http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 9 (6): 1226-1230, 2010
ISSN 1812-5638

Statistical Study on Principal Factors Affecting Employment of Chinese Undergraduates

[1]J. Hua, [2]W. Zhuge, [1]K. Zhou and [1]L. Meng
[1]College of Information Engineering, Zhejiang University of Technology, Hangzhou, 310032, China
[2]College of Foreign Language, Hangzhou Dianzi University, Hangzhou, China

Abstract: Due to the heavy employment pressure in china, the employment of the undergraduates attracts much attention in recent years. Accordingly, this study proposes a SPSS-based statistical method to study the employment issue, where thirteen parameters are carefully chosen to construct the employment database. The proposed method first performs the quantitative and the standardized operations and then calculates the correlated matrix of parameters. Moreover, after proving that the correlated matrix satisfies Kaiser-Meyer-Olkin (KMO) condition, we perform eigenvalue decomposition and compute the variance contribution rate through Principal Component Analysis (PCA) techniques. Both the eigenvalue and the variance contribution rate are used to study the importance of each parameter and finally lead to an importance sort. Therefore, we can quantificationally study the influence of each parameter thrown on the undergraduate employment and find three most important parameters affecting undergraduate employment: university, major and family location.

Key words: SPSS, principal component analysis, statistical study, employment, undergraduate

INTRODUCTION

Recently, the number of Chinese undergraduate grows significantly, with 1.15 million in 2001, 3.30 million in 2003 and 5.59 million in 2009, which burdens the employment market. According to conservative estimation, employing ratio of 70% produces about 1.68 million unemployed undergraduates in 2009 (Wang *et al*., 2009; Dong and Yig, 2009; Rong *et al*., 2009). Moreover, the number of unemployed undergraduates may increase rapidly in the future, which will cause many social problems (Wang, 2007; Lin, 2008).

Chinese scientists had tried to find out some factors affecting employment and then instruct the job-hunting of undergraduates. However, till now there are no authoritative conclusions about these factors. In fact, conventional studies usually concern the data collection and the qualitative research (Gu, 2008), which is simple but cannot deal with complicated datas. In order to account for this problem, people should exploit mathematical tool to analyze the intricate employment data, such as Principal Component Analysis (PCA) (Blanchard *et al*., 2007).

As a popular tool in mathematical analysis, PCA is widely used in statistical study, such as medical signal processing (Castells *et al*., 2007; Lathauwer *et al*., 2000), working efficiency evaluation (Hu *et al*., 2007), biologic data analysis (Reich *et al*., 2008), soil analysis (Kooch *et al*., 2008) and machine learning (Hung and Liao, 2008) (Chinnasarn *et al*.., 2006). However, no one had applied PCA to analyze the employment issues. Hence, we choose PCA as mathematic tool in this study. Moreover, the PCA tools have been integrated into SPSS software (Field, 2009; Miller and Acton, 2009), thus, our investigations will done in SPSS environments. Accordingly, this paper collects some samples about employment and then analyzes the influence of the possible factors affecting undergraduate employment by SPSS.

In this study, we choose thirteen indexes (factors): university, major, family location, academic degree, employer, political status, gender, residence classification, position, related experience, attitude toward employment, additional skill and approach for employment and our mission is to find which index is the most important factor affecting undergraduate employments. To realize this aim, we first perform the parameterization operation to the indexes to obtain the standardized numbers, which we call parameters now. Then, we calculate their correlation matrix and the corresponding eigenvalue. Moreover, variance contribution rates for these parameters are derived in SPSS and both eigenvalue and variance contribution rates are utilized as indicators of importance. Finally, with the help of importance information, we sort the thirteen indexes and find three most important indexes: university, major and family location.

Corresponding Author: Jingyu Hua, College of Information Engineering, Zhejiang University of Technology, Hangzhou, 310032, China

Table 1: Index example (partial)

No.	Gender	Political status	Residence classification	University	Major	Family location	Academic degree
1	Female	CPC	Country	Yangzhou Univ.	Landscape Architecture	Rugao	B.Sc.
2	Female	CCYL	Country	Xian Jiaotong Univ.	Civil Eng.	Huaian	B.Sc.
3	Female	Others	City	Zhejiang Univ. Of Tech	Electronic Engineering	Quzhou	B.Sc.
4	Male	CCYL	Country	Zhejiang Univ.	Bio-Eng.	Hangzhou	B.Sc.
5	Male	Others	City	Suzhou Univ.	Finance	Kunshan	B.Sc.

No.	Employer	Position	Approach for employment	Attitude about employment	Predominance	Related experiences
1	Government	Technique	Campus recruitment	Work for employer	Average	Rich
2	GSI	Technique	Network recruitment	Work for employer	Average	Rich
3	Non-state Enterprise	Technique	Campus recruitment	Work for employer	Average	Few
4	Others	Others	Campus recruitment	Work for employer	Famous University	Rich
5	Non-state Enterprise	Others	Others	Work for himself	Average	Rich

CCYL: Chinese Communist Youth League, CPC: Chinese Communist Party, GSI: Government Sponsored Institution

Table 2: Employment rate (partial)

No.	Gender	Political status	Residence classification	University	Major	Family location	Academic degree
1	0.9699	0.9777	0.9682	0.9000	0.9000	0.8500	0.9769
2	0.9699	0.9685	0.9682	0.9500	0.9455	0.9765	0.9769
3	0.9699	0.9777	0.9682	0.9000	0.9792	0.9815	0.9769
4	0.9729	0.9685	0.9682	0.9500	0.0000	0.9063	0.9769
5	0.9729	0.9676	0.9741	0.9000	0.5000	0.9863	0.9769

No.	Employer	Position	Approach for employment	Attitude about employment	Additional skill	Related experience
1	0.9674	0.9868	0.9739	0.9787	0.9726	0.9745
2	0.9820	0.9868	0.9744	0.9787	0.9726	0.9745
3	0.9674	0.9763	0.9650	0.9787	0.9726	0.9745
4	0.9593	0.9060	0.9739	0.9787	0.9831	0.9745
5	0.9651	0.9060	0.9722	0.9498	0.9726	0.9745

PARAMETERIZATION PROCESS

The choice of indexes: Generally, there are three groups of factors influencing the employment of undergraduates: those related to individuals themselves, schools and the society. Taking all of the above into consideration, we pick up the 13 indexes mentioned in previous section. These indexes are: the employment rate of undergraduates with different genders (A1); the employment rate of undergraduates with different political status (A2); the employment rate of undergraduates with different residence classification (A3); the employment rate of undergraduates from different universities (A4); the employment rate of undergraduates with different majors (A5); the employment rate of undergraduates with different family location (A6); the employment rate of undergraduates with different academic degrees (A7); the employment rate of undergraduates with different employers (A8); the employment rate of undergraduates with different positions (A9); the employment rate of undergraduates with different approaches to the job (A10); the employment rate of undergraduates with different attitudes (A11); the employment rate of undergraduates with different additional superiority (A12); and the employment rate of undergraduates with different trainee experience (A13). With these indexes, we can extract data from database and build a sample database, which is composed of more than one thousand records. Hence, due to the limitation of space, we just give five records in Table 1 as examples. From Table 1, we explicitly see that each index (A_i) must take value from a certain finite alphabet and each possible value in such a alphabet is used for classification criterion of groups.

Parameterization of index: Each index can be parameterized as:

$$S_{ij} = B_{ij} / C_{ij} \quad (1)$$

where, B_{ij}, C_{ij} and S_{ij} represent the number of the undergraduates who have found a job, the total number of undergraduates and the employment rate of Group j in sample A_i respectively. By Eq. 1, Table 1 can modified as Table 2. For example, we take into account the Gender index, i.e., A1 index. There are two possible values for this index: male or female, which means that there are two groups for A1 index. Then we have:

- C_{11}: number of the female undergraduate
- C_{12}: number of the male undergraduate
- B_{11}: number of the female undergraduate who have found a job
- B_{12}: number of the male undergraduate who have found a job

Finally, we can calculate employment rates (S_{11}, S_{12}), i.e., 0.9669 for female and 0.9729 for male, which means that a man is easier to find a job than a woman.

PCA PROCESS

Based on the data matrix above, we can perform PCA operation in SPSS and get test result of KMO and Bartletts (Fig. 1).

$$R = \begin{vmatrix} 1.000 & & & & & & & & & & & & \\ 0.016 & 1.000 & & & & & & & & & & & \\ -0.091 & 0.025 & 1.000 & & & & & & & & & & \\ 0.011 & 0.041 & 0.024 & 1.000 & & & & & & & & & \\ -0.009 & 0.010 & 0.006 & 0.411 & 1.000 & & & & & & & & \\ -0.008 & 0.004 & 0.029 & 0.246 & 0.224 & 1.000 & & & & & & & \\ 0.079 & 0.132 & 0.006 & 0.060 & 0.020 & -0.002 & 1.000 & & & & & & \\ -0.036 & 0.047 & 0.008 & 0.048 & 0.053 & 0.052 & 0.151 & 1.000 & & & & & \\ 0.069 & -0.013 & -0.004 & 0.090 & 0.094 & 0.035 & 0.057 & 0.010 & 1.000 & & & & \\ 0.028 & 0.041 & -0.014 & 0.040 & 0.027 & 0.020 & 0.096 & 0.034 & 0.022 & 1.000 & & & \\ -0.016 & 0.006 & -0.020 & 0.046 & 0.060 & 0.025 & -0.004 & -0.013 & 0.067 & 0.064 & 1.000 & & \\ -0.004 & 0.012 & -0.013 & 0.013 & 0.035 & -0.002 & 0.017 & 0.014 & 0.011 & 0.015 & 0.045 & 1.000 & \\ 0.025 & 0.044 & -0.029 & 0.013 & 0.016 & 0.026 & 0.021 & 0.027 & 0.044 & 0.009 & 0.062 & 0.004 & 1.000 \end{vmatrix} \quad (2)$$

KMO stands for Kaiser-Meyer-Olkin measure of sampling adequacy (Field, 2009). A larger KMO means more common factors among variables, thus more suitable to perform PCA operation. According to Kaiser (Miller and Acton, 2009), if KMO falls below 0.5, its not suitable to do PCA. Here, with a KMO valuing 0.594, we can perform the operation. The correlation matrix can be represented in formula 2.

From matrix R, we can see that there is high correlation between A4 and A5, followed by that between A4 and A6. A5 also correlates with A6, so do A7 and A8. We cannot find such high correlation between other indexes. In fact, this is indeed the case.

We execute the variance analysis by SPSS and obtain Fig. 2, where the eigenvalue of correlation matrix (descending order), the contribution rate of variance as well as the Cumulative contribution rate of variance are shown in the left three columns. Moreover, in order to further study the variance variation, we extract the eigenvalue of correlation matrix, the contribution rate of variance and the Cumulative contribution rate of variance to construct Table 3, where G_i denotes the ith component.

According to Table 3, there are two ways to count the number of principle components. The first is to include all those quartiles with its eigenvalue larger than one. The second is to judge by their cumulative contribution rates. Here we adopt the first method, thus G_1~G_5 become the quartiles, which are, actually,synthetic indexes transformed from the original indexes. Generally, the value of first quartile can be used as the synthetic criteria to judge whether the scheme is good or not, while those of second and other quartiles represent other features waiting to be evaluated and can even serve as supplements when the contribution rate of the first quartiles cannot represent the information of the original index system.

For the convenience of explanation, we use SPSS to obtain the rotated component matrix and the corresponding variance contribution rate in Fig. 3 and 4.

KMO and bartlett's test

Kaiser-meyer-olkir measure of sampling adequancy		**.594**
Bartlett's test of Sphericity	**Approx. Chi-square**	**1943.262**
	df	**78**
	Sic.	**.000**

Fig. 1: The test result of KMO and bartlett

Table 3: Variance analysis table

Component G_i	Eigenvalue λ_i	Contribution rate F_i%	Cumulative contribution rate ΣF_i%
G1	1.683	12.946	12.946
G2	1.263	9.712	22.658
G3	1.149	8.838	31.495
G4	1.063	8.180	39.676
G5	1.005	7.728	47.403
G6	0.988	7.597	55.001
G7	0.974	7.489	62.490
G8	0.966	7.429	69.919
G9	0.895	6.885	76.805
G10	0.851	6.545	83.350
G11	0.806	6.196	89.546
G12	0.775	5.963	95.509
G13	0.584	4.491	100.000

From Fig. 3 and 4, we can see that A4, A5 and A6 compose the first quartile; A7, A8 and A2 are the second quartile; A1, A3 and A9 form the third quartile; A13 and A11make the fourth quartile and A12 and A10 are the fifth quartile. We can also define the major factor as A4, A5 andA6 and rank the thirteen parameters according to their importance as: A4, A5, A6, A7, A8, A2, A1, A3, A9, A13, A11, A12 and A10. That is to say, what affect the employment most are university, major and family location. And University influences undergrads employment most, with major, family location, academic degree, employer, political status, gender, residence classification, position, related

Total variance explained

Component	Initial eigenvalues			Extraction sums of squared loadings			Rotation sums of squared loadings		
	Total	Variance (%)	Cumulative (%)	Total	Variance (%)	Cumulative (%)	Total	Variance (%)	Cumulative (%)
1	1.683	12.946	12.946	1.683	12.946	12.946	1.620	12.462	12.462
2	1.263	9.712	22.658	1.263	9.712	22.658	1.262	9.704	22.166
3	1.149	8.838	31.495	1.149	8.838	31.495	1.135	8.731	30.898
4	1.063	8.180	39.676	1.063	8.180	39.676	1.083	8.328	39.226
5	1.005	7.728	47.403	1.005	7.728	47.403	1.063	8.178	47.403
6	0.988	7.597	55.001						
7	0.974	7.489	62.490						
8	0.966	7.429	69.919						
9	0.895	6.885	76.805						
10	0.851	6.545	83.350						
11	0.806	6.196	89.546						
12	0.775	5.963	95.509						
13	0.584	4.491	100.000						

Extraction method: Principal component analysis

Fig. 2: Variance analysis results

Rotated component matrix[a]

	Component				
	1	2	4	3	5
A4	0.773	0.076	0.040	0.008	0.033
A5	0.764	-0.005	0.020	0.025	0.083
A6	0.614	0.004	-0.072	0.023	-0.082
A7	0.009	0.707	0.201	0.048	0.101
A8	0.095	0.558	-0.154	-0.059	0.036
A2	-0.034	0.556	-0.083	0.149	-0.037
A1	-0.010	0.081	0.766	-0.032	-0.103
A3	-0.069	0.143	-0.575	-0.011	-0.117
A9	0.204	0.006	0.340	-0.329	0.046
A13	-0.035	0.120	0.010	0.786	-0.202
A11	0.047	-0.136	-0.067	0.565	0.481
A12	-0.004	-0.044	-0.041	-0.086	0.715
A10	0.009	0.271	0.130	0.009	0.475

Extraction method: Principal component analysis.
Rotation method: Varimax with kaiser normalization.
a. Rotation converged in 5 iterations.

Fig. 3: The rotated component matrix

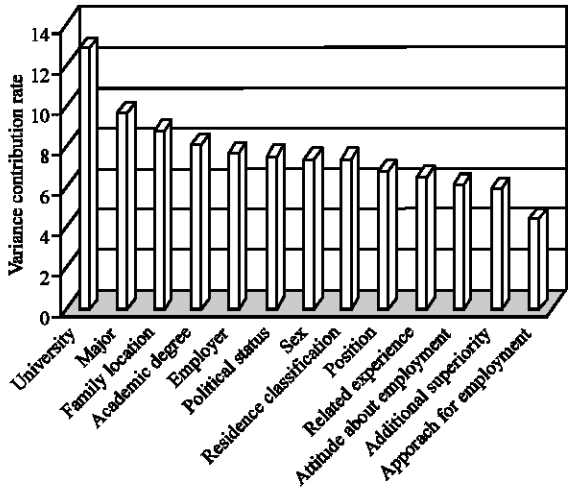

Fig. 4: Variance contribution rate

experience, attitude toward employment, additional superiority and approach for employment following it.

CONCLUSIONS

This study uses PCA to analyze thirteen indexes affecting employment, where we order them according to eigenvalues and contribution rates and find three most important indexes. The result contributes to employment studies and benefits certain decision making departments.

ACKNOWLEDGMENT

This study is sponsored by science foundation for the excellent youth scholars of Zhejiang province (2010) and Zhejiang provincial NSF of China under grant No.Y1090645 (2010-2011).

REFERENCES

Blanchard, G., O. Bousquet and L. Zwald, 2007. Statistical properties of kernel principal component analysis. Machine Learning, 66: 259-294.

Castells, F., P. Laguna, L. Sornmo, A. Bollmann and J. Roig, 2007. Principal component analysis in ECG signal processing. EURASIP J. Applied Signal Process., 2007: 98-98.

Chinnasarn, K., S. Chinnasarn and D.L. Pyle, 2006. Identification of surimi gel strength classes using backpropagation neural network and principal component analysis. J. Applied Sci., 6: 1802-1807.

Dong, Y. and G. Yig, 2009. Research on the employment situation of the first college graduates of athletics education in Dali university. J. Dali Univ., 8: 73-76.

Field, A., 2009. Discovering Statistics Using SPSS. 3rd Edn., Sage Publications Ltd., London.

Gu, A., 2008. The employment situation and countermeasures of college graduates. J. Huizhou Univ. Social Sci. Edn., 28: 90-93.

Hu, J., J. Zheng and Y. Cao, 2007. Application of principal component analysis in performance evaluation of rear depots. Logistics Technol., 26: 120-121.

Hung, Y.H. and Y.S. Liao, 2008. Applying PCA and fixed size LS-SVM method for large scale classification problems. Inform. Technol. J., 7: 890-896.

Kooch, Y., H. Jalilvand, M.A. Bahmanyar and M.R. Pormajidian, 2008. The use of principal component analysis in studying physical, chemical and biological soil properties in Southern Caspian forests (North of Iran). Pak. J. Biol. Sci., 11: 366-372.

Lathauwer, L., B. Moor and J. Vandewalle, 2000. SVD-based methodologies for fetal electrocardiogram extraction. Proc. Acoustics Speech Signal Process. 2000 IEEE Int. Conf., 6: 3771-3774.

Lin, Z., 2008. The coutermeasure on employment of the graduates from the local normal colleges. J. Zhangzhou Normal Univ., 2008: 155-157.

Miller, R. and C. Acton, 2009. SPSS for Social Scientists. Palgrave Macmillan Ltd., USA.

Reich, D., A.L. Price and N. Patterson, 2008. Principal component analysis of genetic data. Nat. Genet., 40: 491-492.

Rong, Y., R.Y. Feng and L.L. Zhang, 2009. Practical significance and implementation approach of colleges and universities to strengthen career guidance curriculum. Res. Teaching, 32: 39-41.

Wang, C.C., X.F. Wang, N. Liu and X. Xu, 2009. Survey and analysis on employment ability training for finance and economics major students. J. Hebei Software Inst., 11: 21-24.

Wang, X.M., 2007. An analysis of undergraduates employment conflicts. J. Chongqing Three-Gorges Univ., 23: 111-113.