

Closeness between Objects and Variables in a Biplot

Seongmo Yoo

Korea University , Department of Informational Statistics
Jochiwon, Chungnam, 339-700, Republic of Korea
syoo@tiger.korea.ac.kr

Sangwoo Kim

Eyeclick, 12th floor, Dae-won Bldg. 946-18, Daechi-Dong
Kangnam-Gu, Seoul, 135-280, Republic of Korea
miner@eyeclick.co.kr

Kangho Choi

Korea University , Department of Informational Statistics
random@apstat.korea.ac.kr

1. Introduction

Biplot is a exploratory data visualization technique to display the multivariate data into a low dimensional scatter plot. The data matrix is either continuous or categorical. In this talk, we define the closeness between an object(or a row) and a variable(or a column) in a biplot for the data matrix. For the illustration of the closeness between an object and a variable, the election results of the 15th and 16th Korean national assembly and Korean president are analyzed and compared.

2. Definition of Closeness between a Object and a Variable in a Biplot

2.1 Biplot

In biplot, usually the data matrix $X_{n \times p}$ is decomposed and approximated as follows:

$$X_{n \times p} = U_{n \times p} D_{p \times p}^a D_{p \times p}^{1-a} V'_{p \times p} = G_{n \times p} H'_{p \times p} \cong G_{n \times 2} H'_{2 \times p}$$

where, U and V are orthogonal matrices, D is a diagonal matrix having eigenvalues of variance-covariance matrix of $X_{n \times p}$ as its diagonal elements, and $G = UD^a$ and $H' = D^{1-a}V'$. The type of biplot is determined by the value of a . When $a=0$ the biplot is called to be a principal component factor biplot, and when $a=1$ the biplot is called to be a principal component biplot.

2.2 Definition of Closeness

If the data matrix is standardized, the Mahalanobis and the Euclidean distances are the same and the (i, j) th element x_{ij}^s of the standardized data matrix can be expressed as

$$x_{ij}^s \cong g_{i1} h_{j1} + g_{i2} h_{j2} = \overset{p}{g}_i \cdot \overset{p}{h}_j = |\overset{p}{g}_i| \cdot |\overset{p}{h}_j| \cos \mathbf{q}_{ij}$$

The closeness r_{ij} between i^{th} object and j^{th} variable is defined as the cosine of two vectors(i.e. $\cos \mathbf{q}_{ij}$) from the principal component factor biplot for the standardized data matrix.

3. Case Study

For the illustration of the closeness between an object and a variable, the election results of recent two Korean national assembly and president are analyzed with following steps. First, the actual votes for political parties in the main provincial regions are standardized based on the assumption that the row categories (regions) and the column categories (political parties) are independent. Second, the column categories (political parties) are again standardized with column means and standard deviations. Third, the factor biplots are obtained for the double standardized data matrix. Finally, the closeness between political parties and provincial regions are calculated and reported in [Table 1] and [Table 2].

[Table 1] The Closeness for the Election of Korean National Assembly

15 th	SL	BS	DG	IC	GJ	DJ	GG	GW	CB	CN	JB	JN	GB	GN	JJ	US
SINHAN	0.522	0.973	-0.483	0.714	-0.768	-0.238	0.999	0.780	-0.175	-0.276	-0.538	-0.749	-0.016	0.967	-0.256	-
KUKMIN	0.338	-0.43	-0.381	0.1	0.98	-0.609	-0.59	-0.976	-0.659	-0.577	0.994	0.986	-0.77	-0.406	0.914	-
MINJU	0.522	0.973	-0.483	0.713	-0.768	-0.238	0.999	0.78	-0.175	-0.277	-0.538	-0.749	-0.015	0.967	-0.256	-
JAMIN	-0.941	-0.422	0.955	-0.83	-0.471	0.999	-0.245	0.455	0.999	0.997	-0.717	-0.498	0.982	-0.445	-0.894	-
16 th	SL	BS	DG	IC	GJ	DJ	GG	GW	CB	CN	JB	JN	GB	GN	JJ	US
HANNARA	0.001	0.899	0.998	-0.641	-0.709	-0.501	-0.016	0.68	-0.409	-0.313	-0.955	-0.914	0.91	0.886	-0.263	0.298
SAEMIN	-0.012	-0.895	-0.997	0.633	0.716	0.492	0.005	-0.672	0.4	0.303	0.958	0.918	-0.9141	-0.882	0.273	-0.288
JAMIN	0.992	-0.321	0.057	0.681	-0.789	0.795	0.99	-0.642	0.854	0.903	-0.414	-0.517	0.526	-0.348	-0.99	-0.91
MINKUK	0.087	0.858	0.998	-0.573	-0.767	-0.425	0.07	0.614	-0.33	-0.23	-0.977	-0.945	0.942	0.844	-0.344	0.215

[Table 2] The Closeness for the Election of Korean President

14 th	SL	BS	DG	IC	GJ	DJ	GG	GW	CB	CN	JB	JN	GB	GN	JJ	US
MINJA	-0.717	0.723	0.947	-0.201	-0.758	-0.159	-0.248	0.025	-0.032	-0.142	-0.778	-0.766	0.985	0.871	0.88	-
MINJU	0.433	-0.441	-0.999	-0.149	0.936	-0.191	-0.101	-0.368	-0.314	-0.207	0.947	0.94	-0.865	-0.648	-0.99	-
KUKMIN	0.583	-0.576	0.458	0.939	-0.757	0.953	0.921	0.992	0.984	0.958	-0.736	-0.751	-0.025	-0.358	0.601	-
15 th	SL	BS	DG	IC	GJ	DJ	GG	GW	CB	CN	JB	JN	GB	GN	JJ	US
HANNARA	0.151	0.766	0.679	-0.131	-0.874	-0.473	-0.237	0.273	-0.233	-0.502	-0.881	-0.881	0.915	0.76	0.573	0.92
KUKMIN HOI	0.178	-0.933	-0.404	-0.197	0.984	0.161	-0.091	-0.571	-0.095	0.194	0.987	0.987	-0.735	-0.93	-0.808	-0.997
KUKMIN SHIN	-0.142	0.958	-0.24	0.756	-0.886	0.476	0.681	0.954	0.684	0.446	-0.88	-0.88	0.167	0.961	0.999	0.834

4. Conclusion

The closeness between an object and a variable in a biplot for the data matrix is defined. For the illustration of the closeness defined in this talk, the election results of the Korean national assembly and the president are analyzed.

REFERENCES

- [1] Gabriel K.R. (1971). The biplot graphic display of matrices with application to principal component analysis, *Biometrika* 58, 3, p.453.
- [2] Huh M. H (1999). Quantification methods for multivariate data, Freedom Academy Pub. Co.

RESUME

Seongmo Yoo is an associate professor of Korea University. Sangwoo Kim is a statistical consultant of Eyecllick Inc. and he got his master degree in applied statistics under the guidance of Prof. Yoo. Kangho Choi is a graduate student of Korea University and he is also under the guidance of Prof. Yoo.