Mini Review

A Note on Factor Analysis Applied in Medical Research

KC Bhuyan*

Department of Statistics, Jahangirnagar University, Bangladesh

*Corresponding author: KC Bhuyan, Professor of Statistics (Retired), Jahangirnagar University, Bangladesh.

Mini Review

The empirical analysis of the data observed in the field of medical science is based on some qualitative as well as quantitative character. But some of the characters are common with other characters of human being. For example, diabetes depends on age, height, weight, food habit, income, occupation, etc. In some studies, in both home and abroad, it was observed that body mass index (BMI) or in other words, higher level of obesity is one of the responsible factors for diabetes [1-6]. But BMI is a character which is measured using the character weight and height and level of obesity is decided using the level of BMI. Hence level of obesity is a common character among the characters height, weight, and BMI. Therefore, to study the impact of height and weight on diabetes, it is better to study the impact of BMI on diabetes. The statistical tool to select such a character among many characters is known as factor analysis.

Factor analysis is a statistical tool used to describe variability among observed, correlated variables in terms of a potential lower number of unobserved variables called factors. The observed variables are modeled as linear combination of the potential factors, plus error terms. This technique of selection of factors is commonly used in biology, medical research, psychometrics, personality theories, educational statistics, etc.

Let us consider that we have a set of observable random variables x_1, x_2, ..., x_k with means M_1, M_2..., M_k. Let us also consider that there be some unknown constants a_{ij} (i = 1, 2, ..., k; j = 1, 2, ..., p) and some unobservable random variables F_1, F_2 ..... F_p. Then

$$x_i - M_i = a_{i1} F_1 + a_{i2} F_2 + ... + a_{ip} F_p + e_i$$

Here e_i is the unobservable stochastic terms with zero means and finite variance which may not be same for all i. Here a_{ij} is the factor weight of F_j which indicates the importance of F_j in explaining the total variation in the variables x_1, x_2.... x_k. The factor F_j is a common factor among the variables x_1, x_2, ... x_k. Like level of obesity is a common factor among the variables BMI, height, and weight. The object of factor analysis is to identify the common factor F_j which is important to study the variability in the data set. In practice, a common factor level of obesity is important to study the variability in the level of blood sugar (diabetes).

Table 1: Important results of factor analysis including Factor loadings in analyzing the data of prevalence of diabetes among adult people of Bangladesh.

Variables	Coefficient for Factor - 1	Coefficient for Factor- 2	Coefficient for Factor - 3	Coefficient for Factor –4	Coefficient for Factor - 5	Communality
Residence	-0.397	0.618	0.219	0.24	0.184	0.577
Age	0.252	-0.125	-0.169	-0.307	0.755	0.71
Gender	0.799	0.132	0.24	0.159	-0.133	0.742
Marital status	-0.016	-0.212	0.384	0.661	0.489	0.502
Religion	0.118	-0.563	-0.134	0.493	-0.271	0.856
Education	-0.694	0.435	0.066	0.1	0.024	0.7
Occupation	0.772	0.38	-0.025	0.177	-0.072	0.666
Type of work	0.612	0.147	0.432	0.111	0.228	0.489
Income	-0.071	0.574-0.584	0.292	-0.059		0.524
Smoking habit	0..493	0.391	0.504	-0.23	-0.87	0.687

One of the techniques of factor analysis is the Principal Component analysis. This technique is helpful to identify the common factor which is sufficient to explain 90% or more variation in the data set. During analysis, we have one analytical result known as Communality, which is the sum of the squared factor loadings for all factors for a given variable. It gives the information that how much of proportion of variation of any x_i is explained by the extracted factor. Due to this character, factor analysis is important to the researchers, specially to the researchers in the field of bio-science. As an example of factor analysis, let us consider the data reported in studying the relationship of level of obesity and socioeconomic variables of a group of adult people of Bangladesh [7]. The variables, except level of obesity, were residence, age, gender, marital status, religion, education, occupation, type of work, income and smoking habit. These variables were sufficient to explain the variation in prevalence of diabetes as KMO = o.633, = 256.371, p-value = 0.000. Here results of the factor analysis were shown below in (Table 1).

From the coefficients of factor–1 it was observed that the variable gender was the most important factor in explaining the variation in the prevalence of diabetes followed by occupation and education. Around 74 percent variation of the variable gender in studying the variation of prevalence of diabetes was explained by the factors which were extracted. Seventy percent variation in the level of education and 66.6% variation in the level of occupation were explained by the factors. From the factor analysis it could be concluded that variation in the level of prevalence of diabetes was mainly for the variation in the levels of gender, education and occupation. The variation in the prevalence of diabetes was also noted due the variation in the level of religion as 85.6% variation of religion was explained by the factors. This factor analysis was satisfactory as no communality was less than 0.40 [8].

Another factor analysis was done using the data presented in another paper [9]. This factor analysis was also satisfactory as no communality [Table 2] was less than 0.40. The analysis had explained 73.098% variation in the data set by 4 factors. From the first factor it was noted that severe obesity among children and adolescents was influenced by the variable father's education followed by mother's education and the time spent by the children and adolescents. The communality related to these three variables indicated that most of the variation in the level of severe obesity could be explained by the factors extracted. Second factor indicated that family income was most responsible variable for the enhancement of level of obesity followed by age of children and their level of body mass index (BMI). The factor analysis had extracted 4 factors which could explain 73.098 % variation in the data set of level of obesity.

The results of another factor analysis are presented here using the data reported in a separate paper [10] in studying the awareness of health hazard of tobacco smoking and smoking habit. The most important variable to explain the variability [Table 3] in the behavior of smoking habit was father's education followed by mother's occupation and father's occupation. The communality related to mother's education indicated that 65.9 percent variation of this variable could be explained by the extracted factors (Table 2 & 3).

Table 2: Important results of factor analysis including factor loadings in analyzing the data of prevalence of diabetes among children and adolescents of Bangladesh.

Variable	Coefficient of Factor - 1	Coefficient of Factor – 2	Coefficient of Factor - 3	Coefficient of Factor - 4	Communality
Father's education	0.884	-0.028	-0.01	0.061	0.786
Mother's education	0.801	-0.257	0.307	0.104	0.813
Father's occupation	0.337	-0.13	0.386	-0.628	0.674
Mother's occupation	0.167	0.54	-0.609	-0.008	0.689
Family income	0.175	0.74	0.375	0.079	0.725
Age of child	0.041	0.585	0.436	0.422	0.713
Time spend	0.559	0.096	-0.63	0.037	0.72
BMI	0.037	-0.556	0.025	0.646	0.727

Table 3: Important results of Factor analysis including Factor loadings in analyzing the data of awareness of health hazard of tobacco smoking and smoking habit.

Variable	Coefficient of Factor - 1	Coefficient of Factor – 2	Coefficient of Factor – 3	Communality
Age	0.233	-0.488	0.085	0.299
Father's education	0.733	-0.032	0.102	0.548
Mother's education	0.214	0.63	0.101	0.453
Father's occupation	0.359	-0.488	0.519	0.636
Mother's occupation	0.792	0.75	0.019	0.659
Family income	-0.389	0.105	0.396	0.369
Residence	-0.295	0.047	0.692	0.568
Gender	0.155	0.6	0.262	0.453

Acknowledgment

None.

Conflict of Interest

No conflict of interest.

References

1. Calamusa G, Amodio E, Costantino C (2012) Body mass index and factors associated with overweight and obesity: A cross-sectional study of adult subjects living in a small city of western Sicily (Italy). Italian Journal of Public Health, 9(3): 1-8.

2. Berg C, Rosengren A, Aires N, Lappas G, Torén K, et al. (2005) Trends in overweight and obesity from 1985 to 2002 in Goteberg, west Sweden. Int J Obes (Lond) 29(8): 916-924.

3. WHO fact Sheet (2018) at http// www.who.int/news - room/factsheets /detal/obesity – and – overweight, dated February.

4. Prasad SB, Fahrtash F, Malaiapan Y, Meredith IT, Cameron J (2009) Obesity and the metabolic syndrome in patients with acute myocardial infraction. Int J Cardiol 144(3): 450-451.

5. Pulgaron ER (2013) Child obesity: a review of increased risk for physical and psychological combordities. Clin Ther 35(1): 18- 32.

6. Mokdad AH, Ford ES, Bowman BA, Dietz WH, Vinicor F, et al. (2001) Prevalence of obesity, diabetes, and obesity related health factors. JAMA 289(1): 76-79.

7. Bhuyan KC, Md Mortuza A, Jannatul F (2018) Discriminating patients suffering from non-communicable diseases: A case study among Bangladeshi adults. Biomed J Sci & Tech Res 10(1): 1-9.

8. Kritika J (2018) Value of Communality, Amity Business School.

9. Urmi AF, Bhuyan KC (2018) Obesity in children and adolescents and the factors responsible for it: A case study among children and adolescents of some affluent families. Integrative Diabetes and Cardiovascular Diseases 2(1): 56-66.

10. Mahfuza K, Bhuyan KC (2014) Awareness of health hazard of tobacco consumption among students of American International University. AJSE 13 (1): 85-91.