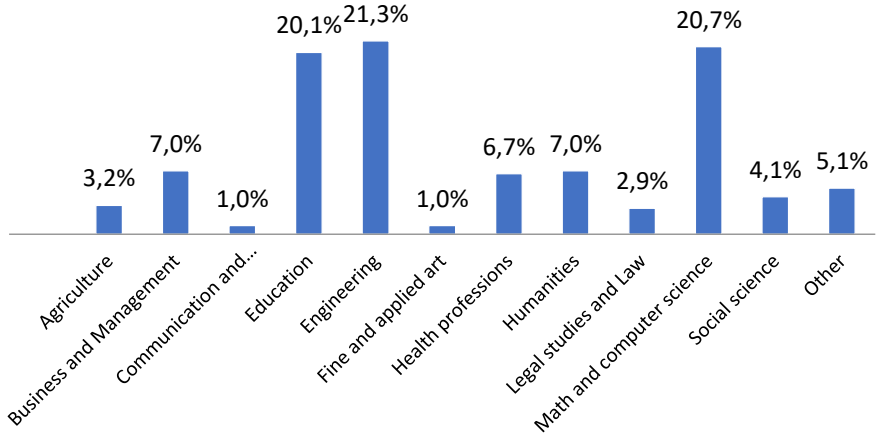
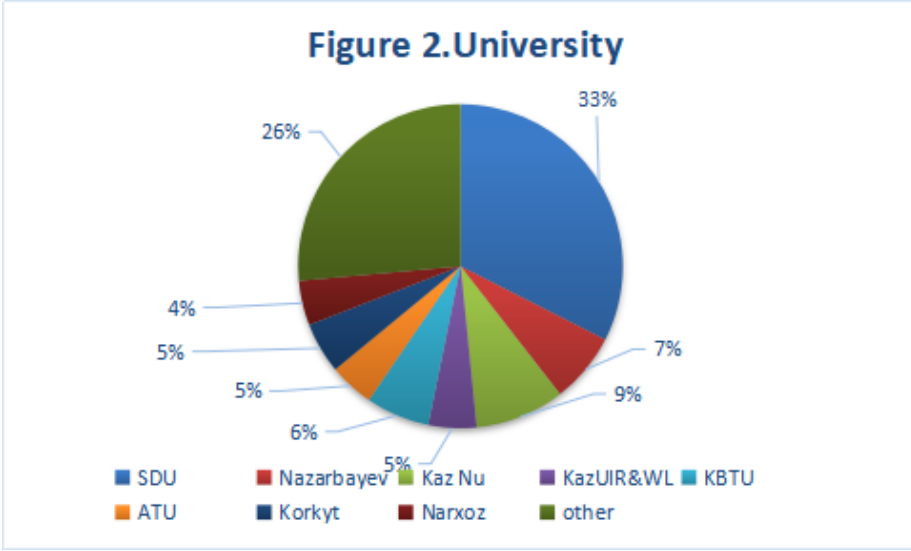


Figure 2 provides information about the number of students surveyed by more than 16 universities. The majority of respondents are students of SDU university.Number of participants from SDU is 197.In the second is KazNu with 28 participants, and the NU with 22. Small number of students from different universities are counted as a one group with 26% from total.



1.2. Instruments

Participants answered to questions by online form. Responses are evaluated using the Likert Scale. Items are graded from 1 to 5 points. Accordingly, 1-'strongly disagree', 2-'disagree', 3-'neutral',4-'agree',5-'strongly disagree'. The answers are translated into Kazakh and Russian languages accordingly with this grading system.

In order to check the internal consistency of scale items, Chronbach alpha reliability analysis is performed.

For dimensionality reduction factor analysis through Principal Component Analysis is implemented with Varimax rotation. The Kaiser-Meyer Olkin sampling Adequacy index is a figure showing the proportion of variation in your variables that could be caused by underlying factors.

Scree plot is used to plot eigenvalues of a data and to determine the number of factors of principal components. By using the rotation methods such as VARIMAX, we have additional tools which make easier the interpretation of the factors, and which can thus improve the relevance of the results.

1.3. Principal Component Analysis

Principal component analysis (PCA) is a technique that is useful for the compression and classification of data variables. The goal is to reduce the

dimensionality of a data set (sample) by grouping the intercorrelated variables , possibly obtaining smaller than the original set of variables, that nonetheless retains most of the sample's information. A PCA is an application of linear algebra where one rotates and shifts the coordinate axes to obtain more suitable representation of data helpful for feature extraction one that presents important information. PCA requires a small background of linear algebra. So, we now discuss some basic concepts of linear algebra, in particular algebra (Lindsay, 2002) used to apply in PCA.

Basic Linear Algebra Review:

Eigenvectors and eigenvalues are important properties of matrices that are fundamental to PCA.

Definition 1. Let A be an $n \times n$ real matrix. A complex number λ is called an *eigenvalue* of a matrix A if there exists an n *dimensional* non-zero complex vector $\vec{x}$, called an *eigenvector,* such that

$$A\vec{x} = \lambda\vec{x}.$$

To determine eigenvalues one needs to solve the characteristic equation:

$$D(\lambda) = \det(A - \lambda I)$$

By solving the equation for λ, we will have eigenvalues $\lambda_1, \lambda_2 \ldots$ .By substituting $\lambda's$ into the vector equation, we can obtain eigenvectors.

Eigenvectors belonging to different eigenvalues are easily seen to be linearly independent. If a matrix is symmetric then in fact distinct eigenvectors are mutually orthogonal. We now make these notions more clearer. Orthogonality is important because it means that you can express the data in terms of these perpendicular eigenvectors, instead of expressing them in terms of the x and y axes. We will be doing this later (Lindsay, 2002).

Definition 2 : A $m \times n$ matrix $A = [a_1, a_2 \ldots\ldots a_n]$ is said to be orthogonal:

$$a_i^T a_j = \begin{cases} 1 & if\ i = j \\ 0 & otherwise \end{cases}$$

where each $a_{i,} i = 1,2,3 \ldots, n$ is a column vector of m rows.

Theorem 1: The inverse of an orthogonal matrix is it's transpose (Shlens, 2014).

Definition 3 : A $m \times m$ square matrix A is said to be symmetric if $A_{ij} = A_{ji}$, i.e., row index and column index are interchangeable: $A^T = A$.

Theorem 2: For any $m \times n$ matrix of real numbers A ,*m* x *m* matrix $A^T A$ and the *n* x *n* matrix AA^T are symmetric (Shlens, 2014).

Proof :

Let's take the transposes of AA^T .We apply properties of transpose operation. Then:

$$(AA^T)^T = A^{TT}A^T = AA^T$$

We repeat this analysis for A^TA :

$$(A^TA)^T = A^TA^{TT} = A^TA$$

Definition 4: A matrix A is said to be diagonalizable if there exists some E such that $A = EDE^T$, where *D* is a diagonal matrix and *E* is some special matrix that diagonalizes A. Additionally, if *E* is orthogonal, then *A* is said to be orthogonally diagonalizable .

Theorem 3: A matrix is symmetric if it is orthogonally diagonalizable (Shlens, 2014).

Proof: Suppose *A* is orthogonally diagonalizable. Let us compute A^T.

$$A^T = (EDE^T)^T = E^{TT}D^TE^T = EDE^T = A.$$

Hence, if A is orthogonally diagonalizable, it must also be symmetric

Theorem 4: If A is symmetric (meaning $A^T = A$), then A is orthogonally diagonalizable and has only real eigenvalues. In other words, there exist real numbers $\lambda_1 \ldots \lambda_n$(the eigenvalues) and orthogonal, non-zero real vectors $\overrightarrow{v_1} \ldots \ldots \overrightarrow{v_n}$(the eigenvectors) such that for each $i = 1,2.....n.$ (Jauregui, 2012):

$$A\overrightarrow{v_i} = \lambda_i \overrightarrow{v_i}$$

Let A be a square n × n symmetric matrix with associated eigenvectors $\{e_i\}_{i=1}^{n}$ and $E = [e_1 \ldots\ldots\ldots e_n]$.

Then:

Theorem 5: A symmetric matrix *A* is diagonalized by a matrix of its orthonormal eigenvectors (Shlens, 2014).

Proof: This theorem asserts that there exists a diagonal matrix *D* such that $A = EDE^T$. Let *A* be any matrix, not necessarily symmetric, and let it have independent eigenvectors e_i (i.e. no degeneracy).

$$AE = [Ae_1 \ldots\ldots\ldots Ae_n] = [\lambda_1 e_1 \ldots\ldots\ldots \lambda_n n] = ED.$$

Since $AE = ED,$ it follows that $A = EDE^{-1}$.

Calculation of PCA:

Step 1. Get some data

Suppose we take n individuals, and on each of them we measure the same m variables. In this case, we say that we have n samples of m-dimensional data. For the $i-th$ individual, record the m measurements as a vector $\overrightarrow{x_i}$ belonging to R^m (Jauregui, 2012).

Step 2. Subtract the mean

Using notation from Step 1, we can store the mean of all m variables as a single vector in R^m:

$$\bar{\mu} = \frac{1}{n}(\overrightarrow{x_1} + \cdots + \overrightarrow{x_n})$$

For PCA to work properly, you should standardize the dataset.The mean subtracted is the average across each dimension. It is common 're-centering' the data so that the mean is zero. It is working by subtracting mean $\bar{\mu}$ from each sample vector $\overrightarrow{x_i}$. Let A be the *m* x *n* matrix whose *i-th* column is $\overrightarrow{x_i} - \bar{\mu}$ (Jauregui, 2012):

$$A = [\overrightarrow{x_1} - \bar{\mu}] \ldots [\overrightarrow{x_i} - \bar{\mu}]$$

Then define covariance matrix.

Step 3. Calculate the covariance matrix

In mathematics and statistics, covariance is a measure of the relationship between two random variables.Covariance is a measure of how changes in one variable are associated with changes in a second variable. Covariance is always measured between 2 dimensions. If you calculate the covariance between one dimension and itself, you get the variance. So, if you had a 3-dimensional data set (x, y, z), then you could measure the $m \times n$ covariance between the *x* and *y* dimensions, the *x* and *z* dimensions, and the *y* and *z* dimensions. In fact, for an n-dimensional data set, you can calculate $\frac{\mathrm{n!}}{(\mathrm{n}-2)!2}$ different covariance values (Lindsay, 2002). Formula for covariance matrix *S* (which will be *m* x *m*) (Jauregui, 2012):

$$S = \frac{1}{n-1} AA^T$$

By Theorem 2 , our S is symmetric. Since S is a symmetric matrix, it can be orthogonally diagonalized by Theorem 3. This connection between statistics and linear algebra is the beginning of PCA. The other point is that since $\mathrm{cov}(a,b) = \mathrm{cov}(b,a)$, the matrix is symmetrical about the main diagonal (Lindsay, 2002).

Step 4. Calculate the eigenvectors and eigenvalues of the covariance matrix

Since the covariance matrix is square and symmetric, we can calculate by Theorem 4 the eigenvectors and eigenvalues for this matrix. This is very important for PCA.Apply the Theorem 4 , and let $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$ be the eigenvalues of S (in decreasing order) with corresponding $eig_1 \ldots \ldots \ldots \ldots \ldots eig_n$ orthonormal eigenvectors by theorem 5. These eigenvectors are called the *principal components* of the data set (Jauregui, 2012).

Step 5. Choosing components and forming a feature vector

If you originally have *n* dimensions in your data, and so you calculate *n* eigenvectors and eigenvalues, and then you choose only the first *p* eigenvectors, then the final data set has only *p* dimensions. What needs to be done now is you

need to form a feature vector, which is just a fancy name for a matrix of vectors. This is constructed by taking the eigenvectors that you want to keep from the list of eigenvectors, and forming a matrix with these eigenvectors in the columns (Lindsay, 2002). .

$$Feature\ \ vector = (eig_1, eig_2, eig_3, \dots eig_n)$$

Step 6. Deriving the new data set

This the final step in PCA, and is also the easiest. Once we have chosen the components (eigenvectors) that we wish to keep in our data and formed a feature vector, we simply take the transpose of the vector and multiply it on the left of the original data set, transposed.

$$Final\ data = Row\ feature\ vector\ \times Row\ data\ adjust$$

where *Row Feature vector* is the matrix with the eigenvectors in the columns transposed so that the eigenvectors are now in the rows, with the most significant eigenvector at the top, and *Row data adjust* is the mean-adjusted data (centered data by Step 2) transposed, i.e. the data items are in each column, with each row holding a separate dimension (Lindsay, 2002).

Final data will give us the original data solely in terms of the vectors we chose. Our original data set had two axes, x and y, so our data was in terms of them. It is possible to express data in terms of any two axes that you like.If these axes are perpendicular, then the expression is the most efficient. This was why it was important that eigenvectors are always perpendicular to each other. We have changed our data from being in terms of the axes *x* and *y,* and now they are in terms of our 2 eigenvectors. In the case of when the new data set has reduced dimensionality,we have left some of the eigenvectors out, the new data is only in terms of the vectors that we decided to keep (Lindsay, 2002).

Step 7. Getting the old data back

Recall that the final transform which can be turned around so that, to get the original data back:

$$RowDataAdjust = RowFeatureVector^{-1} \times FinalData$$

However, when we take all the eigenvectors in our feature vector, it turns out that the inverse of our feature vector is actually equal to the transpose of our feature vector by Theorem 1. This makes the return trip to our data easier, because the equation becomes:

$$RowDataAdjust = RowFeatureVector^{T} \times FinalData$$

However, to get the actual original data back, we need to add on the mean of that original data. So, for completeness (Lindsay, 2002),

$$RowDataAdjust = (RowFeatureVector^{T} \times FinalData) + OriginalMean$$

3. Analysis and Results

In applying the Kaiser-Meyer-Olkin's (KMO) overall measure of sampling adequacy (MSA), a score of 0.850 is recorded which is in the acceptable range based on a KMO overall MSA greater than 0.60 being considered acceptable, (Tabachnick.B.G., 2013).Kaiser-Meyer-Olkin (KMO) Barlett's test of sphericity thr eshold is high and a high significant chi-square (\chi^2= 2102.9 (1 d. p.), p<0.0 001).

Chronbach's Alpha reliability was performed ,to check consistency of the sca le items.Particular sample with the value of 0.856 Chronbach's Alpha shows a high level of internal consistency for our scale.

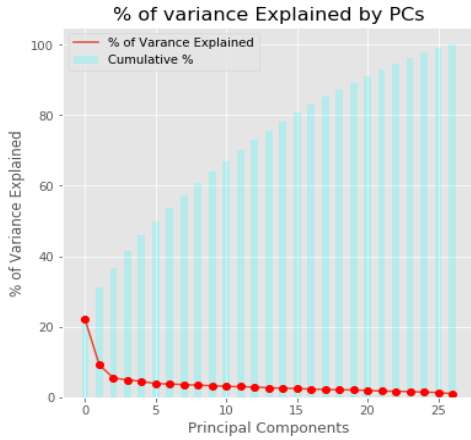
PCA finds principal components in descending order of variations explained. The first components account for more variations than the later ones. The 1st principal component accounts for the maximum amount of variations possible in data, and the 2nd principal component extracts the maximum possible variations in data after excluding what was explained by the 1st component. Extractions can be done until all the by the last principal variations are accounted for components. So, we decided to consider (Kaiser, 1960) the first 6 factors which resulted in 27 Items. From the Table 2 Total Variance Explained it is clear that the 49,97% of the variance is explained by the first six components.

Table 2.Total variance explained

PC#	Eigenvalue	% of Variance Exp	Cumulative %
1	5.951	22.04%	22.04%
2	2.485	9.20%	31.24%
3	1.472	5.45%	36.69%
4	1.327	4.92%	41.61%
5	1.217	4.51%	46.12%
6	1.041	3.86%	49.97%

A scree plot and a bar chart for the cumulated percentage of variance are drawn in the same graph as shown on Figure 3 (Mulhern, et al., 1998).

Figure 3. Principal Components



The next analysis (Table 3) shows how factor loadings. Among the principal c omponents (PCs), at the beginning only the first 6 are selected. The loadings m atrix in output shows the relationship between old variables with new principal components by calculating the coordinate of the old variables along the PC (pri ncipal component) axes:

Table 3.Component matrix

	PC1	PC2	PC3	PC4	PC5	PC6
1	0.182	-0.025	-0.127	0.184	0.137	0.53
2	0.376	-0.399	-0.33	-0.049	0.445	-0.084
3	0.426	-0.295	0.314	0.195	0.256	0.143
4	0.319	0.36	-0.148	0.396	0.291	-0.2
5	0.413	-0.259	0.382	0.175	0.059	-0.241
6	0.458	-0.108	0.459	0.173	0.243	0.138
7	0.432	0.214	0.059	-0.253	0.134	0.071
8	0.442	0.244	-0.471	0.029	0.216	-0.287
9	0.505	0.173	0.029	0.226	0.341	0.072
10	0.561	0.101	-0.132	0.011	0.218	0.118
11	0.553	-0.341	-0.197	-0.248	0.042	0.011
12	0.475	-0.176	-0.27	-0.092	-0.169	0.297
13	0.254	0.497	-0.161	0.15	-0.028	0.308
14	0.609	-0.253	-0.028	-0.22	0.024	-0.213

15	0.434	-0.034	-0.187	0.351	-0.429	0.214
16	0.07	0.66	0.308	-0.216	0.052	0.123
17	0.548	0.138	-0.014	-0.209	-0.223	-0.126
18	0.726	0.13	-0.019	-0.173	-0.196	0.022
19	0.695	0.088	-0.032	-0.197	-0.246	-0.005
20	0.478	0.432	-0.262	0.027	-0.061	-0.227
21	0.134	0.48	0.009	0.153	0.04	0.004
22	0.339	-0.086	0.02	0.561	-0.388	-0.079
23	0.553	-0.189	0.314	0.042	-0.08	-0.084
24	0.606	-0.228	0.117	-0.215	-0.089	0.077
25	0.525	-0.408	-0.06	-0.004	0.004	0.156
26	0.542	0.156	0.267	0.163	-0.137	-0.297
27	0.295	0.492	0.318	-0.289	0.054	0.138

PCA often needs rotation for easier interpretation. The current we used the mo st popular method called Varimax rotation. Varimax orthogonal rotation tries to maximize variance of the squared loadings in each factor so that each factor has only a few variables with large loadings and many other variables with low loa dings (Singh Swapnika), Only loadings greater than |0.40| are considered. Results o f Varimax rotation is shown on Table 4. From rotated component matrix, we e liminated questions 7,10,13,21,26 (Appendix 1) with lowest loadings. We obtai ned components with a Chronbach's alpha greater than 0,6.Components 1,2,3 a re satisfied to condition. We considered each elements in components 4,5,6 sep arately, because these components reliability scale less than 0,6.

Table 4. Rotated Component Matrix

	Component					
	1	**2**	**3**	**4**	**5**	**6**
19.Alumni's presentations influenced my choice	0,682					
18.Presentations of currently enrolled students made a great impact in my choice	0,670					
11.My relatives affected my choice	0,647					
14.I wanted to follow my parents' footsteps	0,628					
24.Opinions of my peers affected my selection	0,617					

17.I was influenced by various advertisement sources (e.g. news, social media, etc)	0,559				
12.University location can be considered as a factor which affected my choice	0,538				
25.Current situation in my family affected my selection	0,489				
6.I was encouraged by a teacher because I was good at my main subjects.		0,692			
3.My high school career advisor influenced my choice		0,647			
5.My religious convictions influenced the selection of my choice		0,612			
23.My high school teacher asked me to specialize in this field as it has high demands nowadays		0,509			
8.Upon graduation good salary affected my choice			0,702		
4.The job's accessibility affected my choice			0,689		
20.Prestige of profession affected my selection			0,565		
9.My academic performance at High School affected my choice			0,460		
16.My personal interest was the strongest factor when choosing a major				0,758	
27.My skills were major effect in my choice				0,675	
2.I was influenced by my parents in my choice				-0,481	
22.My UNT result was important when I selected my major					0,707
15.The major which I had selected provided more state grants than others					0,635
1.University costs played a major role in my choice					0,608

Factor 1: External Influences

External influence factor has 8 items .As a result, external factors play an important role in choosing a profession for a child. It has loadings from 0.682 to 0.489. All these factors are more related to external influences like influence of peers and relatives, and situation of family, also advertisement of specializations. Reliability scale is 0,810(Chronbach alpha).

Factor 2: Teacher influences

Second component gave information that students can influence by school teachers. Reliability scale is 0,638.It is given with loadings 0.692 to 0.509. That is why,parents should make sure that the teacher is a person of good level. Always be in close contact with the teacher.It has 4 items

Factor 3: Influence of occupation salary

The third is important component and it covers job accessibility and prestige of major, also salary. Also, it has 4 items. The child thinks that studying for a prestigious and popular specialty will be received on the highest salary. Influence of occupation salary factor is given with loadings ranging from 0.702 to 0,565.Chronbach's alpha is 0,624.

Since the Chronbach alpha of the other 3 components is very low,we considered each element as a separate factor.Components between 4 and 9 covers only one factor of the study that is:

Factor 4: Personal interest influences

Factor 5: Personal skill influences

Factor 6: Parent's affect

Factor 7: National test affect

Factor 8: State grant affect factor

Factor 9: University cost affect factor

Hence we might conclude that factors between 4 and 9 on a very strong level can influence ones career decision all on its own.

4. Discussion and Conclusion

The factors found in this article show results similar to (Sarwar, et al., 2015) ,despite the fact that two studies used samples from different study fields and

different countries (Kazakhstan, Pakistan). The similarities include the factors such as personal interest affect, skills affect, occupation salary affect, teacher affect, external affect, parent's affect. However, due to the fact that the educational system of the two countries are completely different, (Sarwar, et al., 2015) did not considered some of the facts like UNT results, state grant, cost of the university. Another study which can prove reliability of our study is (William J.Crampton, 2006). They did not work with dimensionality reduction, but defined important factors which influences to choice.

One shortcoming of the study is that majority of participants were from ony city, Almaty. However, we note that Almaty is the largest city in Kazakhstan with many major universities.In order to improve the generalizability the study should be replicated at other universities from different cities. The biggest disadvantage of the our survey is that one third of participants are SDU students. That can change a lot of results. Also, reliability scale of component 2,3 are somewhat low.The results may be improved by increasing number of students participating in the survey and ensuring that there are different universities and majors. In the future work, we can make hypotheses testing between factors and demographic data and determine own concepts. Also, in the future research could look into the relationship between factors on the one hand and students' satisfaction with the choice made and the study success in the bachelor program on the other hand. Also, the way in which data was collected limited the study. Subjects were allowed to sign up to participate in the study and take it online at their own convenience. Administering this type of survey could be more successful if done in person. If any questions arose on the influence listed, having a researcher available to answer questions or clarify the factor listed could provide more accurate data which in turn would lead to more accurate results. Academic consideration factor was not provided in our work. It includes course description, instructors.

The purpose of study was to identify main factors that influence the selection of major field. Principal factor analyses method conducted to reduce number of factors and we decreased factors from 27 to 9.The result of analysis gave 9 main factors. Also, this article shows all the steps needed for PCA along with Python code beyond varimax rotation. Therefore, this would help anyone who wants to run PCA at a deeper level. Teachers and parents can use the results of this study to focus their efforts on supporting students facing the decision about a major. The five-scale translated questionnaires are proved in the Appendix.

References

1 Aldosary, A.S, Assaf, S.A. Analysis of factors influencing the selection of college majors by newly admitted students. *Higher Education Policy,* 3 (9), (2009): pp. 215-220.

2 Alkhateeb H.M. Internal consistency reliability and construct validity of an Arabic translation of the shortened form of the Fennema-Sherman mathematics attitudes scales.*Psychological reports*, 94 (2004): pp. 565-571.

3 Allport G.W. *Attitudes A Handbook of Social Psychology.* Clark University Press, (1935): pp. 219-222.

4 Avellaneda, M , Lee, J.H. Statistical arbitrage in the US equities market. *Quantitative Finance*,10 (2010) :pp. 761-782.

5 Brown, J.D, How are PCA and EFA used in language test and questionnaire development? *Statistics*, 14 (2010): pp. 167-170.

6 Code, W., Merchant, S., Maciejewski, W., Thomas, M., Lo, J. The Mathematics Attitudes and Perceptions Survey: an instrument to assess expert-like views and dispositions among undergraduate mathematics students. *International Journal of Mathematical Education in Science and Technology, 47* (6), (2016): pp. 917-937.

7 Dutton W.H. Measuring attitudes toward arithmetic. *The Elementary School Journal*, 1 (55), (1954): pp. 24-31.

8 Fennema, E., Julia, A. Sherman. "Fennema-Sherman mathematics attitudes scales: Instruments designed to measure attitudes toward the learning of mathematics by females and males". *Journal for research in Mathematics Education,* 7 (5), (1976): pp. 324-326.

9 Fizer, D. Factors affecting career choices of college students enrolled in agriculture. *A research paper presented for the Master of Science in Agriculture and Natural Science degree at The University of Tennessee, Martin* (2013): pp. 51-54.

10 Galotti, K.M. Making a" major" real-life decision: College students choosing an academic major. *Journal of Educational Psychology,* 91 (2), (1999): pp. 379-382.

11 Hyde, J. Sh., et al. Gender comparisons of mathematics attitudes and affect: A meta-analysis. *Psychology of women quarterly,* 14 (3), (1990): pp. 299-324.

12 Jauregui, J. Principal component analysis with linear algebra. *Philadelphia: Penn Arts & Sciences*, (2012).

13 Kaiser, H.F. The application of electronic computers to factor analysis. *Educational and psychological measurement,* 20 (1), (1960): pp. 141-151.
14 Leong, K.E., Nathan, A. College Students Attitude and Mathematics Achievement Using Web Based Homework. *Eurasia Journal of Mathematics, Science & Technology Education,* 10 (6), (2014): pp. 75-78.
15 Li, Han, P., Ralph, L. PCA shows how the effect of population structure differs along the genome. *Genetics,* 211 (1), (2019): pp. 289-304.
16 Liau, A., Kassim, M., Tet, M., Liau, L. Reliability and validity of a Malay translation of the Fennema-Sherman Mathematics Attitudes Scales. *The Mathematics Educator,* 2 (10), (2007): pp. 71-84.
17 Lim S.Y., Chapman, E. Development of a short form of the attitudes toward mathematics inventory. *Educational Studies in Mathematics,* (2013) , 1 (82): pp. 145-164.
18 Lin S.H., Huang Y.C. Development and application of a Chinese version of the short attitudes toward mathematics inventory. *International Journal of Science and Mathematics Education.* (2016), 1 (14): pp. 193-216.
19 Smith, L.I. *A tutorial on principal components analysis*. 2002.
20 Agrey, L., Naltan, L. Determinant factors contributing to student choice in selecting a university. *Journal of Education and Human Development* 3 (2), (2014): pp.391-404.
21 Ma, X., Nand, K. Assessing the relationship between attitude toward mathematics and achievement in mathematics: A meta-analysis. *Journal for research in mathematics education* (1997): pp. 26-47.
22 Melancon, J.G., Bruce Th., Shirley B. Measurement integrity of scores from the Fennema-Sherman Mathematics Attitudes Scales: The attitudes of public school teachers. *Educational and Psychological Measurement,* 54 (1), (1994): pp. 187-192.
23 Michaels, L.A., Robert, A.F. Construction and validation of an instrument measuring certain attitudes toward mathematics. *Educational and Psychological Measurement,* 37 (4), (1977): pp. 1043-1049.
24 Moenikia, M., Adel Zahed-Babelan. A study of simple and multiple relations between mathematics attitude, academic motivation and intelligence quotient with mathematics achievement. *Procedia-Social and Behavioral Sciences,* 2 (2), (2010): pp. 1537-1542.
25 Mulhern, F., Gordon, R. Development of a shortened form of the Fennema-Sherman Mathematics Attitudes Scales. *Educational and psychological Measurement,* 58 (2), (1998): pp. 295-306.

26 Pearson, K. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science,* 2 (11), (1901): pp. 559-572.

27 Pierce, R., Stacey, K., Barkatsas, A. A scale for monitoring students' attitudes to learning mathematics with technology. *Computers & Education,* 48 (2), (2007): pp. 285-300.

28 Richardson, F.C., Richard, M.S. The mathematics anxiety rating scale: psychometric data. *Journal of counseling Psychology,* 19 (6), (1972): pp. 551-554.

29 Sandman, R.S. The Mathematics Attitude Inventory: Instrument and User's Manual. *Journal for research in Mathematics Education,* 11 (2), (1980): pp.148-49.

30 Sanguansat, P. Principal component analysis. *BoD–Books on Demand,* (2012): pp. 1-23.

31 Sarwar, A., Rizwana, M. Factors affecting selection of specialization by business graduates. *Science International*, 27 (1), (2015): pp. 47-50.

32 Schofield, H.L. Sex, grade level, and the relationship between mathematics attitude and achievement in children. *The Journal of Educational Research,* 75 (5), (1982): pp. 280-284.

33 Shlens, J. A Tutorial on principal Component Analysis, (2014): pp. 1-12.

34 Swapnika, S., Kapse, M., Sonwalkar, J. Factors Which Affect the Career and Subject Preference of the Female. *Jounal of Women's Entrepreneurship and Education.* 1 (2), (2011): pp.89-107.

35 Subasi, A., Gursoy, M.I. EEG signal classification using PCA, ICA, LDA and support vector machines. *Expert systems with applications*, 12 (37), (2010): pp. 8659-8666.

36 Tabachnick, B.G., Linda S.F., Jodie, B.U. *Using multivariate statistics*. Boston, MA: Pearson, (2007): pp. 476-480.

37 Tapia M., Marsh G.E. An instrument to measure mathematics attitudes. *Academic Exchange Quarterly,* 2 (8), (2004): pp.16.

38 Turan, C., Kadyrov, S., Burissova, D. An Improved Face Recognition Algorithm Based on Sparse Representation. *2018 International Conference on Computing and Network Communications,* (2018): pp. 32-35.

39 White, A.L., Mathematical attitudes, beliefs and achievement in primary pre-service mathematics teacher education. *Mathematics teacher education and development,* 7 (2005): pp. 33-52.

40 Wigfield, A., Judith, L.M. Math anxiety in elementary and secondary school students. *Journal of educational Psychology*, 2 (80), (1988): pp. 210-216.
41 William, J.C., Kent, W.A., Schambach, T. *Factors influencing major selection by college of business students.* Information system, Illions State University, 7 (2016): pp. 226-229.

Appendix

	English	Russian	Kazakh
1	University costs played a major role	Плата за обучение в университете сыграла большую роль в моем выборе	Менің таңдауымда университеттің оқу ақысы үлкен рөл атқарды
2	I was influenced by my parents	На меня повлияли мои родители в моем выборе	Менің таңдауыма ата-анам әсер етті
3	My high school career adv isor influenced my choice	Консультант по карьере в моей школе повлиял на мой выбор	Менің таңдауыма мектебімдегі мамандық таңдау бойынша кеңесші әсер етті
4	The job's accessibility affe cted my choice	Доступность работы повлияла на мой выбор	Жұмыстың қолжетімділігі таңдауыма әсер етті
5	My religious convictions i nfluenced the selection of my major	Мои религиозные убеждения повлияли на мой выбор	Менің діни сенімдерім таңдауыма әсер етті
6	I was encouraged by a teacher because I was good at my main subjects.	На меня повлиял учитель, потому что я был хорош в своих основных предметах.	Маған мектеп мұғалімі әсер етті, себебі мен негізгі пәндерден жақсы болдым.

7	My Life Experiences have affected me (eg. You want to be a doctor, because a doctor saved someone's life in your family)	Мой собственный жизненный опыт повлиял на мой выбор (напр. Я хочу быть врачом, потому что врач спас чью-то жизнь в моей семье)	Менің өмірлік тәжірибем таңдауыма әсер етті (мысалы, Мен дәрігер болғым келеді, өйткені дәрігер менің отбасымдағы біреудің өмірін сақтап қалды)
8	Upon graduation good sal ary affected my choice	Наличие хорошей зарплаты после окончания учебы повлияло на мой выбор	Оқуды аяқтағаннан кейін жақсы жалақы алу мүмкіндігі таңдауыма әсер етті
9	My academic performance in High School affected the selection	Моя успеваемость в средней школе повлияла на мой выбор	Менің орта мектептегі үлгерімім таңдауыма әсер етті
10	Duration of schooling (e.g . the major will require fur ther training like a master' s degree)	Продолжительность обучения повлияла на мой выбор (например, профессия потребует дальнейшего обучения, как степень магистра).	Оқу ұзақтығы таңдауыма әсер етті (мысалы, мамандық магистр дәрежесі секілді одан әрі оқуды талап етеді
11	Extended family members affected my selection	Мои родственники повлияли на мой выбор	Менің туыстарым таңдауыма әсер етті
12	University location can be considered as a factor which affected my selection	Расположение университета можно рассматривать как фактор, повлиявший на мой выбор	Университеттің орналасуын таңдауыма әсер еткен фактор ретінде қарастыруға болады

13	Reputation of the university was important for me	Репутация университета была важна для меня в моем выборе	Таңдауым үшін университеттің беделі маңызды болды
14	I wanted to follow my parents footsteps	Я хотела пойти по стопам родителей	Мен ата- анамның ізімен жүргім келді
15	The major which I had selected provided more state grants than others	Профессия, которую я выбрал, давала больше государственных грантов, чем другие	Басқаларға қарағанда мен таңдаған мамандық бойынша көбірек мемлекеттік гранттар берілді
16	My personal interest was the strongest factor when choosing a major	Мой личный интерес был самым сильным фактором при выборе специальности	Менің жеке қызығушылығым мамандықты таңдауда ең үлкен фактор болды
17	Academic assessment of the major that I had selected based from printed or online information	На меня повлияли различные источники рекламы (например, новости, социальные сети и т. д)	Маған түрлі жарнама көздері әсер етті (мысалы, жаңалықтар, әлеуметтік желілер және т. б.)
18	Presentations of currently enrolled students made a great impact	Презентации зачисленных студентов оказали большое влияние на мой выбор	Қазіргі таңда сол мамандық бойынша оқып жатқан студенттерінің презентациялары таңдауыма үлкен әсер етті.
19	Alumni's presentations inf luenced my choice	Презентации выпускников повлияли на мой выбор	Түлектердің презентациялары таңдауыма әсер етті

20	Prestige of profession affected my selection	Престиж профессии повлиял на мой выбор	Мамандықтың беделі таңдауыма әсер етті
21	I assumed that professionals in this fi eld can help develop my country	Я считаю, что профессионалы в этой области могут помочь развитию моей страны	Менің ойымша, осы саладағы мамандар еліміздің дамуына көмектесе алады
22	My school graduation exam result was important when I selected my major	Мой результат ЕНТ был важен, когда я выбирал свою специальность	Мамандығымды таңдағанда ҰБТ-ның нәтижесі маңызды болды
23	My high school teacher asked me to specialize in this field as it has high demands nowadays	Мой учитель средней школы попросил меня специализироваться в этой области, поскольку она имеет высокие требования в настоящее время	Менің орта мектеп мұғалімім маған осы саланы меңгеруге кеңес берді, себебі ол қазіргі уақытта жоғары сұраныста бар сала
24	Opinion of my peers affected my selection	Мнения моих сверстников повлияли на мой выбор	Менің құрдастарымның пікірлері таңдауыма әсер етті
25	Current situation in my family affected my selection	Текущая ситуация в моей семье повлияла на мой выбор	Менің отбасымның сол уақыттағы жағдайы таңдауыма әсер етті
26	Famous personalities who had the same specialization in that field affected my major selection	Знаменитые личности, которые имели ту же специализацию в этой области, повлияли на мой основной выбор	Осы салада маманданған танымал тұлғалар негізгі таңдауыма әсер етті

27	My skills were major effect in my choice	Мои навыки оказали большое влияние на мой выбор	Менің қабілеттерім таңдауыма үлкен әсер етті