Understanding costumer satisfaction with fixed internet services in Mexico: a factor analysis using polychoric correlations

José Alberto Candelaria Barrera
jose.candelaria@ift.org.mx. Research Centre, Mexican Federal Institute of Telecommunications, Mexico City[1]

Abstract

The aim of this research consists on determine the factors that explain costumer satisfaction in the Mexican household fixed Internet service industry. The impossibility to measure directly our variable of interest entails the use of a factor analysis methodology which examines latent variables indirectly by employing measurable variables that are believed to be indicators of customer satisfaction. Data gathering is based on an anonymous questionnaire survey of 1481 adults interviewed by telephone conducted by a private consultant by request of the Mexican Federal Institute of Telecommunications on February 2016. The survey includes eight questions that gauge respondent's perceptions about the quality of services offered by his internet service provider. Subsequently, those opinions are coded on Likert scale of scores $1-5$. These ordered-category data require the use of Polychoric correlation in order to find patterns between the indicators that allows us to infer their relationship to our unknown latent costumer satisfaction variable. Futher, the data gathered includes demographic characteristics of individuals and detailed information of internet service providers' characteristics, such as internet connection speed. The factor analysis was conducted with varimax rotation to obtain variable loads onto one factor as high as possible. The findings suggest the prevalence of two main factors: the first one related with technical features of the internet service and the second one with the service features provided by the internet service company.

Keywords: Factor analysis, Polychoric correlation, Customer satisfaction, Internet service industry.

[1]**Disclaimer:** The content of this article is of the author's sole responsibility. It does not represent the views of the Instituto Federal de Telecomunicaciones (IFT) or its staff. Any inquiries related to the article's content and all permission to cite should be directed to the author.

JEL Classification:

Part I

Introduction

Internet market is a fast growing segment in Mexico's telecommunications sector. In 2020, the fixed broadband household penetration is projected to reach the 51.09 percent [2]. Further, it is predicted that fixed broadband subscriptions will surpass the 19 millions, with annual net additions ranging from slightly more than 1 million by 2017 to 737 thousand by 2020. Under this optimistic scenario in telecommunications sector, it acquires great relevance the analysis of consumers's opinion about perceived quality and satisfaction with household fixed internet service. Additionally, telecommunications sector firms have realized that an analysis of consumer's perceived satisfaction is a necessary condition in order to guarantee their own continuity and development. This situation is linked with company's image, brand positioning and the valuation of intangible property (Villafañe, 2003). Also, it is worth mentioning that mexican consumers are becoming more well-informed, therefore it is expected an increased demand of quality services.

Mexican Federal Telecommunications and Broadcasting Law published on July 14, 2014 (Ley Federal de Telecomunicaciones, 2014), establishes the responsabilities of Mexican Federal Telecommunications Institute (hereafter, Institute) and its composition through Article 15, fraction $XLVII$. Among them, the Institute must establish quality targets by service to which telecommunications carriers must adhere. Furthermore, Article 145, fraction IV settles a normative for all Internet carriers and operators with regards to the preservation of minimum quality standards. Thus, a proposal of a factor analysis approach that explains consumers' perceived satisfaction with household fixed internet, acquires a great relevance since it extends the analysis beyond universal Internet service coverage. Also, it recognizes that the degree of consumers' perceived quality is a key element in the development of telecommunications sector.

On the other hand, the issue is a recurrent field of research in other industrial sectors. For instance, Manisera *et al.* (2010) identify the components of job satisfaction in Italy by analyzing survey data and applying a non-linear principal components methodology. Other authors, such as Brough and Frame (2004) test the influence of organizational and individual variables, such as sexual harassment and the perception of social support, respectively, on job satisfaction and labor mobility in New Zealand police department. The authors use a factor analysis and a structural equation model (SEM). The medical field is another sector where is usual to find statistical analysis with

[2] https://www.ovumkc.com/

latent variables in order to study the satisfaction with a service. In this respect, Beattie *et al.* (2002) analyze patient's satisfaction with outpatient physical therapy in four regions of the United States. Their conclusions indicate that patient's satisfaction is mainly associated with the perception about physical therapist´s professionalism and the time dedicated to their rehabilitation. In a much lesser extent patients appreciate the location of medical facilities, parking or the type of medical equipment.

Other works, such as the one of Lee *et al.* (2001) focuses on the analysis of the link between users' satisfaction and their loyalty in the telecommunicatins sector, specifically for the market of mobile telephony in France. These authors use a principal components analysis and they deduce that the conecction between users's satisfaction and loyalty is a function of market regulation, loyalty programs, product differentiation, proper technology and switching costs. Furthermore, Yu and Dean (2001) extend the analysis by proposing a factor analysis of an emotional component. They found that both of them, users' positive and negative emotions have a significant effect on the explanation of loyalty to a given service. Additionally, in the restaurant industry it has been verified the existence of a strong relationship between both variables (users' satisfaction and loyalty). Some authors as Saad Andaleeb and Conway (2006) conducted a factor analysis with a varimax rotation to investigate consumers' satisfaction in the American full service restaurant industry. Their conclusion is that the more weighted factor on consumers' satisfaction is the quality of service offered, prevailing over the prices and food quality.

Part II

Factor Analysis Model

An unobserved variable of interest such as *customer satisfaction* can be defined as a *latent variable*, *i.e.* it can only be indirectly measured. In this regard, factor analysis is a mathematical technique used to analyze patterns of correlation between indicators known as *manifest* or *observed variables* in order to infer their relationship to *latent variables*. The model specifies the observed measured variables as function of unobserved common factors. We then have the following multiple linear regression model :

$$x_1 = \lambda_{11} f_1 + ... + \lambda_{1k} f_k + u_1$$

$$x_2 = \lambda_{21} f_1 + ... + \lambda_{2k} f_k + u_2$$

$$\vdots = \qquad \vdots$$

$$x_p = \lambda_{p1} f_1 + ... + \lambda_{pk} f_k + u_p$$

Where $x = (x_1, ..., x_p)'$ are the observed variables; $f = (f_1, ..., f_k)'$ are the common factors; $u = (u_1, ..., u_p)'$ represent the specific factors and λ_{ij} are the factor loadings. The matricial form is represented as:

$$x = \Lambda f + u \qquad (1)$$

WhereΛ is a $p \times k$ matrix containing the $\lambda'_{ij}s$. While, the covariances of the specific factors u and of the observed variables x, are expresed in the following form:

$$Cov\,(u) \equiv \Psi = diag\,(\psi_{11}, ..., \psi_{kk})$$

$$Cov\,(x) = \Sigma$$

The Ψ term represents the diagonal matrix of specific or *unique* variances, *i.e.* it represents the variance of variable x_i that is not shared with the rest of variables. Furthermore, the variance of x_i is composed of two parts, which are the following:

$$Var\,(x_i) = \sum_{j=1}^{k} \lambda_{ij}^2 + \psi_{ii}$$

The term $\sum_{j=1}^{k} \lambda_{ij}^2$ is known as the *communality* of x_i. It represents the variance of x_i that is shared with the rest of variables through the common factors.

On the other hand, if G is a $k \times k$ orthogonal matrix, we have the following:

$$x = \Lambda G G' f + u \qquad (2)$$

Now, both models (1) and (2) are indistinguishable, because f and $G'f$ are not correlated in their components. It means that both models are equivalent in order to explain the covariance matrix of the observed variables. Thus, the covariance matrix of variable x has the following structure:

$$\Sigma = (\Lambda G)\,(G'\Lambda') + \Psi = \Lambda\Lambda' + \Psi$$

This expression implies that the covariance matrix of the observed variables admits a decomposition as a sum of two matrices (Peña, 2002). Moreover, the covariance matrix Σ is usually estimated by the sample covariance matrix S. While, the correlation matrix is estimated by the sample correlation matrix, R. Thus, it can be used S or R in order to estimate $\hat{\Lambda}$ and $\hat{\Psi}$, such that $S\,(or\ R) \approx \hat{\Lambda}\hat{\Lambda}' + \hat{\Psi}$. Tyically, the number of parameters in $\hat{\Lambda}$ and $\hat{\Psi}$ is smaller than the number of parameters in S.

On the other hand, the two main methods to estimate $\hat{\Lambda}$ and $\hat{\Psi}$ are the iterated principal factors method and maximum likelihood estimation. Additionally, it is necessary to determine the value of k; the number of factors. In this regard, there is some controversy about the procedure to select the number of factors. Some authors consider that there is no need to find the correct number of them but the number that is worth conserving (Cattell, 1996). On this subject, a recognized selection rule of factors is the criterion of *screeplot,* which is useful in order to determine the number of significant factors. Another popular criterion is the percentage of variance explained by determined number of factors.

Part III

Data

The subjects in this study were $1,481$ survey respondents from 31 Mexican federal entities. The survey, "Encuesta 2016. Usuarios de Servicios de Telecomunicaciones" pose a series of questions in order to obtain survey respondents' opinion regarding their perception of the quality of household fixed internet services offered by their Internet service provider. The number of questions are eight (view Table 1) which range from the perception of quality of download speed of data files, to the clarity of billing or customer service. Survey respondents are able to choose among five possible answers: *i*) Very Good, *ii*) Good, *iii*) Neither Good nor Bad, *iv*) Bad, *v*) Very Bad.

	Survey Questions
P33_1	How would you rate the quality of services offered by your internet provider regarding ...? Download speed of data files
P33_2	How would you rate the quality of services offered by your internet provider regarding ...? Upload speed of data files
P33_3	How would you rate the quality of services offered by your internet provider regarding ...? Quality video reproduction
P33_4	How would you rate the quality of services offered by your internet provider regarding ...? Clarity of billing
P33_5	How would you rate the quality of services offered by your internet provider regarding ...? Internet without interruptions
P33_6	How would you rate the quality of services offered by your internet provider regarding ...? Internet performance during peak hours
P33_7	How would you rate the quality of services offered by your internet provider regarding ...? Internet tariff plans information provided by the internet service provider
P33_8	How would you rate the quality of services offered by your internet provider regarding ...? Customer service

Table 1. Survey questions

These type of categorical answers can be measured on a five point Likert scale: 5−Very Good, 4−Good, 3−Neither Good nor Bad, 2−Bad and 1−Very Bad. Likert (1932) developed a procedure for measuring attitudinal scales that facilitates survey construction, data coding and analysis. Likewise, Likert's response categories are useful in order to capture the intensity of survey respondent's feelings toward the question.

In exploratory factor analysis Pearson correlation matrices are commonly used in order to obtain factor solutions. However, the method is not convenient when studying the degree of association between categorical variables. The reason is that Pearson correlation assumes that variables are

continuous and follow a multivariate normal distribution. Further, Guilley and Uhlig (1993) state that the use of Pearson correlations would impose artificial restrictions derived by the categorization. Then, if our model includes variables that are ordered-category a factor analysis can be performed using a Polychoric correlation matrix . The Polychoric correlation is considered as an excellent way to measure rater agreement (Uebersax, 2015). Additionally, Polychoric correlation estimate the correlation between raters as if ratings were made on a continuous scale.

Part IV

Results

The Polychoric correlation between variables is shown in Table 2. As we can observe there exists a strong correlation between the variables that rate the quality of the download speed of data files (P33_1) and the quality of upload speed of data files (P33_2) (0.7781) . Also, there is a strong correlation between variable P33_1 and the variable that rate the quality of video reproduction (P33_3) (0.6348); and between variables P33_2 and P33_3 (0.6455). Moreover, correlations between variables P33_1 and P33_2 with respect to the variables that rate the quality of internet without interruptions (P33_5) and the internet performance during peak hours (P33_6) are all positive and above 0.60. The table also reports a correlation coefficient of 0.7293 between variables P33_5 and P33_6.

	P33_1	**P33_2**	**P33_3**	**P33_4**	**P33_5**	**P33_6**	**P33_7**	**P33_8**
P33_1	1							
P33_2	0.7781	1						
P33_3	0.6348	0.6455	1					
P33_4	0.5335	0.4198	0.4968	1				
P33_5	0.6353	0.6311	0.5199	0.4354	1			
P33_6	0.6081	0.6383	0.5405	0.4223	0.7293	1		
P33_7	0.5438	0.5584	0.5873	0.5607	0.5067	0.5518	1	
P33_8	0.5243	0.5057	0.4914	0.5154	0.5498	0.5038	0.6337	1

Table 2. Polychoric Correlation Matrix

Choosing the optimal number of factors in our exploratory factor analysis is done using the criterion of cumulative variance explained. Factors 1 and 2 jointly account for the 71.59 percent of total variability (view Table 3). The retention of two factors is confirmed by screeplot criterion (view Figure 1) which suggest the recovery of two factors. Cattell (1966)suggest that the optimal number

of retained factors can be identified when we observe an inflection point in the graph from which the curve is transformed into a relatively straight line

Factor	Eigenvalue	Proportion	Cumulative
Factor 1	4.9409	0.6176	0.6176
Factor 2	0.78611	0.0983	0.7159
Factor 3	0.59	0.0738	0.7896
Factor 4	0.48335	0.0604	0.85
Factor 5	0.4083	0.051	0.9011
Factor 6	0.34054	0.0426	0.9437
Factor 7	0.25077	0.0313	0.975
Factor 8	0.20003	0.025	1

Table 3. Cumulative Variance

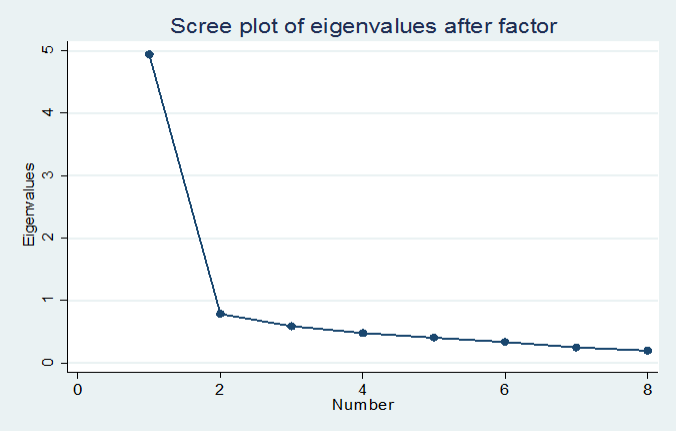

Figure 1. Screeplot

The factor analysis is conducted with a *varimax* rotation. The rotation is practiced because it is useful in order to make interpretation easier of the correlations between the variables and the factors. Then, it is now possible to calculate the factor loadings of the orthogonal solution, which represent how the variables are weighted for each factor. From Table 4 we can observe that variables P33_1, (0.7562) ; P33_2, (0.8102) ; P33_3, (0.5743) ; P33_5, (0.7086) and P33_6, (0.7011) are heavily weighted for factor 1. While, variables P33_4, (0.6021) ; P33_7, (0.7352) and P33_8 (0.6321)are strongly weighted for factor 2. Figure 2 depicts a cluster graph where factor 1 is indicated in X-axe and factor 2 in Y-axe. It is clear that variables P33_4, P33_7 and P33_8 form a cluster that is sloped towards the Y-axe. On the other side, variables P33_1, P33_2,

P33_3, P33_5 and P33_6 form a second cluster that is sloped towards the X-axe. Thus, we can conclude that factor 1 is determined by the variables that are related to technical issues of fixed internet service; while factor 2 is determined by variables related to customer service (tariff plans information, billing clarity, etc). In this respect, we can name both factors according to their characteristics. Then, factor 1 is titled as *satisfaction with technical issues of fixed internet service*. While, factor 2 is named as *satisfaction with the customer service offered by the Internet service provider*. Finally, our coefficient of reliability, the Cronbach's alpha is equal to 0.8494 for the variables that conform factor 1. While, Cronbach's alpha is equal to 0.7288 for the variables that determine factor 2.

Variable	Factor1	Factor2	Uniqueness
P33_1	**0.7562**	0.3748	0.2876
P33_2	**0.8102**	0.2997	0.2538
P33_3	**0.5743**	0.4682	0.4509
P33_4	0.3356	**0.6021**	0.5249
P33_5	**0.7086**	0.343	0.3803
P33_6	**0.7011**	0.3479	0.3874
P33_7	0.3935	**0.7352**	0.3046
P33_8	0.4018	**0.6321**	0.4391

Table 4. Factor analysis of independent variables with varimax rotation (method: iterated principal factors)

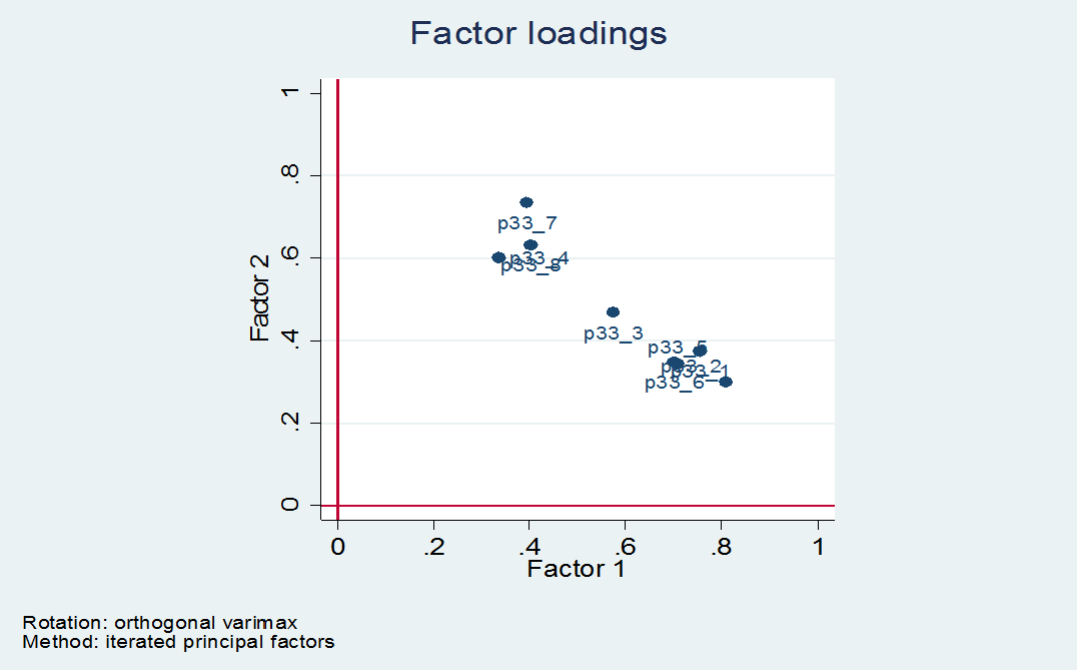

Figure 2. Factor Loadings

Scatter plot of factors 1 and 2 shows that does not exist any positive or negative tendency between both factors, *i.e.* they are independent or orthogonal.

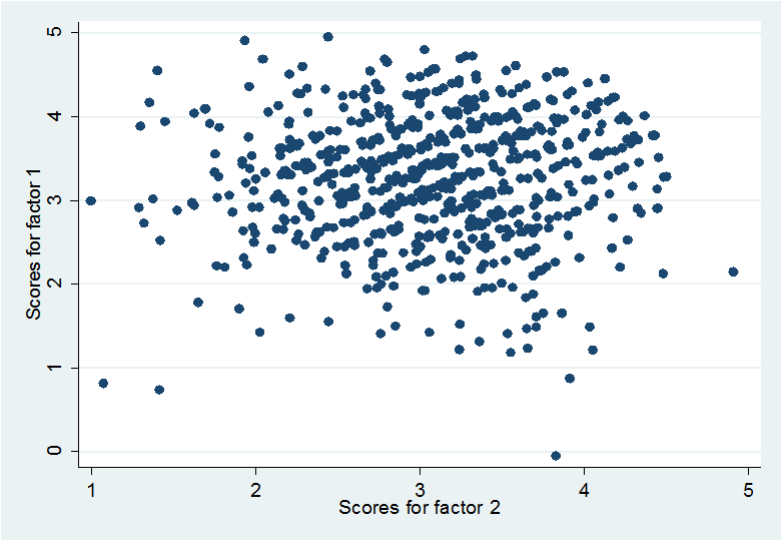

Figure 3. Scatter plot of factors 1 and 2

Finally, factors are scored for each variable (view Table 5), which implies that factors are scaled to have variance equal to 1. Then, we have a mathematical expression for each factor:

$$F1 = 0.27\,P33_1 + 0.45\,P33_2 + 0.05\,P33_3 - 0.07\,P33_4 + 0.22\,P33_5 + 0.22\,P33_6 - 0.18\,P33_7 - 0.07\,P33_8$$

$$F2 = -0.02\,P33_1 - 0.24\,P33_2 + 0.10\,P33_3 + 0.24\,P33_4 - 0.04\,P33_5 - 0.06\,P33_6 + 0.55\,P33_7 + 0.29\,P33_8$$

Variable	Factor1	Factor2
P33_1	0.26952	-0.02153
P33_2	0.45028	-0.24177
P33_3	0.05693	0.10402
P33_4	-0.07596	0.24915
P33_5	0.22893	-0.04908
P33_6	0.22641	-0.06487
P33_7	-0.18738	0.55578
P33_8	-0.07726	0.29354
Scoring coefficients (method = regression; based on varimax rotated factors)		

Table 5. Factor scores

Now, we analyze the dispersion of factors 1 and 2 by using a set of variables included in the survey. These variables are demographic as gender, region, age groups or level of income. While, two other variables are Internet speed (in megabytes) and the Internet service provider. Then, we construct box plots of each factor in order to show the dispersion around the median. Figures 4*a* and 4*b* depict the distribution of factors 1 and 2, respectively, by gender. As we can observe, the minimum value of factor 2 is below the minimum value of factor 1, for each gender; being below a 2 score in the case of males. Also, the maximum value of factor 2 is below the maximum value of factor 1, for each gender. Thus, we can conclude that among both men and women, the factor of satisfaction related with the customer service has lower levels than factor 1, related with the technical issues. The same situation is presented when we analyze the demographic variable, region. Figures 5*a* and 5*b* show that there exists lower levels of satisfaction related with the customer service than with technical issues, among people living in rural and urban regions.

The next box plot depicts the dispersion of factors 1 and 2 by another demographic variable, age groups (view Figures 6*a* and 6*b*). It is clear that there is more dispersion for the second factor. This is especially true in the case of three age groups: $35-44$ years old, $45-55$ years old and $56-65$ years old. Then, we have the dispersion of factors 1 and 2 by income groups. Again, it is notorious that factor 2 exhibits lower minimum values as long as more dispersion than the factor related with technical issues. It is indistinguishable for high or low income levels. For instance, for the ranges from $\$1,601-\$4,000$ Mexican pesos and $\$4,000-\$6,400$ Mexican pesos the minimum score for factor 2 is below a value of 2. While, in the case of factor 1 for the same two ranges, it is above a value of 2.

The last two variables used in our analysis are Internet speed and Internet service providers. We can observe that especially in the case of high Internet speeds (from 21 to 60 megabytes and from

61 to 100 megabytes) the dispersion increases when we analyze the satisfaction with customer services offered by the Internet service provider. Additionally, if we pay attention to Telmex as the largest Internet service provider in Mexico, it is clear that its minimum and maximum values for factor 2 are below the ones of factor 1 (view Figures 9*a* and 9*b*). All the evidence indicates that customer satisfaction with household fixed Internet services in Mexico shows differences when we focus in two main aspects, satisfaction with technical issues and satisfaction with the customer service offered by the Internet service provider. Further, those differences show that there are lower levels in the scores for factor 2 than with respect to the ones of factor 1.

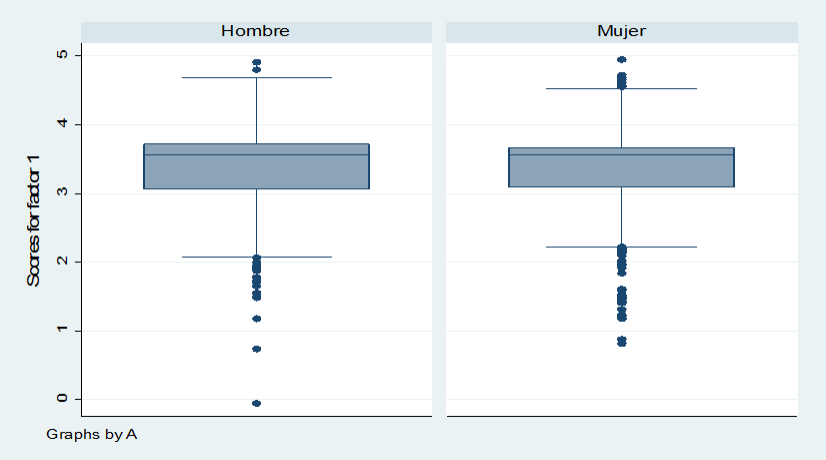

Figure 4a. Box plot of factor 1, by gender.

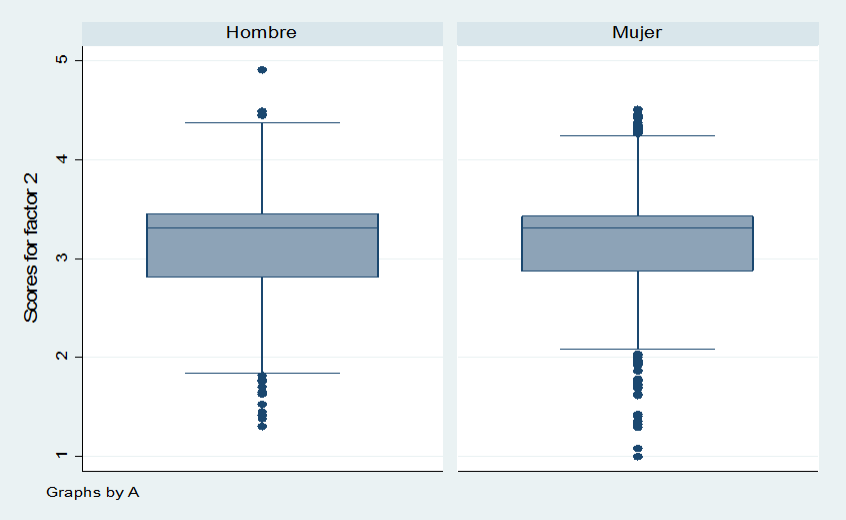

Figure 4b. Box plot of factor 2, by gender.

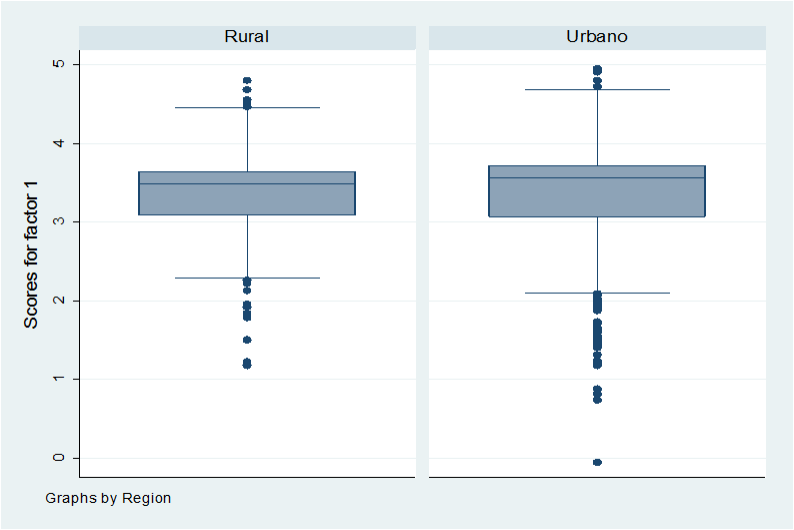

Figure 5a. Box plot of factor 1, by type of region.

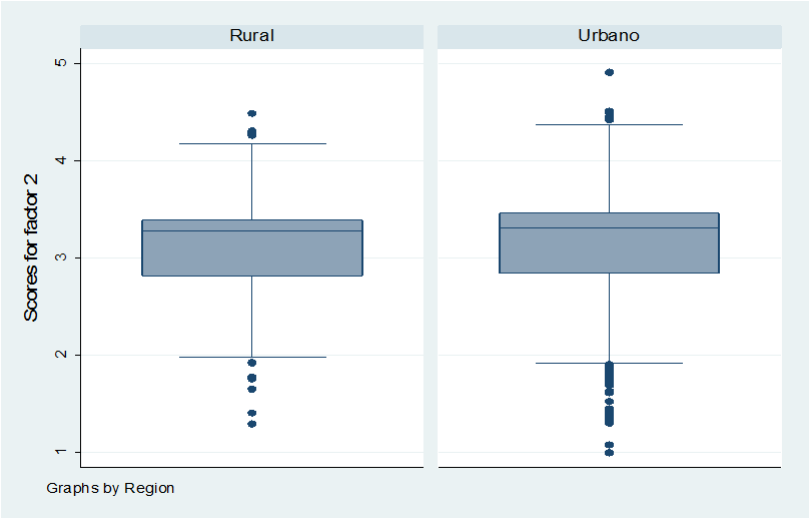

Figure 5b. Box plot of factor 2, by type of region.

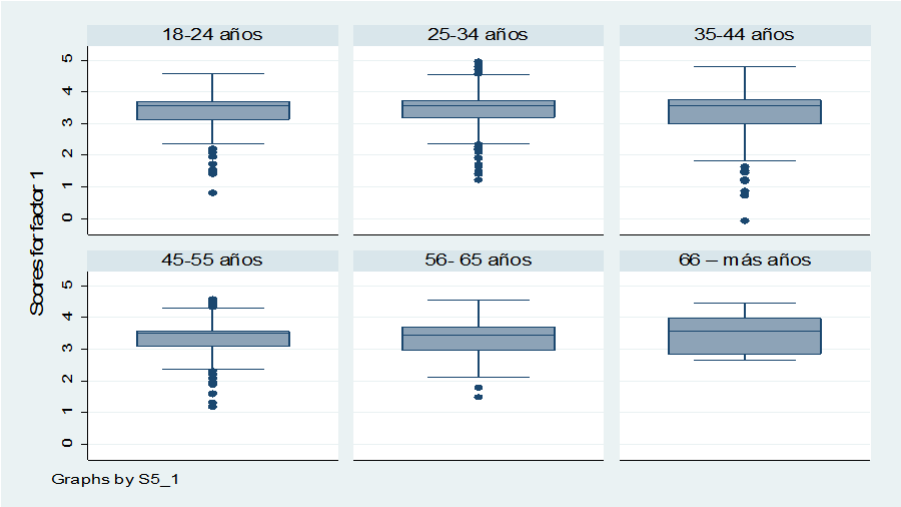

Figure 6a. Box plot of factor 1, by age groups.

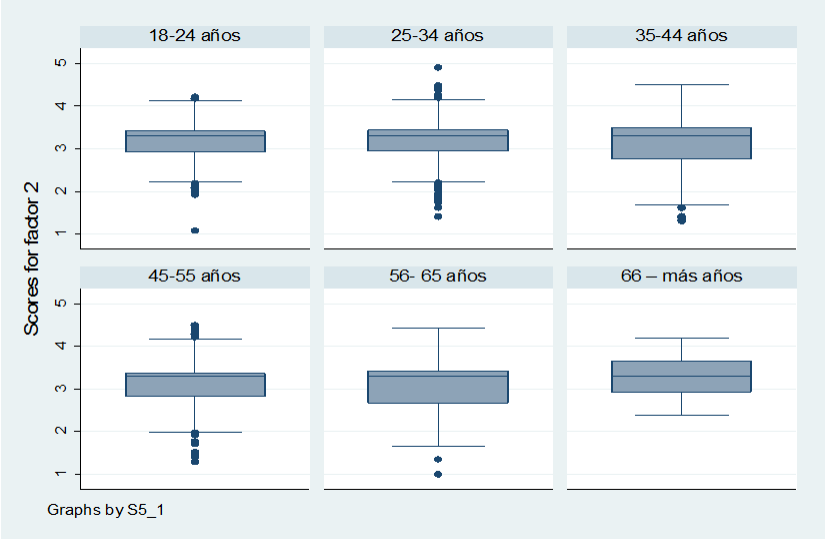

Figure 6b. Box plots of factor 2, by age groups.

Figure 7a. Box plot of factor 1, by income groups.

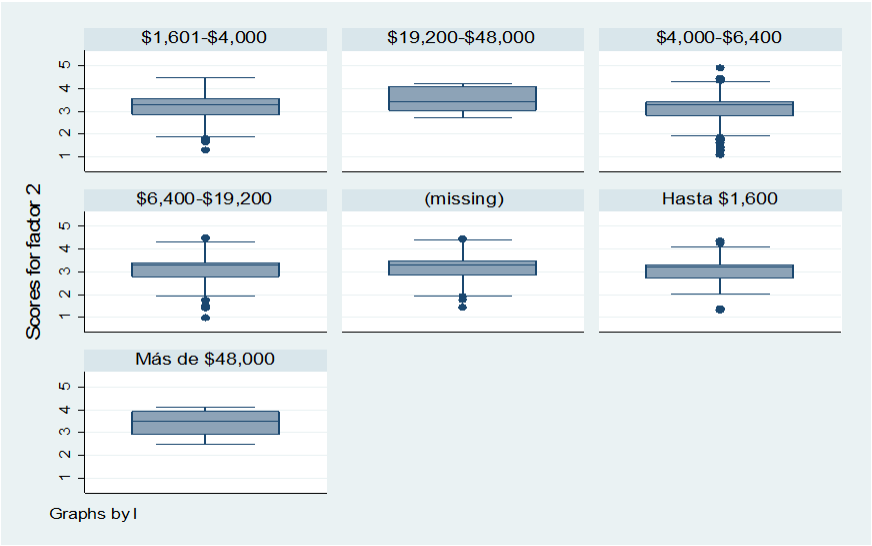

Figure 7b. Box plots of factor 2, by income groups.

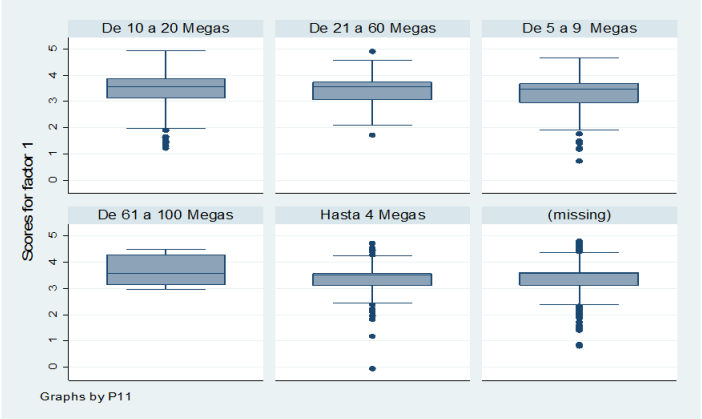

Figure 8a. Box plot of factor 1, by Internet speeds.

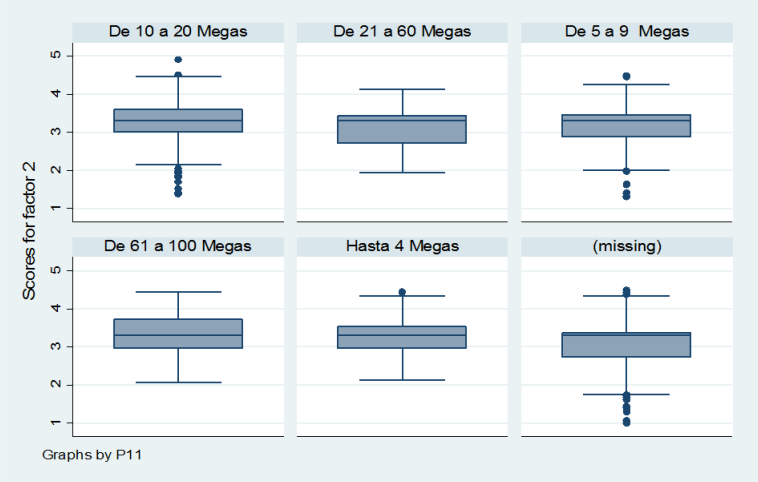

Figure 8b. Box plot of factor 2, by Internet speeds.

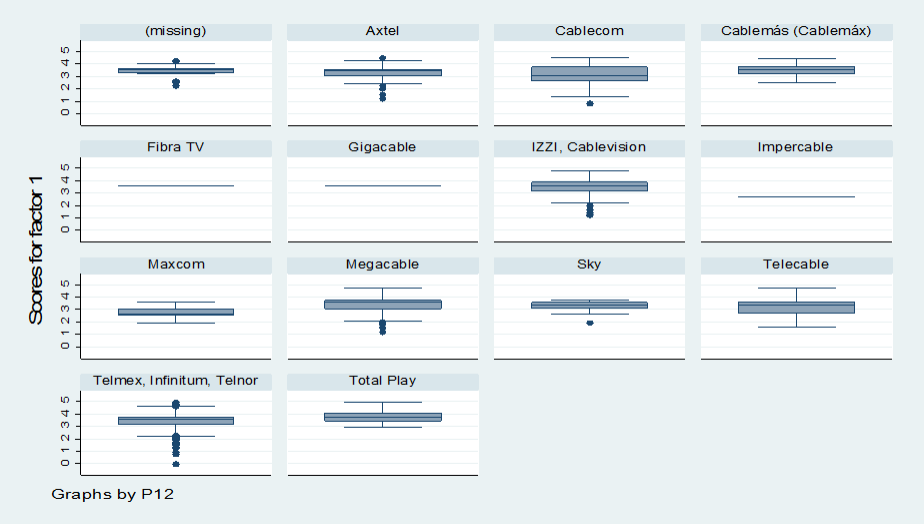

Figure 9a. Box plot of factor 1, by Internet service provider.

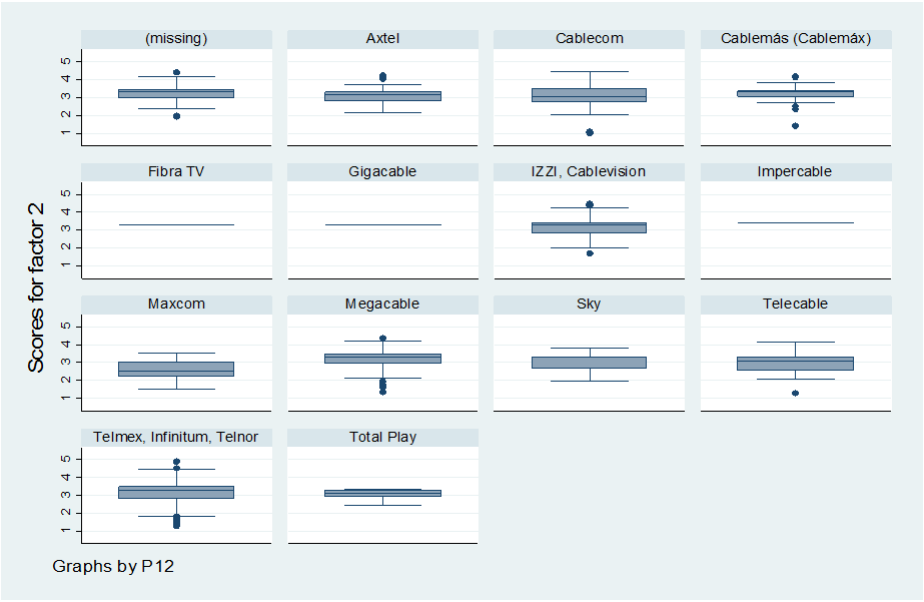

Figure 9b. Box plots of factor 2, by Internet service provider.

Part V

Conclusions

In this article we have applied a methodology of factor analysis in order to analyze a latent variable as satisfaction with the quality of household fixed Internet services offered by the Internet service provider, in Mexico. Factor analysis methodology examines latent variables indirectly by employing variables that are believed to be indicators of customer satisfaction. By using the survey "Encuesta 2016. Usuarios de Servicios de Telecomunicaciones" we identified eight questions that collect survey respondents' opinion regarding their perception of the quality of household fixed internet services offered by their Internet service provider. The election of two factors was confirmed by the application of two criterions: the cumulative variance explained which accounted for the 71.59 percent of total variability and the screeplot criterion. Additionally, the factor analysis was conducted with a *varimax* rotation. The estimation of the factor loadings concluded that variables P33_1, P33_2, P33_3, P33_5 and P33_6 are heavily weighted for factor 1. On the other hand, variables P33_4, P33_7 and P33_8 determine factor 2. Thus, we deduce that factor 1 is related with customer's perception of satisfaction with technical issues of fixed Internet service, while factor 2 is related with the perception of satisfaction with customer service offered by the Internet service provider. The scores of both factors obtained for each observation were graphed using additional demographic variables included in the survey. These demographic variables include gender, region, age groups and income groups. Two other variables that were used are Internet speed and Internet service providers. The results obtained implied the observation of a higher dispersion for factor 2

than for factor 1, in each case. This being so, it is clear that even when customers have contracted a high speed Internet service (e.g. 21 to 60 megabytes or 61 to 100 megabytes) the satisfaction with technical issues presents higher levels than with respect to the satisfaction with the customer service. Thus, we can conclude that customer disatisfaction with customer service must be taken into consideration by operators, especially in order to reduce the cost associated with the churn or customer retention. On the other hand, it is a good opportunity for the Mexican regulator in order to establish quality targets for customer service that must be mandatory for operators. It will allow an improvement in the quality of services offered by the Internet service providers in Mexico.

References

[1] Beattie, Paul; Pinto, Mary; Nelson, Martha and Nelson, Roger (2002). Patient Satisfaction with Outpatient Physical Therapy: Instrument Validation. Physical Therapy. Volume 82, Number 6, pp. 557-565.

[2] Brough, Paula and Frame, Rachael (2004). Predicting Police Job Satisfaction and Turnover Intentions: The Role of Social Support and Police Organisational Variables. New Zealand Journal of Psychology. Volume 33, Issue 1, pp. 8-16.

[3] Cattell, Raymond (1966). The Scree Test for the Number of Factors. Multivariate Behavioral Research. Volume 1, Issue 2, pp. 245-276.

[4] Guilley, W., Uhlig, G. (1993). Factor analysis and ordinal data. Education 114(2), 258–264

[5] Lee, Jonathan; Lee, Janghyuk and Feick, Lawrence (2001). The Impact of Switching Costs on the Customer Satisfaction-Loyalty Link: Mobile Phone Service in France. Journal of Services Marketing. Vol. 15, No. 1, pp. 35-48.

[6] Ley Federal de Telecomunicaciones y Radiodifusión (2014). Diario Oficial de la Federación. México

[7] Likert, Rensis (1932). A Technique for the Measurement of Attitudes. Archives of Psychology. Vol. 140, No. 55.

[8] Manisera, Marica; Van der Kooij, Anita and Dusseldorp, Elise (2010). Identifying the Component Structure of Satisfaction Scales by Nonlinear Principal Components Analysis. Quality Technology and Quantitative Management. Vol. 7, No. 2, pp. 97-115.

[9] Peña, Daniel (2002). Análisis de Datos Multivariantes. McGraw-Hill Interamericana de España S.L.

[10] Saad Andaleeb, Syed and Conway, Carolyn (2006). Customer satisfaction in the restaurant industry: an examination of the transaction-specific model. Journal of Services Marketing Vol. 20, Issue 1, pp. 3–11

[11] Uebersax, John (2015). Introduction to the Tetrachoric and Polychoric Correlation Coefficients. http://john-uebersax.com/stat/tetra.htm

[12] Villafañe, Justo (2003). La Buena Reputación. Claves del Valor Intangible de las Empresas. Ediciones Pirámide, España.

[13] Yu, Yi-Ting and Dean, Alison (2001). The Contribution of Emotional Satisfaction to Consumer Loyalty. International Journal of Service Industry Management. Vol. 12, No. 3, pp. 234-250.