Marmara Üniversitesi
İ.İ.B.F. Dergisi
YIL 2005, CİLT XX, SAYI 1

APPLICATION OF HOMOGENITY ANALYSIS FOR QUESTIONNAIRE RELIABILITY

Dr. Dicle Taspinar*

Abstract

Using questionnaire is the most commonly used technique for data collecting in marketing and social researches. The most important points in collecting data by using quesionary method is that the attributes must be reliable and the people must give importance while filling it and be painstaking. A deterministic criteria in questionnaire reliability is Cronbach's alpha coefficient. As the value of Cronbach's alpha increases, questionnaire reliability gets to be more reliable.

In this paper, we will prepare a simulation questionnaire in order to dewell upon determining the choices of the responder which would affect the consistency and therefore reliability of the questionnaire.

Giovanni, Bernardi [2] and Barnette [3] also studied that Cronbach's alfa coefficent can be affected by the observations in previous works. Giovanni [4] studied observations and the correlations among variables together with the variance of these variables using a Fortran computer program. He then have built a program which can calculate the point where Cronbach's alfa would be maximum. Bernardi [2] tried to show that some relations can be found between the variables and the observations even if Cronbach's alpha is small. Barnette [3] worked out a simulation and found a way to increase and decrease Cronbach's alfa by looking at the answers of the replier responder.

We will dwell upon the reliability of the questionnaire and the Cronbach's alfa which is the unit of measurement of this. We will show that it is possible to find out the observations which causes decrease in consistency by using homogeneity analysis which is an optimum scaling method; and it is then also possible to increase the questionnaire reliability by just taking out those observation(s).

* İstanbul Ticaret Üniversitesi, Öğretim Üyesi

Keywords : Cronbach's alpha, consistency, reliability, homogenity analysis

1. CRONBACH'S ALFA COEFFICIENT

Cronbach's alpha is developed to measure the reliability of the tests (like psychological tests) which are created especially to meausure specific characteristics of people. There are two important requirements for a test to be successfull: It must give the same result when it is applied on the same person at different times, as well as when it is applied on the same person by different interviewers. There are some coefficients which designates how reliable a test is. Cronbach's alpha can be used to measure the reliability of a test with ordinal scaled data.

Cronbach's alpha is developed by Cronbach [13] in 1951 and is calculated as follows:

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum s_i^2}{s_t^2}\right)$$

Number of treatments is expressed by k; variance of the i'th treatment is expressed by s_i^2, variance of the whole experiment is expressed by s_t^2. It is clear that α increases as s_t^2 increases while Σs_i^2 decreases. Normally, value of α is always between [0, 1]. As α gets closer to 1, the reliability of the test increases; and as α gets closer to 0, the reliability of the test decreases. If there are negative correlations between treatments, then value of α will be negative. The recommended solution to increase the reliability is either to increase n, or to cancel the threatment with the highest s_i^2.

The α coefficent can also be used to increase the reliability while preparing the attributes and using them in a plot study. The important point in here is that, the attributes must be prepared with consistency which helps to observe the subject interested. As an example; let's assume that we are preparing a questionarrie in order to measure the responder behaviour. If we want to see the influence of a promotion, the attribute group prepared to measure this influence must also designed only to measure this point. Also the correlation between the attributes must be positive. Therefore if a responder has selected "I agree" in one attribute, it is ideal that s(he) also responded "I agree" in all attributes. The responder who responded this way is assumed to be consistent, and would be easily categorised. The result would be that the requested benefit would be fully gained this way. But if the attributes consist one which is different than the main theme and therefore possibly wouldn't measure the requested value; when the responder comes to that attribute, (s)he won't be

consistent with the previous answers. Therefore, the responses might be different. Cronbach's alpha would definitely decrease in this case. In other words, the consistencies between the answers would be lowered. As the reliability decreases, the attributes won't reflect the demanded characteristics of the responder. In this case, this attribute should simply be discarded. If we discard the attribute which has the highest variance, we will observe that the reliability of the questionarrie would increase. After we make a plot study which results in elimintaing the inconsistent attributes and the answers from the questionnarrie, the remaining attributes and answers would highly satisfy the goal.

2.ANOTHER FACTOR AFFECTING CRONBACH'S ALPHA

It is argued in the previous section that as the Cronbach's alfa coefficient increases in a questionarrie, the consistency and therefore the reliability of the test increases. Also the test is assumed to measure the target characteristics in the marketing research questionnarie. It is also shown that as the variance of a attribute with the biggest variance increases, Cronbach's alfha decreases resulting in decreasing the reliability.

Cronbach's alpha coefficient will be studied from the observations view also. When working on the questionarries, it is clearly known that there are certain problems such as the responders might not understand the attributes, they may not be so enthusiastic to answer them, causing them not to answer the proper one which best represents their thoughts. This makes it impossible to take out the correct results from the faulty choices they have made. Expressing the situation in another way might be that the results might not be reliable at all. Similar case may be valid in marketing researchs where Cronbach's alpha coefficient can be used. When we go one step further, if a responder answers an attribute randomly which does not represent the actual thought, the consistency would be effected. After concluding that the reliability of the test is decreased by this careless responder, removing the answers of this responder would be a fair operation. This would save us to cancelling out a proper attribute where the rest of the responder have given proper answers. Even if the attribute would be judged to be not proper, taking out this responder would increase the reliability of the whole test.

3. A HYPOTHETICAL PRACTICE

We have built a simulated test which is applied on a group of 35 people. The test consists of 4 attributes which have 5 optional answers, i.e: ["I totally agree", "I aggree", "Undecided", "I disagree", "I totally disagree"]. We assume that the responders are all consistent with their individual answers. This concludes that if a

responder has choosen "I agree" in an attribute, s(he) must also choose the same options from the other attributes in order to be the consistent as wel; and this is assumed the case. When we applied reliability analysis to these data by using SPSS program, we obtained 1.00 as Cronbach's alpha value as expected. Therefore the consistency between the answers is high as well. When we change one answer of one responder, we immediately observe the decrease in Cronbach's alpha. If we change other responders answers, we observe lower Cronbach's alpha values. In this case, even if most of the responders have understood one attribute througly, and only a couple of them didn't understand so, or they just answered the attribute unconcerned , then the variance of that attribute might be high. In order to increase the Cronbach's alpha, regular action would be cancelling the attribute resulting in increasing the reliability of the whole test. We propose removing answers of those responders who didn't payed enough attention to the attributes to increase the reliability of the whole test. This would save the whole attribute and the consistent answers to it by sacrifisinga couple of responders answers.

We need to find out the responders who handled the test casually and therefore behaved differently from the majority in this phase this way. The second movement would be excluding them from the test. Several methods can be used to point those people. Using optimal scores which can be derived by optimal scaling will be studied in this paper.

Optimal scaling methods are being used especially for quantification to qualitative data as being analysis techniques used in qualitative data. The phases of the operations using qualitative data are: 1: scaling the categories of the ordinal (or nominal) variables, 2: selecting the most appropriate scale among these, and 3: deciding which quantifications the observations will consist instead of the qualifications.

The multivariate techniques based on this approach are called non-linear multivariate analysis techniques. If the technique uses qualitative data, the technique is then called linear multivariate analysis techniques. When qualitative data are being used, quantification the qualitative observation data would give the same result, resulting an unnecessary process. But if me make a close observation, we will see that this process requires a linear transformation. Therefore the analysis techniques used with qualitative data are called linear multivariate analysis techniques. When we think about the same case for the quantitative data, non-linear transformations are being used in order to quantify the observation data. The methods which uses quantifying methods are called as non-linear multivariate analysis techniques. The samples of this technique are Correspondence Analysis,

Homogeneity Analysis, Non-Linear Principal Component Analysis and Non-Linear Canonical Correlation Analysis.

When examining Cronbach's alpha coefficient which is the unit of measurement of the reliability of the tests applied in marketing researches; more than two categorical variable are assumed to be directed the same subject, therefore examing the observations with the homogeneity analysis would be appropriate. Homogeneity analysis reveals the fit between the variables and the observations by using Alternating Least Square (ALS). This causes to homogeneity analysis to be called by HOMALS (Homojenity Analysis and Alternating Least Square). This technique is developed by Benzecri [1973] and is also called as Multiple Correspondence Analysis. When the number of the parameters are more then two, homogeneity analysis would be most appropriate to use.

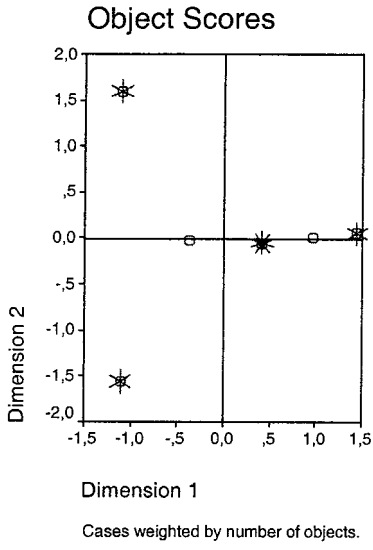
3.1. Examining the observations by homogeneity analysis

When the consistency is highest for each responder in the group with 35 people, itt has been mentioned that the value of Cronbach's alpha would be 1.00 which is the highest possible. We have made some changes on the answers of responders #21 and #22 which are randomly chosen. The change decreased these consistency of the answers, and decrease Cronbach's alpha down to 0.97.

It can easily be seen that the changes in the observations #21 and #22 effect Cronbach's alpha. A decrease in alpha is also observed when number of unconsistent observations are increased with similar behaviour. This concludes that when number of unconsistent observations are small with respect to total number of observations, these observations can just be ignored. In order to do this, we have used homogeneity analysis.

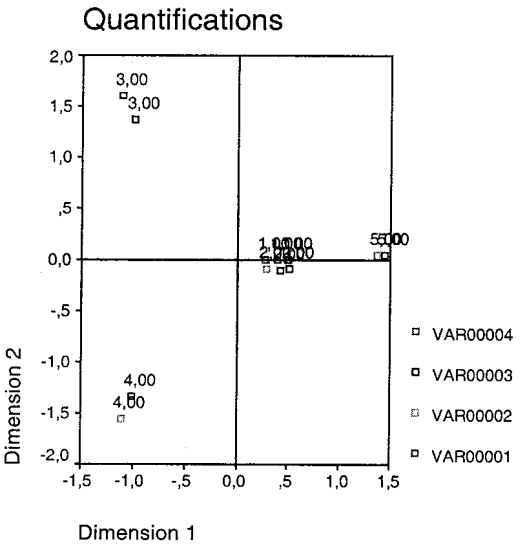
When homogeneity analysis is applied to the data, two-dimentional fit is found to be 1.94. This value is judged to be harmonious since it is very close to 2.00 which is the maximum value in two-dimentional case. Graphics 1 shows the object scores in two-dimention.

Graphics 1 : Object Scores



When we look at graphics 1, we see that 33 out of 35 observations are grouped together at 4 points. It is required that these people are the ones who replied the attributes with one of the following 5 choices: ["I totally agree", "I aggree", "Undecided", "I disagree", "I totally disagree"]. This concludes that 5 group was expected to be formed instead of 4. Graphics 2 shows the category qualification. When we observe graphics 2 closely, it can be seen that the reponders who has choosen 1 and 2 are marked very closely. Observing graphics 1 also shows that the group which is most close to the center has the highest population. The most important observation we have is that actually there are responders from two different groups who gathered together on the graph. When we inspect the coordinates of the two, we see that these are 21st and 22nd observations.

Graphics 2: Category Quantifications



As we observe this example, when the fit is high in a homogeneity analysis; if most of the observations gather together while a couple of them stays alone we can justify that these ones behave differently from the others and belong to a minority. This means that it is possible to get better results just by ignoring the minor observations and re-calculating Cronbach's alpha.

4. AN ACTUAL APPLICATION

Until now, we have built a hypothetical questionarrie and conclude changes in Cronbach's alpha with assumptional changes in responses of a couple of responders. We will now try to see the effects in real observation which has higher population. We have chosen to use a questionnarie which consists of 38 attributes about consumption of fruit juice which is applied to 500 responders . The 38 attributes are designed to get information about 4 different subject. The grouping of attributes into 4 subjects are shown in Table 1.

Table 1 : Grouping the attributes of fruit juice questionnarie.

	Taste Related	Quality Related	Price Related	Health Related
	attribute1 attribute 3 attribute 5 attribute 7 attribute 9 attribute 11 attribute 12 attribute 14 attribute 16 attribute 22 attribute 23 attribute 24 attribute 28 attribute 33	attribute 18 attribute 24 attribute 27 attribute 29 attribute 32 attribute 34 attribute 36 attribute 37 attribute 38	attribute 6 attribute 15 attribute 21 attribute 32 attribute 35	attribute 2 attribute 4 attribute 8 attribute 10 attribute 13 attribute 15 attribute 17 attribute 19 attribute 20 attribute 26 attribute 31
Cronbach's alpha	0,7566	0,6798	0,4993	0,7631

Cronbach's alpha is calculated and given at the bottom of the table which shows the reliabilty of each group. It can be easily seen that all of the groups have higher consistency except Price-Related one. This also means that this group has a consistency problem. Computing the variance of the 5 attributes [5, 15, 21, 32, 35] which belong to Price-Related group shows that the values are very close to eachother. Attributes category quantifications by using HOMALS is depicted in Figure 3

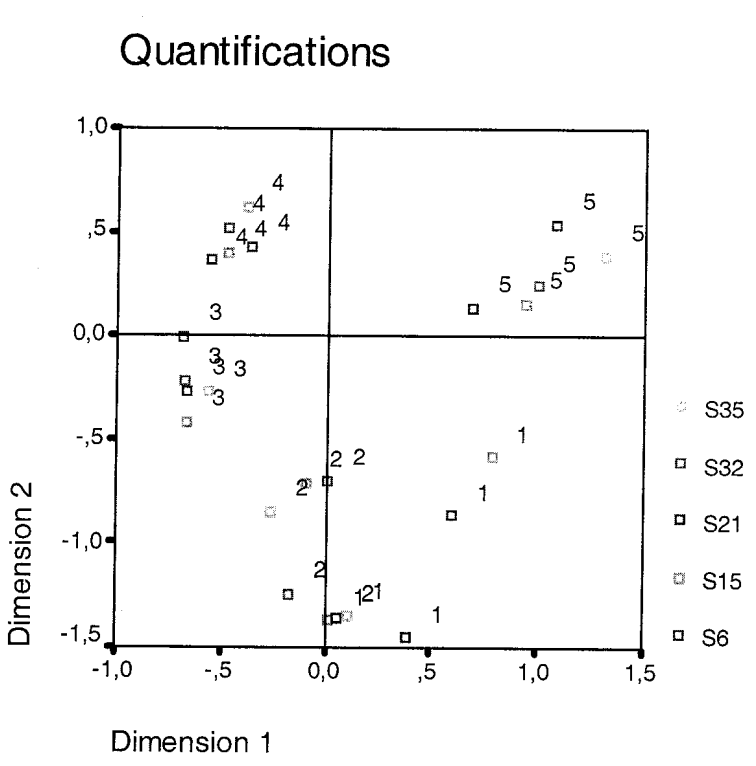


Figure 3: Category quantification related with Price Related Attributes

The locations of each category can be seen in Figure 3. We justify that the observations which are located between category 1 and category 5 groups are not consistent. In other words, the responders who are in this region responded "I totally aggree" to some attributes, while responding "I totally disagree" to others. Examining the observation scores, the identities of the observations in this region is required to be determined.

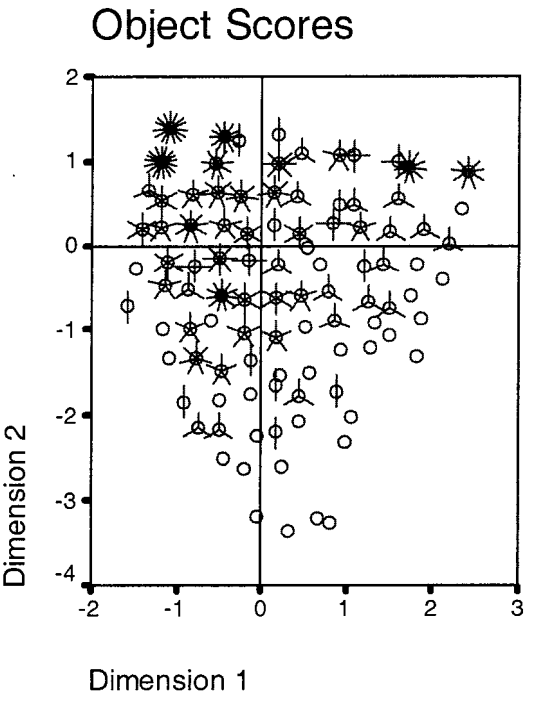


Figure 4 : Observation scores related to Price Attributes.

Examining the coordinates gives us 15 observations in this region. At the beginning (step 0), cronbach's alpha equals 0.4993. At the 1st step, we excluded observation #474 and calculated Cronbach's alpha as 0.5020. At the 2nd step, we have excluded observation #457 as well as #474 and calculated Cronbach's alpha as 0.5038. We have repeated this process 15 times so that in each step, we have excluded all the observation in previous steps together with a new observation. The resultant Cronbach's alpha values are shown in Table 2.

Table 2 : Cronbach's alpha values calculated by ignoring the observation cumulatively.

Step #	Excluded Observation #	Cronbach's alfa
0		0.4993
1	474	0,5020
2	457	0,5038
3	346	0,5071
4	260	0,5086
5	239	0,5104
6	228	0,5123
7	202	0,5150
8	201	0,5161
9	119	0,5182
10	118	0,5193
11	94	0,5215
12	76	0,5241
13	73	0,5266
14	46	0,5282
15	22	0,5295

We can see that ignoring one observation, namely #474 gives us Cronbach's alpha value of 0.5020 while ignoring all of the 15 inconsistent observations gives us 0.5295. This shows that ignoring the 15 inconsistent observations increase the reliability from 0.4993 to 0.5295.

5. CONCLUSION

Measuring the consistency of the attributes is an important topic then ensuring reliability to the decision-makers. Increasing the reliability of the whole questionarrie by ignoring the attributes with high variance is the regular solution. We have shown that instead of ignoring the whole attribute and losing a valuable data, analyzing and checking the consistency of the observations and ignoring the ones with inconsistent observations may be an important method and can be important alternative. Moving ahead from this point, homogeneity analysis is thought to be an important method and can be used to pick the observations which are carelessly answered. The study shows that homogeneity analysis points out the observations in the minority group who responded carelessly. An important thing in

here would be that increase in the number of the observations makes it harder to justify the graphics with human eyes. A solution to this problem might be that coordinates of observation scores might be still hard to analyze the coordinates of the observations when the amount is high, where a computer program may be a solution to this problem.

References :

[1] BECKER, Gilbert, *Creating Comparability Among Reliability Coefficients: The Case Of Cronbach's alfa And Cohen Kappa,* Psychological Reports, Dec2000 Part 2, Vol. 87 Issue 3, P1171, 12p

[2]BARNETTE J (1999). *Nonattending Respondent Effects On The İnternal Consistency Of Self-Administered Surveys: A Monte Carlo Simulation Study.* Educational and Psychological Measurement 59(1),38-46

[3] BERNARDİ, Richard, *A Validating research results when Cronbach's alfa is below .70: A methodological procedure,* Educational & Psychological Measurement, Autumn94, Vol. 54 Issue 3, p766, 10p

[4] FLEBUS, Giovanni Battis, *A program to select the best items that maximize Cronbach's alfa*; Educational & Psychological Measurement, Winter90, Vol. 50 Issue 4, p831

[5] GIFI, Albert, *Nonlinear Multivariate Analysis*, Wiley&Sons, Netherland, 1990.

[6] Giovanni Battis Flebus, *A Program to Select the best Items That Maximize Cronbach's Alfa*, Educational&Psychological Measurement, Winter90, Vol. 50, Issue4, p.831,3p.

[7] GREENACRE, J. Michael, *Correspondence Analysis In Practice*, Academic Press, Harcourt Brace & Company, Publishers, London, 1993

[8] KOPALLE, Praveen K.; LEHMANN, Donald R, *Alfa inflation? The impact of eliminating scale items on Cronbach's alfa,*. Organizational Behavior & Human Decision Processes, Jun97, Vol. 70 Issue 3, p189, 9p

[9] Van De GEER, J., *Multivariate Analysis of Categorical Data: Theory*, Sage Publications, USA, 1993.

[10] Van De GEER, J., *Multivariate Analysis of Categorical Data: Application*, Sage Publications, USA, 1983.

[11] TENENHAUS, M. And YOUNG, W:F., *An Analysis and Shenthesis of Multiple Correspondence Analysis, Optimal Scaling, Dual Scaling, Homojeneity Analysis and Other Methods for Quantifying Categorical Multivariate Data*, Psychometrica, Vol. 50, no.1, 91-119, March 1985.

[12] BAGOZZI, R.P., *Principles of Marketing Research*, Edited by Bagozzi, Blackwell Business, USA, 1994, s18

[13] CRONBACH, L.J. (1951)., *Coefficient alpha and the internal structure of tests*, Pyschometrika, 16(3), 297-334.