Computer Life (CPL) ISSN: 1819-4818
DELIVERING QUALITY SCIENCE TO THE WORLD

Wage Distribution based on Principal Component Analysis

Xinxin Bian*

College of Mathematics and Systems Science, Shandong University of Science and Technology,

Qingdao, Shandong, 266590, China

* Corresponding author's e-mail:13061485005@163.com

Abstract: Total wage distribution is a management element closely related to enterprise human resource strategy. According to the year 2018 26 branches operation statistics data, using the regression analysis method to analyze and 16 indicators, the extraction effect of 26 provinces and cities each branch of the main factors of the distribution of total wages, then using the principal component analysis (pca), on the premise of minimum information loss, will be integrated into many of the original variables is less a few comprehensive indexes, according to related data to evaluate the comprehensive scores of provinces and cities formulated in early 2018 branch whether the allocation of total wages. Secondly, based on the analysis of the main factors, building set up based on the improved entropy method of evaluation index system, the evaluation index of each was to quantify the information of evaluation objects and synthesis method of empowerment, to avoid the interference of man-made factors, ensure the rationality evaluation results, and to establish a reasonable mathematical model of reasonable allocation scheme are presented.
Keywords: Wage Distribution, Principal Component Analysis, regression analysis.

---

1. INTRODUCTION

Total wage distribution is a management element closely related to enterprise human resource strategy. The total salary distribution mechanism of enterprises is very important to the development of enterprises. It not only affects the motivation, regulation and security management of employees, but also helps enterprises to achieve strategic goals, improve business performance, improve market competitiveness and strengthen corporate culture. How to establish a set of scientific and reasonable total wage distribution scheme is a new and significant subject for state-owned enterprises. Total wage refers to the total amount of labor remuneration that each unit pays directly to all its employees within a certain period of time. It is an important index of salary plan management and the basis of calculating average salary.

2. MODELING

This paper is based on the data attached to the second postgraduate mathematical contest in hebei province in 2019, this paper establishes an appropriate mathematical model to analyze the main factors affecting the total amount of wages distributed by branches in 26 provinces and cities. As the units of different indicators given are greatly different, firstly, SPSS is used to conduct standardized

processing of data, so as to reduce errors as much as possible. Secondly, regression analysis is carried out by using the inherent relationship between the original indexes.

When using SPSS for data analysis, when the index has the serious multicollinearity between case, typically through the principal component analysis method to extract the comprehensive index, and according to the variance contribution rate, in the selected principal components variance contribution ratio of the number of indicators under the principle of "$\geq 85\%$", replace the original indexes is analyzed with comprehensive indexes, to eliminate the harm of multicollinearity.

The principal component analysis method is mainly used to reduce the dimension of data sets and convert more indexes into fewer comprehensive indexes, so as to simplify the computation amount through dimension reduction in the case of large data.

Suppose there are n samples, each sample has p variables, forming a data matrix of $n \times p$. Remember that the original variables are $X_1, X_2, \cdots, X_P$, and set their comprehensive index after dimension reduction, namely, the new variables are $Z_1, Z_2, Z_3, \cdots, Z_P (m \leq p)$ or less. It can be proved mathematically that they are the eigenvectors corresponding to m larger eigenvalues of the correlation matrix respectively. Calculation steps of principal component analysis:

(1) Calculate the correlation coefficient matrix, and the calculation formula is:

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix}$$

(2) Calculate eigenvalues and eigenvectors $|\lambda I - R| = 0$ Solution characteristic equation $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_P$, Usually jacobian(Jacobi), Find the eigenvalues and arrange them in order of size; Find the eigenvector $e_i (i = 1, 2, \cdots, p)$ corresponding to the eigenvalue λ_i, namely $\sum_{j=1}^{p} e_{ij}^2 = 1$.

(3) Calculate the principal component contribution rate and cumulative contribution rate, and generally take the characteristic value with the cumulative contribution rate of 85-95%.

Contribution:

$$\frac{\lambda_i}{\sum_{k=1}^{p} \lambda_k} \quad (i = 1, 2, \cdots, p)$$

Cumulative contribution rate:

$$\frac{\sum_{k=1}^{i} \lambda_k}{\sum_{k=1}^{p} \lambda_k} \quad (i = 1, 2, \cdots, p)$$

(4) Calculate principal component load:

$$l_{ij} = p(z_i, x_j) = \sqrt{\lambda_i} e_{ij} (i, j = 1, 2, \cdots, p)$$

(5) Score of each principal component::

$$Z=\begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1m} \\ z_{21} & z_{22} & \cdots & z_{2m} \\ \vdots & \vdots & & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{nm} \end{bmatrix}$$

3. SOLUTION METHOD FOR MODEL

As the units of 16 different indicators are greatly different from each other, there may be huge errors in the direct analysis of specific statistical data between indicators. Therefore, SPSS was used to conduct partial correlation analysis on the data before data analysis, so as to reduce the error as much as possible. The correlation coefficients between the total wage distribution scheme and the 16 indicators were obtained as shown in table 1.

Table 1. * means Sig=0.05, ** means Sig=0.01.

	The correlation		The correlation
Regionalpopulation(10,000)	0.841**	Average selling price of commercial housing (yuan)	0.375
Per capita income of urban and rural residents (yuan)	0.493*	Total business income (ten thousand yuan)	0.986**
Per capita consumption expenditure in urban and rural areas (yuan)	0.435*	Total business cost (ten thousand yuan)	0.989**
Number of urban residents (10,000)	0.936**	Fixed assets for production (ten thousand yuan)	0.960**
Income of urban residents (ten thousand yuan)	0.943**	Return on equity	0.354
Per capita disposable income of urban residents (yuan)	0.608**	Personnel cost ratio	-0.384
GDP (ten thousand yuan)	0.942**	Labor productivity	0.639**
Average salary of employees in other state-owned enterprises (yuan)	0.384	Cost ratio	-0.539**

According to the correlation strength analysis table, we select approximately 0.6 correlation index as the main influencing index, we can see from the results, urban and rural residents per capita income, urban and rural per capita consumption expenditure, other state-owned enterprise on-the-job worker average wage, cost rate, return on equity, the personnel cost rate, commodity house average sales price and the total wages scheme of correlation coefficient is less than 0.6, there are nine other factors between the strong positive correlation.

Due to the strong multicollinearity among the index factors selected in this paper, we converted several indexes with strong correlation in the index factors into a new set of comprehensive indexes without multicollinearity, and then conducted regression analysis to obtain tables 2 and 3.

Table 2. KMO and Bartlett spherical tests

KMO 和 Bartlett 的检验

取样足够度的 Kaiser-Meyer-Olkin 度量。		.763
Bartlett 的球形度检验	近似卡方	508.301
	df	36
	Sig.	.000

It can be seen from table 2 that KMO is 0.763>0.6, and the significance value of Bartlett test is 0.00<0.05, which both indicate that the data can be subject to principal component analysis. As can be seen from table 3, the principal component analysis extracted two principal components whose

variance contribution rate was 80.797% and 11.563%, respectively, and whose characteristic value was greater than 1. The cumulative contribution rate of variance of the two principal components was 92.360%. According to the principle of "$\geq 85\%$", the extracted two principal components could represent the original nine indexes.

Table 4 for the factor loading matrix, it reflects the extraction of the two main components of the correlation between the indexes and strength, all nine indicators can be found with the first principal component has a high load, and urban per capita disposable income, cost rate of main components of the two measures factor and the second load is also relatively high, so by the principal component analysis (pca) to extract the two principal component can reflect the original nine indicators reflected by the factor of information, and can replace the original nine indicators.

Table 3.Total variance analysis table

解释的总方差

成份	初始特征值			提取平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	7.272	80.797	80.797	7.272	80.797	80.797
2	1.041	11.563	92.360	1.041	11.563	92.360
3	.422	4.687	97.047			
4	.131	1.453	98.501			
5	.053	.585	99.086			
6	.045	.499	99.584			
7	.031	.340	99.924			
8	.007	.074	99.998			
9	.000	.002	100.000			

提取方法：主成份分析。

Table 4. Factor loading matrix

成份矩阵[a]

	成份	
	1	2
地域人口	.852	-.469
城镇居民数	.950	-.284
城镇居民收入	.960	.140
城镇居民人均可酉收入	.653	.695
GDP	.971	-.104
业务总收入	.989	-.069
业务总成本	.990	-.051
生产用固定资产	.944	-.030
劳动生产率	.707	.467

提取方法 :主成分分析法。

a. 已提取了 2 个成份。

According to the results of principal component analysis, the extracted two principal components can replace the original nine index factors. The relationship between the two principal components and the nine index factors can be obtained as follows:

$$Z_1 = 0.32X_1 + 0.35X_2 + 0.36X_3 + 0.24X_4 + 0.36X_5 + 0.37X_6 + 0.37X_7 + 0.35X_8 + 0.26X_9$$

$$Z_2 = -0.46X_1 - 0.28X_2 + 0.14X_3 + 0.68X_4 - 0.10X_5 - 0.07X_6 - 0.05X_7 - 0.03X_8 + 0.46X_9$$

Thus, the 9-dimensional variable is changed into a 2-dimensional variable to obtain the principal component synthesis model:

$$F = \frac{7.272}{7.272+1.041} Z_1 + \frac{1.041}{7.272+1.041} Z_2$$

According to the principal component synthesis model, the comprehensive principal component value can be calculated and sorted according to the comprehensive principal component value, and the comprehensive evaluation and comparison can be conducted for each region. The results are shown in table 5:

Table 5. Comprehensive evaluation comparison table and total wage distribution ranking of each region

Save the company	Composite principal component F	ranking	Save the company	The total wages	ranking
Guangdong	6.05	1	Guangdong	384414	1
Jiangsu	5.24	2	Jiangsu	348110	2
Shandong	3.65	3	Zhejiang	279441	3
Zhejiang	3.02	4	Shandong	243007	4
Beijing	1.88	5	Sichuan	218626	5
Sichuan	1.12	6	Beijing	180483	6
Hebei	0.69	7	Hubei	177829	7
Hubei	0.65	8	Hebei	156861	8
Fujian	0.43	9	Heilongjiang	155617	9
Hunan	0.37	10	Fujian	155049	10
Liaoning	0.09	11	Hunan	152664	11
Heilongjiang	-0.47	12	Liaoning	124977	12
Shanxi	-0.66	13	Shanxi	116547	13
Tianjin	-0.77	14	Guangxi	108714	14
Chongqing	-0.83	15	Shanxi	106275	15
Jiangxi	-0.88	16	Chongqing	105175	16
Guangxi	-1.01	17	Jiangxi	93782	17
Shanxi	-1.12	18	Jilin	89285	18
Jilin	-1.21	19	Inner Mongolia	83697	19
Yunnan	-1.46	20	Yunnan	81921	20
Guizhou	-1.58	21	Guizhou	68214	21
Inner Mongolia	-1.68	22	Tianjin	63806	22
Hainan	-2.42	23	Gansu	55790	23
Gansu	-2.52	24	Hainan	30790	24
Ningxia	-3.13	25	Qinghai	21518	25
Qinghai	-3.45	26	Ningxia	18099	26

Contrast table can be found, we a year according to the 2018 26 branches of statistics of the operation situation of regional comprehensive evaluation score ranking and formulate the distribution of total wages at the beginning of 2018's ranking is not consistent, such as tianjin ranking is 14 in the table, the table of tianjin is ranked no. 22, so at the beginning of 2018 the total wages of branches distribution is not reasonable.

REFERENCES

[1] Oulton, N.2001, Must the Growth Rate Decline Baumol's Unbalanced Growth Revisited [J]. Oxford Economic Papers, 2001, 53(4): 605-627.

[2] Ning Wang, Fei Sun, Xiao Hong Shan. Analysis and Short-Term Prediction of Information Service Industry's Prosperity Index[J] Advanced Materials Research.2014(10)：2971-2976.

[3] Hyunsoo Kim, Seo Young Kim. Introduction of the 'Service Industry Development Index'： Initial

thoughts on the new paradigm shift [J] Journal of Service Research. 2016(8): 85-94.

[4] Zhou Xuexin, Zhou Juan. Empirical research on the impact of China's industrial structure changes on economic growth -- based on the panel data analysis of China's eastern and western regions [J]. Business economics, 2008 (7): 22-24.

[5] Xu Feng. Research on evaluation system of comprehensive development level of modern service industry based on fuzzy evaluation [D]. Wuhan: Huazhong University of science and technology, 2006.