

Can I predict the happiness of My Customer?

Preamble

The case writer explains the reader the importance of using factor analysis in regression analysis. It talks about how to link both the analyses to make appropriate inferences and also build a suitable model. Hence, the situation and the analysis was appropriately designed by him, to facilitate the discussion. Towards the end of the case, list of questions that an instructor can ask, are given. These questions are commonly asked by the students and play an important role in the analysis.

Introduction

Mr. Sam, who is the CEO of SLN ANALYTICS* is very anxious that day. In his capacity as a CEO he had handled many projects. But, the current project is a different one. The client wants them to find out the factors responsible for the customers to choose their products to the competitor and also find whether the customers are happy with them. This is a different challenge to SLN ANALYTICS. The earlier clients have either asked them to do a market survey on customers and give the insights of the same or do a site based survey at their organization and give them a solution. But, the current client wants to find the factors that are responsible for the customer satisfaction and also find whether a customer is happy or not. The customer satisfaction is measured on a 10-point rating. This is a challenge because, the factors, which are latent, have to be identified first from the survey and then the same have to be used to find the satisfaction of the customers. A well-structured questionnaire is used to collect the responses from the customers on several variables and the survey was completed successfully. Now the challenge the team has is, to analyse the data appropriately and give the client the necessary results. Mr Sam, who is heading the team for this project felt that a new approach has to be used to get the necessary results. For this, he had approached a Statistician who is helping them for many years.



Copyright © 2017 Shri Dharmasthala Manjunatheshwara Research Centre for Management Studies (SDMRCMS), SDMIMD, Mysore. This case is published as a part of 'Cases in Management Volume 6 (2017)' with ISBN 978-93-83302-29-1.

The case writer(s) G. Srilakshminarayana, Assistant Professor - Quantitative Methods, SDMIMD, Mysore, may be reached at lakshminarayana@sdmimd.ac.in Author(s) have prepared this case as the basis for class discussion rather than to illustrate either effective or ineffective handling of the situation. This case is fictionalized and any resemblance to actual person or entities is coincidental. This publication may not be digitized, photocopied, or otherwise reproduced, posted, or transmitted, without the permission of SDMRCMS, SDMIMD, Mysore. For Teaching Notes please contact sdmrcms@sdmimd.ac.in.

* SLN Analytics is a fictitious Company created by the author for the discussion

SLN ANALYTICS analytics

Mr. Sam, who has graduated from one of the reputed B-schools has started his career at an analytics firm. In the initial stages, he was trained in the advanced techniques and also received certification from the top universities. His actual abilities have come out when he had handled a critical project for his firm. The CEO of the firm was impressed by his efforts and has made him the head of the analytics team. In his position, Mr Sam had successfully completed several projects. After few years of his experience, he felt that it is time for him to be on his own. Hearing this, the CEO of the firm has expressed that he can partner with him instead of starting a new firm. Mr Sam who felt that he has to go back to his home town to take care of his parents and also to fulfil his dream of providing the employment to his people, informed the CEO the same. After hearing this, the CEO felt happy for him and expressed that still his firm can Collaborate with Mr Sam's new firm in future projects. Mr Sam proceeded to his home town and started his firm in the year 2010 along with his ex-friends at the B-school. In a span of two years, he could complete critical projects and could gain the confidence of the clients.

Organizational culture

Mr Sam has taken all the measures to ensure that his team gets all those training and developmental opportunities to grow. He also encouraged few of the team members who are talented, to start their own firms at their places. He has a different view to his work. He wishes that his firm should be a centre for the future analytics firms and had a vision of seeing his firm as the training point for the future. He is flexible in receiving the suggestions from his team members on the projects and also on expanding the firm. He feels that right people have to be appointed for the right job and never has taken up any project that is beyond their capacities. But, in later years, he felt to take up such projects in collaboration with few consultants. One among such consultants is his friend who is a statistician and also good at training young minds in statistical methods. Whenever SLN ANALYTICS has taken up the projects that needs intense statistical analysis, Mr Sam approaches his friend for help. Also, he started collaborating with other consultants to get solutions on aspects related to marketing, Finance etc. In few years, he could get a good circle of consultants who works for him and also who are willing to collaborate with his firm. Now, he could fulfil his dream of providing employment to the people around him and also expand his services to all other disciplines.

They follow a three step approach: 1. Solve 2. Simplify and 3. Sustain. Using this they could successfully complete several projects and it has made them one of the top analytics firm in India.

The current project

The current project is quite different from the earlier projects. The client wishes to find the factors that are motivating their customers to choose their products and also find whether the customers are happy or not. As mentioned, the satisfaction is measured on a 10-point rating scale. The client wants SLN ANALYTICS to conduct a market survey and then give them the required results. To achieve this, Mr Sam and his team has constructed a detailed questionnaire. In the first stage, they have conducted a pilot study to estimate the sample size required, test the reliability of the questionnaire, and estimate the budget etc. For this, they have taken the help of the statistician and the following is the details of the pilot study.

Pilot Study

Before conducting the actual survey, a pilot study was conducted and the pilot sample is 91. The team has ensured that the pilot study will be done more sincerely than the actual study. The pilot sample respondents were selected based on judgemental sampling and the responses were collected through the questionnaire designed. The results of the pilot study also helped to decide upon the budget required for the actual study, check whether the same questionnaire can be used for the final survey. To check this, Cronbach alpha is computed. If the levels of the Cronbach alpha are more than 0.75, then one can conclude that the questionnaire has the desired levels of reliability. The following table gives the Cronbach alpha value for the current project, based on the pilot results.

Table 1 : Reliability Statistics

Cronbach's Alpha	No of Items
0.913	22

Source : From data analysis.

From the above table, one can note that the questionnaire has the desired level of consistency and hence reliable to capture the factors appropriately. Mr Sam and his team have realized the importance of Cronbach alpha. One can discuss here on the importance of Cronbach alpha and also alternatives for other types of scaling. The same questionnaire can be used to conduct the actual survey. The pilot study is also used to decide upon the sampling design and on the budget necessary to conduct the survey. The team members also have decided upon the time horizon to conduct the survey, number of additional members to be involved in the survey. Mr S has helped them in estimating the sample size needed to conduct the survey.

Estimation of the sample size

In order to estimate the sample size, the results of the pilot study are used. In the first step the key variables of the study are considered and the corresponding estimates are obtained. The estimate needed is the standard deviation and the sample size is estimated at 95% level of confidence, 1.96 as the normal critical value and the level of precision at the level of 0.09. The discussion can be on how to decide upon how to decide upon this value. The following table gives the details of the sample size.

Table 2 : Sample size determination

Variable	Std. Deviation	Sample size
V11	.7336	255.261201
V12	.8134	313.748908
V13	1.0788	551.984974
V14	.958	435.133836
V15	.8561	347.567542
V16	.8376	332.742936
V17	.8309	327.415342
V18	.9553	432.809032
V19	.9912	465.932764
V20	1.0260	499.288129
V21	.9234	404.433808
V22	.7971	301.356463
V23	1.0259	499.172312
V24	.9350	414.625725
V25	.9065	389.725018
V26	1.0977	571.442271
V27	1.1677	646.723478
V28	.9147	396.78987
V29	.8804	367.603925
V30	.7463	264.179128
V31	1.001	474.922914
V32	1.0850	558.354922
	Max=	646.723478

Source: From Pilot Study Results

The final sample size is the maximum of the individual sample sizes. From the above table, one can note that the maximum sample size is 647. Mr Sam and his team had decided to target minimum 647 customers, to conclude the survey with 95% confidence. **One can discuss on the logic behind this.**

Once the sample size is estimated, the team has gone to the market to collect the data from the customers. The population for the survey is the segment of customers that have already purchased the products of the clients at least once. The population is decided based on the pilot study conducted and the final survey was successfully conducted. In order to get the suggestions from the statistician, Mr Sam and his team has met the statistician (Mr S).

Sampling design

Mr Sam, based on his past experience, had decided to adopt a non-probability sampling design. This is because of non-availability of a sample frame and also he felt that using this, they can reach the maximum. **Is it necessary to have the sampling frame always? What are the advantages of a non-probability sampling design?**

Solution to the problem

Mr Sam: Sir, good morning.

Mr S: Good morning team. Hope you are doing well in your projects. I learnt about the current project from Mr Sam. I have understood the problem clearly and I had planned the methods to analyse the data. Let me explain the methods I have planned to give you a solution.

Team: We are happy to meet you to learn the methods. The survey questionnaire and the data was mailed to you. Hope you have received the same.

Mr S: Yes. I have received the same and I will show you the analysis of the same today.

I have divided the requirements of your client into two parts / objectives. The first one deals with finding the factors responsible for the customers to choose the client's products and the second one deals with finding whether the customers are happy or not. To achieve the first objective of the study, one has to use the "Factor analysis". Factor analysis extracts the latent factors that are responsible for the customers to choose the products. Once the factors are identified, the corresponding factor scores can be generated for each of the customers and

the same can be used to find whether they are happy or not. For this, one has to use “**Factor analysis Regression (FAR)**”. This is one type of regression that is used when there exists the problem of multi-collinearity. Let me first explain the factor analysis in detail and then FAR. Before I explain the FAR, I will also explain the ordinary regression analysis and the reasons that lead to FAR. This will give you a clear understanding on the difference between both and also why FAR is relevant for the current project.

Factor analysis

This is one technique that is used to extract the latent factors or constructs, from the observed variables. The principle on which the method is constructed is, significant correlation between the variables. Mr S had used this to find the factors and the following are the results.

Table 3 : KMO and Bartlett’s Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.859
Bartlett's Test of Sphericity	Approx. Chi-Square	9185.545
	Df	231
	Sig.	.0001

Source : From data analysis.

Mr Sam : What is the KMO test and why this is needed? How should we use this in the current project? (To be discussed in class)

Table 4 : Communalities

	Initial	Extraction		Initial	Extraction
V11	1.000	.750	V22	1.000	.605
V12	1.000	.591	V23	1.000	.680
V13	1.000	.615	V24	1.000	.649
V14	1.000	.727	V25	1.000	.763
V15	1.000	.517	V26	1.000	.748
V16	1.000	.780	V27	1.000	.781
V17	1.000	.657	V28	1.000	.798
V18	1.000	.825	V29	1.000	.631
V19	1.000	.876	V30	1.000	.676
V20	1.000	.828	V32	1.000	.761
V21	1.000	.715	V31	1.000	.726

Source : From data analysis.

Mr S: The above table gives the communalities. They are like R-square in the regression analysis. The values given in the last column gives the percentage explained in each variable by the factor analysis (To be discussed in class)

Table 5 : Total variance explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	9.217	41.896	41.896	9.217	41.896	41.896	4.628	21.035	21.035
2	1.646	7.482	49.378	1.646	7.482	49.378	3.146	14.298	35.333
3	1.436	6.527	55.904	1.436	6.527	55.904	2.487	11.304	46.637
4	1.200	5.455	61.359	1.200	5.455	61.359	2.449	11.130	57.768
5	1.104	5.018	66.377	1.104	5.018	66.377	1.673	7.606	65.374
6	1.097	4.986	71.363	1.097	4.986	71.363	1.318	5.989	71.363

Source:From data analysis.

The table above gives the total variance explained by the six factors out together.

Table 6 : Rotated Component Matrix

	Component					
	1	2	3	4	5	6
V11				.769		
V12			.579			
V13					.698	
V14					.663	
V15	0.436					
V16				.764		
V17			.610			
V18		.774				
V19		.807				
V20		.836				
V21	.610					

Contd...

V22	.609					
V23	.625					
V24	.625					
V25	.836					
V26	.662					
V27			.824			
V28				.585		
V29	.557					
V30				0.493		
V32						-.852
V31	.716					

Source:From data analysis.

The values in the above table gives the factor loadings and they act as the correlations between the variables and the factors.

Based on the analysis, factors scores are computed and the following table gives the same.

Table 7 : Factor scores

Factor-1	Factor-2	Factor-3	Factor-4	Factor-5	Factor-6
-0.16874	-0.15706	-0.65545	-0.93309	-1.77756	1.03512
0.852	0.22866	1.4313	-0.15637	-1.84831	0.85027
0.78558	0.2054	-1.55287	-0.32056	-2.14123	0.14282
-0.13661	-1.39692	-0.73526	-0.52625	-1.74126	0.27338
3.16799	0.5214	-0.54843	-0.88466	-0.41497	-1.48069
1.23133	0.02629	1.27518	-1.94198	0.19818	-0.80538
2.67037	0.56028	-1.76032	0.61953	2.43395	0.75815
-0.18292	-1.42206	-0.38408	-0.14729	0.88168	-0.39394
-0.49276	-0.06156	0.39837	-1.08052	0.27841	0.1672
-0.56889	-0.20624	0.0218	-0.98776	-0.18326	-0.16114
1.20321	1.03548	0.53908	0.15916	-1.54465	1.08122
0.39008	-1.02611	0.25733	0.7027	1.40312	0.20464
-1.11945	-0.53984	-0.73491	0.97084	2.21907	0.09052
-0.72081	-0.30655	-0.67177	2.2569	-0.7437	-0.36557
-0.56323	1.81466	-0.15016	0.33749	0.53775	0.24596

Source:From data analysis.

Ordinary regression

Regression analysis is used to build a model between the response (dependent variable) variable and the regressor variables (independent variables), which can be used to understand the behaviour of the response with the corresponding changes in the regressor variables. The impact of each variable can be studied separately and the set of variables that are significantly impacting the response variable can be identified. It is an important method that gives the analyst an opportunity to understand the behaviour of response variable against each of the regressor variable and also to generalize the model built to understand the reality. The process starts with building a model, testing for the significance of the model, testing the model assumptions, building the interval bands to understand the sensitivity of the model etc. One key assumption is that the regressor variables are independent and if this assumption is violated, then the model fit cannot give reliable results. This is called as problem of multi-collinearity. One can discuss more on this issue while discussing the case.

There exists a solution to this. That is, grouping the variables that are dependent as components and using these components in regression analysis. Note that, these components are orthogonal to each other and this makes the components independent. Also note that, the original variables are replaced by these components and hence the variance explained by the original variables has to be preserved to an extent in the components identified. In order to ensure that the variance explained in these components is maximum, the components have to be extracted such that, each component explains the maximum variance of the situation. To achieve this, one can use Principal Component Analysis (PCA) and this method gives the components in such-a-way that, the variance explained by the components is maximum. As per the method, the first component has a linear combination of variables such that, it explains the maximum variance and the other components will have the subsequent variances. The most important aspect is the total variance explained. The current project demands identifying the factors and then use the same in studying the customer satisfaction levels. The factors are latent and has to be identified using the original variables. These factors have to be extracted and the original variables can be replaced by these factors. These factors are expected to contain the maximum variance of the given situation. For this, PCA is used to extract the factors that contain the maximum variance. For the current project, factor analysis is applied and the above section gives the results.

Mr Sam: Excellent way of identifying the factors and using them in regression analysis is a new aspect.

Factor analysis regression

Mr S: Now, let us look at the FAR. The following tables give the results of the same. The basic difference between the ordinary regression and the FAR is explained in the previous discussion.

From the above factor analysis, one can note that six factors have been extracted by the analysis and all the six put together explains 71% of the total variance. Here, one can discuss on the level of variance that the factors have to explain for the analysis to be a successful one.

For the current project, the model fit is assessed using adjusted R-square and it happens to be 0.82. This indicates that the linear model used to link the customer satisfaction and the factors is a good fit. Note that, all other assumptions of linear regression analysis are satisfied by the data. The model significance is tested and the p-value happens to be 0.0001. This indicates that the model is significant and can be used to estimate/predict the satisfaction of the customers. The following is the model

$$\text{Customer Satisfaction} = 3.21 + 0.185 F_1 + 0.245 F_2 + 0.345 F_3 + 0.175 F_4 + 0.125 F_5 + 0.185 F_6$$

Using the above model, one can get the customer satisfaction.

Mr Sam: This is an interesting analysis and we have learnt many aspects related to the statistical methods. But, we have the following questions.

Questions for discussion

1. While estimating the sample size, how can one decide upon the level of precision?
2. What is Cronbach alpha and how it is used in testing the questionnaire?
3. Can't we use always a non-probability sampling design to collect the data?
4. What is multi-collinearity and why it has to be given importance?
5. Are there alternative methods to extract the factors, other than PCA?
6. Why the factor rotation is important in the factor analysis?
7. Is there any threshold on the total variance explained in factor analysis?
8. Is the sample size sufficient to run the factor analysis?
9. How can one use the model to estimate/predict the customer satisfaction?
10. What is the managerial implication of the same and is this process useful to the client of SLN ANALYTICS to understand the customer satisfaction?