



# Journal of Applied Sciences

ISSN 1812-5654

**science**  
alert

**ANSI***net*  
an open access publisher  
<http://ansinet.com>

## Factor Analysis Evidence in Describing Consumer Preferences for a Soft Drink Product in Malaysia

Lazim Abdullah and Hasliza Asngari

Department of Mathematics, University of Malaysia, Terengganu, Malaysia

---

**Abstract:** Soft drink products are one of the many popular beverages that easily available in today's market. Somehow, it is quite ambiguous to reason out what factors that influenced the popularity of the products. This paper focuses on factor analysis model and its application to identify consumer preferences for a popular soft drink product in Malaysia. Factor analysis is used to extract factors in consumer preferences and items' loading factors with data obtained from a survey. The survey conducted via in-person questionnaires at small district in Peninsular Malaysia. Factor analysis results showed that the consumers' preferences were characterized by a four-factor: branding, validation and prices, packaging and taste, respectively. Rotational factors successfully extracted the factor of branding as the dominant factor. Therefore, factors analysis successfully extracted the factors that play enormous role in elevating the products particularly in the Malaysian market.

**Key words:** Latent variables, marketing, soft drink, loading factors

---

### INTRODUCTION

It is becoming increasingly difficult to ignore the existence of soft drink in today's markets. Since the inception of soft drink in the 1830's, its consumption has steadily increased with technological advances in production and increased product availability. Recent statistics from the United States Department of Agriculture reported a per capita increase in regular soft drink consumption from 28 gallons per person in 1986 to 41 gallons per person in 1997 (Yule, 2002). In Malaysia, with the same trend it was reported that the number of people consumes soft drink is very high up to 1000 cans per minute (Bemama, 2007). Soft drink products have been well accepted by consumers and gradually overtaking hot drinks as the biggest beverage sector in the world. In the midst of the rapidly growing soft drink demand, the industry on the whole is encountering new opportunities and challenges. Changing consumer demands and preferences require new ways of maintaining current customers and attracting new ones. Amid ever-increasing competition, beverage companies must intensely court customers, offer high quality products, efficiently distribute them, ensure safety and keep prices low all while staying nimble enough to exploit new markets by launching new products.

In order to survive in this environment, manufacturers must consider the market trends that will likely shape the industry over the next few years. This will

help soft drink companies to understand the challenges they will encounter and to turn them into opportunities for improvement, enhanced flexibility and, ultimately, greater profitability. Market trends for the soft drink industry can be characterized by six fundamental themes (Doilette, 2007). Out of six themes, there is one that directly focuses to consumers preferences. This is the one out of many challenges that has to be faced by manufacturers. Recent developments in soft drink consumption and challenges in marketing have heightened the need for searching the consumers' needs and preferences. In food industry, taste and other sensory characteristics of foods occupy a key position. Taste has to be faultless since it strongly influences food choices (Arvola and Tuorila, 1999), in many cases surpassing health issues (Glanz *et al.*, 1999; Tepper and Trail, 1998). However, it is still unclear whether these factors also affect consumers' acceptability of soft drink.

Consumer preferences have been studied across food categories from beverages to fast food with a number of methods. In a study conducted by Nelson *et al.* (2005), three attributes of roasted peanuts were evaluated using conjoint analysis. The three predetermined attributes were dry-honey roasted, country of origin and price. In another study to understand consumer attitudes and preferences for chocolate milk, again the predetermined attributes such as visual, flavour and mouth feel were differentiated between a group of adults and children using descriptive statistics (Thompson *et al.*, 2007). Another example of

predetermined attributes of mandarins was studied by Wei *et al.* (2003). The attributes include the appearance, taste, texture and overall quality of fruit segments and skin colour were considered to determine consumer preferences and compared between domestic and imported mandarins. All these studies used predetermined attributes or factors to determine consumer preferences. Furthermore, statistical approaches mainly descriptive and inferential statistics were tabled and tested to meet the objectives. The problem of determining factors that influenced consumer preferences has been given somewhat less attention. One of the methods in multivariate analysis that meant for extracting variables or attributes is factor analysis. Thus, the present study focuses on statistical evidence provided by factor analysis in describing consumer preferences with respect to the choices of a soft drink product at a small town in Malaysia. These latent variables will help us to identify consumer preferences of a soft drink product.

**FACTOR ANALYSIS MODEL AS MEASURING CONSUMER PREFERENCES**

Factor analysis is a set of techniques for determining the extent to which variables that are related can be grouped together so that they can be treated as one combined variable or factor rather than as a series of separate variables. Perhaps the most common use of factor analysis in the social and behavioural sciences is to determine whether the responses to a set of items used to measure a particular concept can be grouped together to form an overall index of that concept (Duncan, 2003). Factor analysis is often used in the empirical research in social sciences (Harman, 1976; Kim and Mueller, 1990; Hatcher, 1994). Political scientists, when comparing the attributes of nations in terms of a variety of political and socio-economic variables, have applied factor analysis in an attempt to determine characteristics that are the most important in classifying nations (Rummel, 1979). Alternatively, sociologists have determined friendship groups' by examining which people associate most frequently with each other (Asher, 1976). In addition, psychologists have used this statistical technique so as to study a given individual's intelligence dimensions (Thomson, 1951) and to assess how people perceived different stimuli and categorize them into different response sets (Stukat, 1958). Finally, economists have used factor analysis in the study of consumer behavior, namely in assessing the individual consumer living standards and individual consumer charity behaviour

(Schokkaert and van Ootegem, 1990). These are among the examples of the broad range applications of factor analysis. A brief theoretical view of factor analysis model is relevant in explaining how factors can be extracted from considerable numbers of latent variables.

**FACTOR ANALYSIS MODEL**

Factor analysis begins with number of variables  $X_1, X_2, \dots, X_p$ .

$$\begin{aligned} X_1 &= 1_{11}\lambda_1 + 1_{12}\lambda_2 + 1_{1m}\lambda_m + \ell_1 \\ X_2 &= 1_{21}\lambda_1 + 1_{22}\lambda_2 + 1_{2m}\lambda_m + \ell_2 \\ &\dots \\ &\dots \\ X_p &= 1_{p1}\lambda_1 + 1_{p2}\lambda_2 + 1_{pm}\lambda_m + \ell_p \end{aligned} \tag{1}$$

Equation 1 can be simplified in matrix form:

$$\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix} = \begin{pmatrix} 1_{11} & 1_{12} & 1_{1m} \\ 1_{21} & 1_{22} & 1_{2m} \\ \vdots & \vdots & \vdots \\ 1_{p1} & 1_{p2} & 1_{pm} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_m \end{pmatrix} + \begin{pmatrix} \ell_1 \\ \ell_2 \\ \vdots \\ \ell_p \end{pmatrix} \tag{2}$$

where,  $X_1, X_2, \dots, X_p$  are known variables,  $1_{ij}$  is a constant represents loading for i-th and j-th factor.  $\lambda_j$  is j-th factor.

Similarly, the Eq. 2 can be expressed in the matrix notation:

$$x = \Lambda f + e \tag{3}$$

Where:

- $\Lambda$  =  $\{1_{ij}\}$  is a  $p \times k$  matrix of constants, called the matrix of factors loadings
- $f$  = Random vector representing the  $k$  common factors
- $e$  = Random vector representing  $p$  unique factors associated with the original variables

The common factors  $f_1, f_2, \dots, f_k$  are common to all  $X$  variables and are assumed to have mean = 0 and variance = 1. The unique factors are unique to  $X_i$ . The unique factors are also assumed to have mean = 0 and are uncorrelated to the common factors.

Equivalently, the covariance matrix  $\Sigma$  can be decomposed into a factor covariance matrix and an error covariance matrix:

$$\Sigma = \Pi \Pi^T + \Psi \tag{4}$$

where:

$$\Psi = \text{Var} \{u\}$$

$JJ^T$  is the transpose of  $J$

The diagonal of the factor covariance matrix is called the vector of communalities  $h_i^2$  where:

$$h_i^2 = \sum_{j=1}^n \lambda_{ij}^2$$

The factor loadings are the correlation coefficients between the variables and factors. Factor loadings are the basis for imputing a label to different factors. Analogous to Pearson's  $r$ , the squared factor loading is the percentage of variance in the variable, explained by a factor.

The sum of the squared factor loadings for all factors for a given variable is the variance in that variable accounted for by all the factors and this is called the communality. The factor analysis model does not extract all the variance; it extracts only that proportion of variance, which is due to the common factors and shared by several items. The proportion of variance of a particular item that is due to common factors (shared with other items) is called communality. The proportion of variance that is unique to each item is then the respective item's total variance minus the communality.

The solution of Eq. 4 is not unique (unless the number of factors = 1), which means that the factor loadings are inherently indeterminate. Any solution can be rotated arbitrarily to obtain a new factor structure. The goal of these rotation strategies is to obtain a clear pattern of loadings, i.e., the factors are somehow clearly marked by high loadings for some variables and low loadings for other variables.

Rotation factors can be found as follows:

$$\begin{aligned} Z_1 &= b_{11} z_1 + b_{12} z_2 + \dots + b_{1p} z_p \\ &\dots \\ &\dots \\ Z_p &= b_{p1} x_1 + b_{p2} z_2 + \dots + b_{pp} z_p \end{aligned} \tag{5}$$

$\therefore b_{ij}$  is an element of eigenvector in correlation matrix.

Due to orthogonal transformation from X-score to Z-score, the Eq. 5 becomes:

$$\begin{aligned} X_1 &= b_{11} x_1 + b_{12} x_2 + \dots + b_{1p} x_p \\ &\dots \\ &\dots \\ X_p &= b_{p1} x_1 + b_{p2} x_2 + \dots + b_{pp} x_p \end{aligned} \tag{6}$$

With only extracting  $m$  principle components, Eq. 6 yields:

$$\begin{aligned} X_1 &= b_{11} z_1 + b_{12} z_2 + \dots + b_{m1zm} - e_i \\ &\dots \\ &\dots \\ X &= b_{1p} z_1 + b_{2p} z_2 + \dots + b_{m1zm} - e_p \end{aligned} \tag{7}$$

$\therefore e_i$  is a linear combination of principle components  $Z_{m+1}$  to  $Z_p$ .

$\therefore z_1$  is a new factor and  $b_{ij}$  is a new loading factor.

The eigenvalue for a given factor reflects the variance in all the variables, which is accounted for by that factor. A factor's eigenvalue may be computed as the sum of its squared factor loadings for all the variables. The ratio of eigenvalues is the ratio of explanatory importance of the factors with respect to the variables. If a factor has a low eigenvalue, then it is contributing little to the explanation of variances in the variables and may be ignored.

In finding eigenvalue, consider the Eq. 8:

$$\begin{aligned} A_1 &= a_{11} x_1 + a_{12} x_2 + \dots + a_{2n} x_n \\ A_2 &= a_{21} x_1 + a_{22} x_2 + \dots + a_{2n} x_n \\ &\dots \\ &\dots \\ A_p &= a_{p1} x_1 + a_{p2} x_2 + \dots + a_{pn} x_n = \lambda x_n \end{aligned} \tag{8}$$

where,

$$\begin{aligned} (A - \lambda I)x &= 0 \text{ or } A_p = \lambda x_n \\ \therefore I &\text{ is an identity matrix } n \times n \\ \therefore 0 &\text{ is a zero vector with } n \times 1 \text{ dimensions} \end{aligned}$$

By finding the determinant of matrix  $(A - \lambda I)x = 0$ , then  $\lambda$  can be determined and it is an eigenvalue of  $A$ . An empirical example is presented to find eigenvalue, loading factors and new factors.

### AN EMPIRICAL STUDY

In this empirical study, an application of factor analysis to identify consumers' preferences with respect to the choices of a soft drink will be explored. In other words, understanding and listing of consumer valuation for the choices of a soft drink are proposed. The present

survey instrument contains a list of 23 attitudinal items were given out to 190 respondents at Sabak Bernam, Selangor, Malaysia. Each item was represented by a sentence as presented in the instrument survey. The respondent expresses her opinion towards a soft drink product by classifying each sentence using a five point semantic differential or Likert scale: I completely disagree, I disagree, Sometimes I agree, sometimes I disagree, I agree and I completely agree. Formally, the factor analysis model that confined to the following steps was employed. In the empirical work, two major analyses steps were taken. In the first analysis an extraction of the initial common preferential factors were performed. In the second analysis, a loading factor for each common preferential factor was provided. The Eq. 1-8 were used in the two analyses. Calculations were made with the used of SPSS 11.0.

**EXTRACTION OF THE INITIAL COMMON PREFERENTIAL FACTORS**

The first step in the empirical analysis consists in extracting of a latent preferential construct with factors. Since the object of factor analysis is to reduce the number of variables that to be handled, this would not be achieved if all of them were used. Consequently, the next step is to decide how many factors should be kept. This really is a question of how many of the smaller factors should be retained, since the first few which explain most of the variance would be kept. The first factor will always explain the largest proportion of the overall variance; the second factor will explain the next largest proportion of variance that is not explained by the first factor and so on, with the last factor explaining the smallest proportion of the overall variance. Each variable is correlated with or loads on each factor. Because the first factor explains the largest proportion of the overall variance, the correlations or loadings of the variables will, on average, be highest for the first factor, next highest for the second factor and so on. To calculate the proportion of the total variance explained by each factor, the loadings of the variables on that factor is squared, add the squared loadings to give the eigenvalue or latent root of that factor and divide the eigenvalue by the number of variables. As there are as many components as variables, some criterion needs to decide how many of the smaller factors should be ignored, as these explain the least amount of the total variance. One of the main criteria used is the Kaiser or Kaiser-Guttman criterion, which is that factors that have eigenvalues of one or less should be ignored. As the maximum amount of variance that can be explained by one variable is one, these factors effectively account for no

Table 1: Kaiser test

Factor	Eigenvalues ( $\Delta$ )
1	7.020
2	1.935
3	1.707
4	1.529
5	1.103
6	0.983

more than the equivalent of the variance of one variable. The rationale for the Kaiser test is as follows: each observed variable contributes one unit of variance to the total variance in the data set. Thus, any component that displays an eigenvalue greater than one accounts for a greater amount of variance than had been contributed by one variable. In addition, a component that displays an eigenvalue less than one accounts for less variance than had been contributed by one variable.

Since, the purpose of the factor analysis is to assess a number of reduced components (or factors), this cannot be effectively achieved if one retains factors that account for less variance than had been contributed by individual variables. According to the estimation results, it is sufficient to retain five factors since  $\Delta_6 = 0.983 < 1$  (Table 1).

Furthermore, the pattern of loadings of the retained five-factor model shows a common conceptual meaning.

**ROTATIONAL FACTORS**

In the first step of factor analysis, the extraction (or identification) of the preferential factors is preceded. At this stage the number of factors to be rotated is not specified. Since the final objective is the reproduction of the covariances and correlations, the sample correlation matrix of rotated factor loadings as the primary data to be used in the analysis is accounted. Estimation results are presented in Table 2. Printed results are multiplied by 100 and rounded to the nearest integer. The values of below 0.50 are omitted (Lazim *et al.*, 2005; Lazim and Kamaludin, 2007). This matrix represents the product-moment correlation between the observable variable and the underlying factor. The factor loadings are analogous to the standardized regression coefficients as obtained in regression analysis. In other words, dropping an attitudinal item that does not score above 0.50 (assumed as the minimum correlation bound) means that an exploratory rule based on the magnitude of the estimated regression coefficients, which is characterized by rejecting all the items which indicate low correlations with the common factors are followed.

As can be seen from Table 2, the factor model specification is characterized by five common factors: Factor 1, 2, 3, 4 and 5. A closer inspection to the

Table 2: Rotated factor pattern (five common factors)

Item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
M7	68.9				
M10	68.7				
M9	65.0				
M15	64.0				
M19	59.4				
M16	54.0				
M22	52.0				
M21		79.3			
M18		74.2			
M20		70.2			
M17		67.3			
M4			82.6		
M3			80.7		
M2			76.5		
M13				86.5	
M1				75.6	
M12				66.9	
M23					75.0
M6					62.7
M5					55.9

attitudinal items polled under the factor 5 reveals that they do not share any common conceptual meaning: M23 focuses on the trend in choosing soft drink of the days; M6 and M5 focus on the accessibility in the market. The M23, M6 and M5 are interpreted as not being pure measures underpinning any common latent construct. Therefore, whether a four-factor model performs an adequate representation of the data is tested.

Factor 1 collects a number of items related to the respondent's general attitude with respect to promotions and popularity of the product. Therefore this latent construct can be labelled as branding. A higher number of items in this factor indicate a strong trustworthiness to brand.

Factor 2 collects a number of items that are related to respondent's belief toward certification by authority and also confirmation of prices that comparable to other brands. Therefore this factor is labelled as the validation and prices. A high score on this factor reveals that the respondent experiences good value for money that makes them feel happy with the product.

Factor 3 collects a number of items that related to the physical appearance of the product. For this reason the third factor is labelled as the packaging. Respondents who score high on packaging have strong beliefs with respect to the well presented and palatable product.

Factor 4 collects two items that related to the tasty of the product. Item M13 and item M1 focus on the taste of the product. Therefore, this factor is labelled as taste. A higher score in the item M13 of thirst quencher was due to the strong belief consumers in taste of the product.

These statistical evidences provide a robust tool in identifying four conceptual factors and loading for each item in describing consumers preferences toward choice of a soft drink.

## CONCLUSION

The results successfully showed the application of factor analysis model in extracting new factors together with their respective loading factors. These results have important implications not only to manufacturers and marketers but also to the researchers. First, one can conclude that any policy action in marketing strategy must consider the four factors. Secondly, the evaluation results clearly indicate that the Malaysian customers prefer to keep branding as the first priority. The other factors are also important but with a slight reduction in the degree of importance. These results can positively be transferred to other areas of marketing policy. Indeed, identification and characterization of consumer preferences are expected to play a crucial role in stimulating consumer awareness for worth value products. Summarily factor analysis model have proven its applicability in identifying factors influencing customers and surely could be extended into other areas where factors extracting is highly needed.

## REFERENCES

- Arvola, L.L. and H. Tuorila, 1999. Predicting the intent to purchase unfamiliar and familiar cheeses: The effects of attitudes, expected liking and food neophobia. *Appetite*, 32: 113-126.
- Asher, H.B., 1976. *Causal Modeling*. 2nd Edn., Sage Publications, Sage University Paper Series on Quantitative Applications in the Social Sciences, Beverly Hills, CA.
- Bernama, 2007. Rakyat Malaysia Tidak Jaga Makan. <http://www.bernama.com/bernama/v3/bm/news.php?id=284195>.
- Doilette, 2007. Profitable Growth and value creation in the soft drink industry. <http://www.deloitte.com/dtt/cda/doc/content>.
- Duncan, C., 2003. *Advanced Quantitative Data Analysis*. 1st Edn., GBR, McGraw-Hill Education, Berkshire, ISBN: 978-0-335-20059-7.
- Glanz, M., E. Basil, J. Maibach Goldberg and D. Snyder, 1999. Why Americans eat what they do: Taste, nutrition, cost, convenience and weight control concerns as influences on food consumption. *J. Am. Diet Assoc.*, 98: 1118-1126.
- Harman, H.H., 1976. *Modern Factor Analysis*. Chicago University Press, Chicago, ISBN: 0226316521.
- Hatcher, L., 1994. *A Step-by-Step Approach to Using the SAS System Factor Analysis and Structural Equation Modeling*. SAS Publishing, Cary, USA., ISBN: 1555446434.

- Kim, J.O. and C.W. Mueller, 1990. Factorial Analysis: Statistical Methods and Practical Issues, Quantitative Applications in Social Sciences. 2nd Edn., Sage University Papers Series, Beverly Hills, CA.
- Lazim, M.A, W.A. Wan Salihin and M.T. Abu Osman, 2005. A new look at the students' attitudes toward scientific calculators. *Malaysian Online J. Instruct. Technol.*, 2: 17-26.
- Lazim, M.A. and A.S. Kamaludin, 2007. The perceptions of Malaysian undergraduate students about a set of generic skills. *J. Institut. Res. South East Asia*, 5: 15-26.
- Nelson, R.G., C.M. Jolly, M.J. Hinds, Y. Donis and E. Prophete, 2005. Conjoint analysis of consumer preferences for roasted peanut products in Haiti. *Int. J. Consumer Stud.*, 29: 208-215.
- Rummel, R.J., 1979. *Applied Factor Analysis*. Northwestern Amsterdam University Press, Amsterdam, ISBN: 0-810102544.
- Schokkaert, E. and L. van Ootegem, 1990. Sen's Concept of the Living Standard Applied to the Belgium Unemployed. *Economic Research of Louvain, University* © Catholique de Louvain, USA.
- Stukat, K.G., 1958. *Suggestibility: A Factorial and Experimental Analysis*. 2nd Edn., Stockholm, Wiksell.
- Tepper, B.J. and A.C. Trail, 1998. Taste or health: A study on consumer acceptance of corn chips. *Food Qual. Pref.*, 9: 267-272.
- Thompson, L., P.D. Gerad and M.A. Drake, 2007. Chocolate milk and the hispanic consumer. *J. Food Sci.*, 72: 666-675.
- Thomson, G.H., 1951. *The Factorial Analysis of Human Ability*. 5th Edn., University of London Press, London.
- Wei, S., S. Singgih, E.J. Woods and D. Adar, 2003. How important is appearance? Consumer preferences for mandarins in Indonesia. *Int. J. Consumer Stud.*, 27: 406-411.
- Yule, A., 2002. Increased soft drink consumption is contributing to an increased incidence of obesity. *Nutr. Bytes*, 8: 1-7.