maserati

[Draw your reader in with an engaging abstract. It is typically a short summary of the document. When you’re ready to add your content, just click here and start typing.]

[Document title]

[Document subtitle]

**APPLYING THE SHAPLEY VALUE REGRESSION METHOD TO MARKETING RESEARCH**

Contents

[I. Abstract 2](#_Toc8302944)

[II. Introduction 3](#_Toc8302945)

[Advantage of the CSAT model 3](#_Toc8302946)

[Limitations of the CSAT model 3](#_Toc8302947)

[The variables used in our study 3](#_Toc8302948)

[III. Multicollinearity 5](#_Toc8302949)

[Consequences 5](#_Toc8302950)

[How to detect Multicollinearity? 5](#_Toc8302951)

[Examination of Correlation Matrix: 5](#_Toc8302952)

[Variance Inflation Factor: 5](#_Toc8302953)

[Eigensystem Analysis of Correlation Matrix: 5](#_Toc8302954)

[Solutions : Shapley value regression : 6](#_Toc8302955)

[IV. Linear regression and multicollinearity 7](#_Toc8302956)

[Overview the data 7](#_Toc8302957)

[OLS regression’s results 8](#_Toc8302958)

[Multicollinearity 9](#_Toc8302959)

[V. Application of SVR 10](#_Toc8302960)

[SV and reg interperetation 10](#_Toc8302961)

[Computing SV 10](#_Toc8302962)

[VI. Methodology and results 11](#_Toc8302963)

[VII. Code SAS explanation 14](#_Toc8302964)

[VIII. Conclusions 16](#_Toc8302965)

[IX. BIBLIOGRAPHY 17](#_Toc8302966)

[ANNEXES 18](#_Toc8302967)

# Abstract

As soon as you work with customer’s perception, variables are often correlated. The clients gave a satisfaction rating on a scale from 1 to 10.

So what is the issue with multicollinearity? It is not a bias issue, but a precision issue.

**The solution to solve this issue:**

The *first* solution is obtain more data. It is the best answer to the issue. Why?:

Modelling on perceptual data often happen in a barometric context. Brand preference drivers, satisfaction drivers,.. don’t vary significantly across a few months or even across years in some sectors. Cumulate waves of a barometer to get more data, and thus more precision in the modelling, is an excellent answer to the challenges raised by multicollinearity

The *second* solution is reduce the number of estimated in the model by testing coefficients equality. By imposing constrainst on the coefficients, the dimensionality of the model is reduced end experience shows that confidence intervals around the coefficients shrink.

But in this paper, we will talk about the *third solution* which based on a technique used in Game Theory – Shapley value. So, Shapley Value is also a solution to explain how strong contribute the variable. The Shapley value, from economics, provides an alternative way to define variable importance. As we describe below, Shapley value provides a way to attribute the value created by a team to its individual members. In our context the members are individual input variables. The Shapley Value creates a score for each factor in a game which represents the worth of that participant over all possible combination of their coalitions contribution to the total value of the game. This applies to cooperative games.

# Introduction

* **SNCF Customer Satisfaction Study (CSAT).**

As part of our study on SNCF customer satisfaction, it seemed logical to choose one of the most widely used models in market research, the CSAT or Customer Satisfaction Score model.

CSAT measures a customer's satisfaction following a specific interaction with the company.

The database of our study represents a sample of SNCF customers, who answered CSAT's questions on whether the customer is satisfied with the company's contact services, etc. Leaving scores on a scale of 10 for each question asked. The higher the number, the higher the customer satisfaction. We note:

- 1 to 6: detractor note

- 6 to 7.5 : neutral note

- 7.5 to 10: promoter note

## Advantage of the CSAT model

The advantage of this model is that it produces a rate that is simple to visualize and interesting to use for real-time measurements.

In addition, CSAT is versatile because it allows you to ask a variety of questions to clients.

## Limitations of the CSAT model

Despite the ease of implementation of this model, the CSAT has limitations such as:

- Focuses on a specific interaction and not on a broader relationship with the company.

- Users give an overall rating, it does not detail their experience so it only focuses on the main point of satisfaction or dissatisfaction.

- Users only judge the present time: the customer does not project himself into the future. However, even a positive rating will not allow us to make a definitive conclusion on customer satisfaction.

## The variables used in our study

The questions asked in the CSAT, representing the variables of our study, are as follows, how satisfied are you with: (all on a scale of 10)

Q1a - The orientation in the station?

Q1b – The ease of finding information about your train (time, track...)?

Q1c - The ease of locating the services available at your station?

Q2 - The transfer to and from the platform?

Q3a - The cleanliness of the station?

Q3b – The security of the station?

Q3c - The cleanliness of the toilets in the station?

Q4a - The comfort of waiting in the station?

Q4b - The time spent in the station?

This is a measure of client’s satisfaction and it’s a continuous variable. Here, we have 4 groups of questions in total:

1. Information (q1a, q1b, q1c)

2. Transfer (q2)

3. Cleanliness and security (q3a, q3b, q3c)

4. Comfort (q4a, q4b)

Our expectation is all those variables have a positive impact on the satisfaction variable.

# Multicollinearity

Multicollinearity is a statistical phenomenon which defined as the incidence of a high degree of correlation between some or all of the predictor variables:

The multicollinearity of empirical data violates the assumption of independence between the explanatory variables of a linear regression model and also with variable dependent.

## Consequences

* Difficult to come up with reliable estimates of their individual coefficients.
* Incorrect conclusions about the relationship between outcome variable and predictor variables.
* Lack of statistical significance of individual predictor variables even though the overall model may be significant caused by by inflating the standard error of the estimates of the estimated regression coefficients.
* Serious problems with the estimation of β and the interpretation. ( multicollinearity affects the sign of regression coefficients)

## How to detect Multicollinearity?

### Examination of Correlation Matrix:

* + Large correlation coefficients in the correlation matrix of predictor variables indicate multicollinearity.
  + If there is a multicollinearity between any two predictor variables, then the correlation coefficient between these two variables will be near to unity.

### Variance Inflation Factor:

* + The Variance Inflation Factor (VIF) quantifies the severity of multicollinearity in an ordinary least- squares regression analysis.
  + VIF is an index which measures how much variance of an estimated regression coefficient is increased because of multicollinearity.
  + ***Rule of Thumb***: If any of the VIF values exceeds 5 or 10, it implies that the associated regression coefficients are poorly estimated because of multicollinearity (Montgomery, 2001).

### Eigensystem Analysis of Correlation Matrix:

* + The eigenvalues can also be used to measure the presence of multicollinearity.
  + If multicollinearity is present in the predictor variables, one or more of the eigenvalues will be small (near to zero).
  + Let λ1………λp be the eigenvalues of correlation matrix. The condition number of correlation matrix is defined as K = √(λmax / λmin) & Kj = √(λmax / λj), j=1,2,…….,p.
  + Rule of Thumb: If one or more of the eigenvalues are small (close to zero) and the corresponding condition number is large, then it indicates multicollinearity (Montgomery,2001)

## Solutions : Shapley value regression :

Shapley value regression is a technique that solves the weaknesses of linear regression. This method calculates the relative importance of the independent variables (predictors or inputs) that are used to explain the values of the dependent variables in the linear regression.

The regression of the Shapley value considerably reduces the deleterious effects of collinearity on the estimated parameters of a regression equation.

Shapley's concept of value was introduced into game theory (cooperative collusion) in which agents form a collusion and cooperate with each other to increase the value of a game in their favour and then divide it among themselves.

# Linear regression and multicollinearity

| **Variable** | **Mean** | **Minimum** | **Maximum** |
| --- | --- | --- | --- |
| |  | | --- | | **Satisfaction** | | **q1a** | | **q1b** | | **q1c** | | **q2** | | **q3a** | | **q3b** | | **q3c** | | **q4a** | | **q4b** | | |  | | --- | | 7.5218412 | | 8.2382671 | | 8.2851986 | | 7.3516245 | | 7.9054152 | | 8.0231047 | | 7.4994585 | | 6.2588448 | | 7.3435018 | | 7.4119134 | | |  | | --- | | 6.0000000 | | 6.3000000 | | 7.1000000 | | 5.7000000 | | 6.6000000 | | 6.2000000 | | 5.1000000 | | 4.8000000 | | 6.2000000 | | 6.1000000 | | |  | | --- | | 8.3000000 | | 9.0000000 | | 9.0000000 | | 8.8000000 | | 8.6000000 | | 9.0000000 | | 8.9000000 | | 7.2000000 | | 8.2000000 | | 8.2000000 | |

## Overview the data

As we can see here, the average of Satisfaction is 7,5. It considered a Promoter note. The minimum note is 6 and the highest note is 8,3.

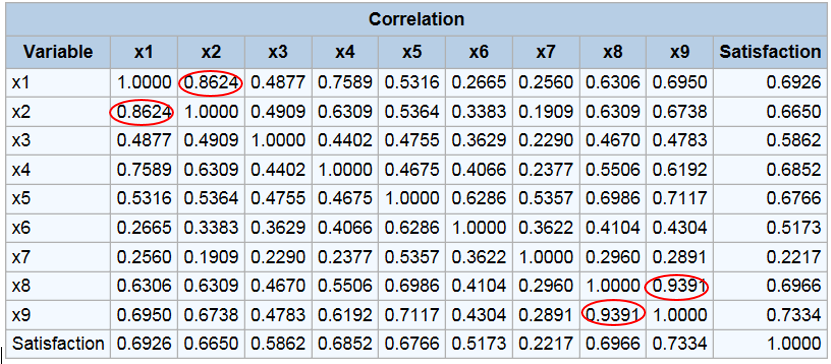
The first group is Information. This group have a high average note compared to another one: Ease of locating the service is 8,3 and finding information about the train is 8,2. We can conclude that the station layout is quite reasonable, enabling customers to capture information about their train journey easily.

Move to the next group, transfer represents the period when you move from the waiting platform to your train and vice versa. It has an average score of 7,9. It’s a good score.

The third group is Cleanliness and Security. This is the most problematic group in our study. The cleanliness of the station are highly appreciated by the clients but not for the toilet. The minimum note of the question about the toilet is 4,8. It is a detractor note. So clearly, we need a plan to ameliorate the situation. For example, build more toilet, upgrade equipment, increase the number of cleaning times a day. The average note of security just enough to reached the promotor level.

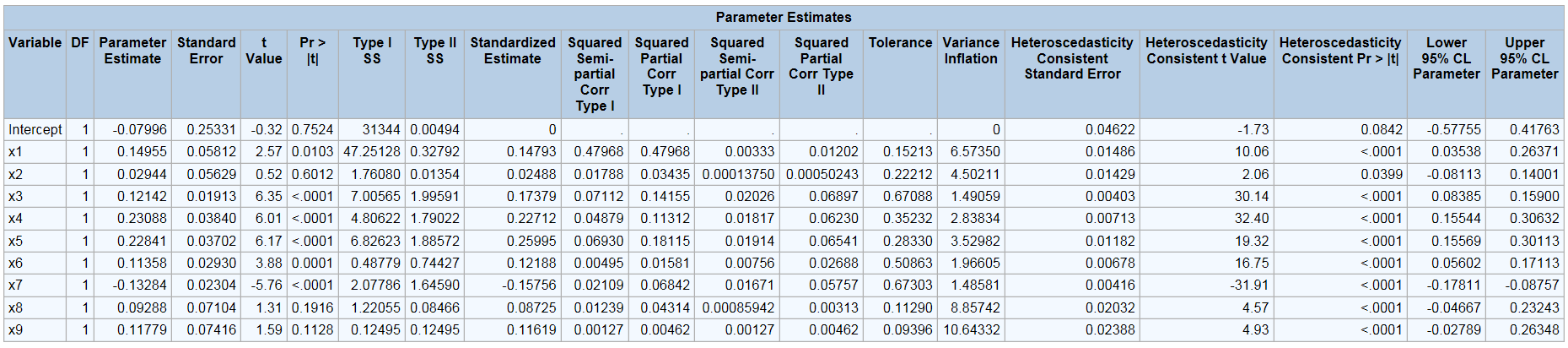
And the last group is comfort. The average note is note high enough to be a promotor note, but also about to reach the target. So what we can do here maybe analyse the verbatims of clients to find their needs, and have strategy to develop the comfort in train station (build a lounge for V.I.P clients, put more mobile’s charger station, etc.)

**Are there any correlated variables?**



As can be seen from this table, some variables are correlated with each other, notably Q1a and Q1b and Q4a and Q4b. This result suggests that we are facing a multicollinearity problem.

## OLS regression’s results



At the 5% level, everything else is equal:

When the rating of Q1a gains 1 pt on a scale of 10, customer satisfaction increases by almost 0.15 pts. I would like to point out that this variable is not significant because it is correlated with Q1b.

When the Q1b score gains 1 pt on a scale of 10, customer satisfaction increases by 0.03 pts. Just as variable Q1ariable Q1b is not significant.

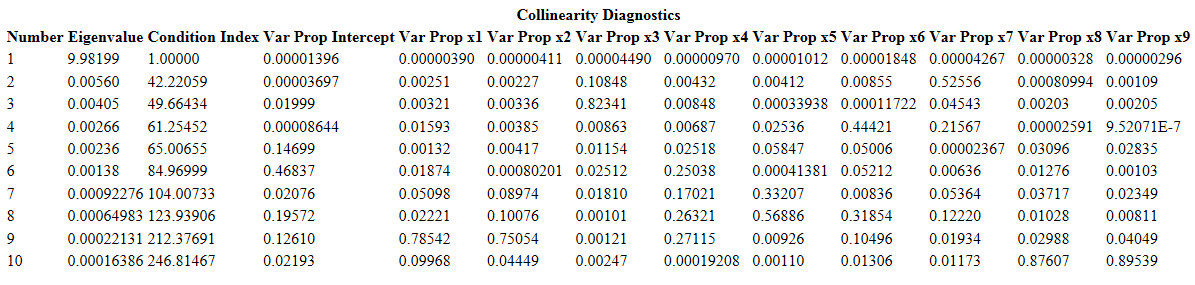
When the score of Q1c gains 1 pt on a scale of 10, customer satisfaction increases by 0.12 pts. Q1c is significant and this is also the case for Q2, Q3a, and Q3b.

When the Q3c rating gains 1 pt on a scale of 10, customer satisfaction drops by 0.13 pts which is unlikely but despite this it is significant

When the Q4a score gains 1 pt on a scale of 10, customer satisfaction increases by almost 0.09 pts, but it is not significant because of the collinearity between Q4a and Q4b

## Multicollinearity

Observe the Variance Inflation value (VIF), we found Q1a, Q4a values exceeds 5 and and Q4b exceed 10, which means the multicollinearity is present.



Next, based on the following table, we can see that some variables are colinear.

Like we can see here condition index in this model is extreme large. It begin with 1, then 42,2 and finish at 246,8. The corresponding of the eigenvalues are really small also. At the end the eigenvalues equal 0,0001636. So it indicates multicollinearity.

# Application of SVR

Shapley value regression consists of two steps:

First, each regressor’s importance in the model is assessed based on averaging R2 increases over all orderings of regressors.

Second, the coefficients in the model are adjusted to match the regressor importance assessment.

## SV and reg interperetation

The SV for each contributor is defined as:

The final results is the evaluations of the worth of contributor j over all possible combinations of their coalitions. This is the biggest difference of SVR from other variable importance measurement methods

## Computing SV

For each set of independent variables, S, and the interactions containing those variables, a linear regression is run and the value of is retained. Letting

the increase in from the contribution of adding variable i to S is computed by the γ-weighted difference. These values are computed for every set S not containing i and summed. The sign and value of k depend on the number of variables in a regression and whether variable i is in the regression.

# Methodology and results

We create several datasets by pulling repeated 100 bootstrap samples from the original dataset for each sample sizes of: N = 553 (original dataset size), N=450, N=300 and N=150.

We run Linear Regression and Shapley Value Regression and compare the results.

Firstly, we observe the result of the full size dataset.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| N = 553 / OLS | Sample 1 | Zeroed | Importance | Sample 2 | Zeroed | Importance | Gap |
| Orientation in the station | 0.07158 | 0.07158 | 6% | 0.23676 | 0.23676 | 21% | 15% |
| Ease of finding information | 0.08142 | 0.08142 | 7% | -0.02915 | 0 | 0% | 7% |
| Ease of locating the services | 0.1095 | 0.1095 | 10% | 0.1201 | 0.1201 | 11% | 1% |
| Ease of locating arrival and departure | 0.2831 | 0.2831 | 25% | 0.17889 | 0.17889 | 16% | 9% |
| Cleanliness of the station | 0.20444 | 0.20444 | 18% | 0.22719 | 0.22719 | 20% | 2% |
| Security | 0.12321 | 0.12321 | 11% | 0.14285 | 0.14285 | 13% | 2% |
| Cleanliness of the toilet | -0.17057 | 0 | 0% | -0.15564 | 0 | 0% | 0% |
| Comfort of waiting in the station | 0.09374 | 0.09374 | 8% | 0.08403 | 0.08403 | 7% | 1% |
| Time spent in the station | 0.15011 | 0.15011 | 13% | 0.1378 | 0.1378 | 12% | 1% |
| Average Gap between Sample 1 and 2 | | | | | | | 4% |
| Maximum Gap between Sample 1 and 2 | | | | | | | 15% |

For the purpose of demonstrate the gap importance between the first and the second sample, negative coefficients now are set to “0”. The average gap found between them is 4%, and the maximum is 15%

We produce now exact same test for each sample size. The table below shows the results of all methods.

|  |  |  |
| --- | --- | --- |
| SAMPLE SIZE | OLS | SVReg |
| Average Gap Between Sample 1 & 2 | | |
| N = 553 | 4% | 1% |
| N = 450 | 7% | 2% |
| N = 300 | 10% | 1% |
| N = 150 | 6% | 2% |
| Maximum Gap Between Sample 1 & 2 | | |
| N = 553 | 15% | 2% |
| N = 450 | 18% | 4% |
| N = 300 | 26% | 4% |
| N = 150 | 21% | 5% |

So the result of SVReg is clearly better than OLS. The average gap is constant. Maximum gaps of SVReg in the meantime are at least four times smaller than those of OLS.

Based on these new coefficients of Shapley Value Regression (Sample N = 554):

| **SV** | |
| --- | --- |
|  | **SV** |
| Q1A | 0.0937245 |
| Q1B | 0.0790006 |
| Q1C | 0.0781057 |
| Q2 | 0.1060902 |
| Q3A | 0.1004004 |
| Q3B | 0.0583481 |
| Q3C | 0.0152333 |
| Q4A | 0.0906286 |
| Q4B | 0.1049373 |

| **RSquare** |
| --- |
| 0.7264688 |

The OLS regression has a good coefficient of multiple determination R2 = 0.73, the quality of SV model is R2 = 0.7264688, so about 99% of OLS. It is a price in the trade-off for obtaining the models with meaningful coefficients useful both for analysis and prediction.

| **R2Share** | |
| --- | --- |
|  | **Share(%)** |
| Q1A | 12.90138 |
| Q1B | 10.874605 |
| Q1C | 10.751421 |
| Q2 | 14.603549 |
| Q3A | 13.820331 |
| Q3B | 8.0317445 |
| Q3C | 2.0969021 |
| Q4A | 12.475218 |
| Q4B | 14.444848 |

We can see clearly, the transfer to and from the platform (q2) is the most important point to clients. When the train come and go, the more streamlined the platform is, the easier it is for customers to find their train. Time to take old passengers out and pick up new passengers is shortened. Time is saved. The average score of q2 is not bad: 7,9 points, but we can always optimized it.

The second in the most important factor is time spent in the station (q4b). Waiting for the train is one of the inevitable things when you take the train. The company's mission is to make customers most comfortable during the waiting period. For example, design more waiting seats in the station, build a lounge exclusively reserved to loyal customers. The average score of q4b is 7,4 which mean it needs a plan to improve because its score is below the average satisfaction level (7,5).

In 9 questions, the one with the lowest average score is q3c – cleanliness of the toilet in the station (6,3 points). But luckily, in the point of views of clients, the important of this point is only 2%, so the score doesn’t have much effect on overall satisfaction.

From the above result, the company can make strategic plans based on the importance of the factors.

# Code SAS explanation

The code is quite long and complexe. For this reason, the instruction below cover only the major points.

First step, we use the PROC SURVEYSELECT statements to select a probability sample of satisfaction’s note from the original dataset by using simple random sampling. We modify the option “sampsize” four times by 553, 450, 300, 150 in order to have all the results needed, and “num” ,which present the sample number, by 1 and 2.

PROC MEANS with option “*cv”* is in order to look up coefficients of variation of each touchpoints.

Next, the code fragment below provides an implementation in the IML matrix language of the SAS system. It is embedded in a SAS macro with parameters for the database name, the dependent variable, and the regressors and group correspondence. It outputs the OLS results along with the decomposition of the goodness-of-fit.

This macro makes extensive use of the %sysfunc() and %qscan functions.

Now we count words in a macro variable, that is, strings of characters separated by blanks. Function %qscan() is used to select words when they are delimited by blanks. So each word is an independant variable name. Then, variable *&nb* is created to automatically count the number of variables used in the case.

Function %qscan with second argument &nb and third argument %str( ) is used to select individual variables when they are delimited by blanks..

Firstly, we get the results of linear model using PROC REG. Then with the macro, PROC REG runs all subsets regressions of &indepvar. The SubsetSelSummary option of ods select and SubsetSelSummary = RsquareValues ods output statement outputs selection summary for R², Adj-R² and Cp methods to the work library under the name RsquareValues. The option selection = Rsquare on the MODEL statement selects the maximum R² method of variable selection. These options help all subsets regressions to be performed and outputs R² of each.

Secondly, RsquareValues is split into two data: one “with i” and another “without i”. For example, “R with x1” contain every subsets with x1 in it. In opposite, “R without x1” contain the remaining subsets.

Thirdly, in order to find the actual weight of xi in each subset, the subtraction calculation of R² of subset with xi and the one without xi is set. The result is output to the data named R2\_delta.

And finally, a PROC IML used to call produces the Shapley values. It creates a matrix of indicators of the presence of independent variables to select the appropriate weights. For each variable, the R² of each regression is multiplied by its weight. The weights are summed for each variable as variable RSquare. The largest R² is the RSquare of the whole regression, so RSquare is divided by it to form R2Share.

A final PROC DATASETS call deletes the intermediate datasets created by the macro.

# Conclusions

# BIBLIOGRAPHY

Jane Tang and Jay Weiner, Multicollinearity in CSAT studies

Stan Lipovetsky and Michael Conklin, 2001, Analysis of regression in game theory approach, Applied Stochastic Models in Business and Industry, 17, 319-330.

Stan Lipovetsky and Michael Conklin, 2005, Incremental net effects in multiple regression, International Journal of Mathematical Education in Science and Technology, Vol. 36, No. 4, 361-373.

Stan Lipovetsky, 2006, Entropy Criterion In Logistic Regression And Shapley Value of Predictors, Journal of Modern Applied Statistical Methods, Vol 5, Issue 1, Article 9.

Charles D. Coleman, 2017, Decomposing the R-squared of a Regression Using the Shapley Value in SAS, SESUG Paper SD-61-2017

# 

# ANNEXES

LIBNAME in "C:\Users\Dan Vu\Desktop\Shapley value\data";

**PROC** **IMPORT** OUT=in.data

DATAFILE= "C:\Users\Dan Vu\Desktop\Shapley value\data\org2.csv" DBMS=csv Replace;

DELIMITER=",";

GETNAMES=YES;

**RUN**;

%let rep\_data\_in = C:\Users\Dan Vu\Desktop\Shapley value\data;

libname in "&rep\_data\_in.";

%let indepvar = q1a q1b q1c q2 q3a q3b q3c q4a q4b; /\*Independent Variables\*/

%let depvar = Satisfaction; /\*dependent Variable\*/

%let Size = 554; /\*Orignal size = 554 obs / Samplesize test: 554, 450, 300, 150\*/

%let NumSamples = 100;

%let num = 1; /\*1 and 2\*/

**proc** **surveyselect** data=org2 NOPRINT seed=**1**

out=BootCases(rename=(Replicate=SampleID))

method=urs /\* resample with replacement \*/

sampsize=&Size

reps=&NumSamples; /\* generate NumSamples bootstrap resamples \*/

**run**;

**DATA** BootcasesCV;

SET BootCases;

KEEP &indepvar.;

**RUN**;

ods html file = 'C:\Users\Dan Vu\Desktop\Shapley value\data\554.html' ;

**PROC** **MEANS** data=BootcasesCV;

**RUN**;

**%macro** BC(sample = &num);

DATA BootCasesn&num.;

SET BootCases (WHERE = (SampleID = &num.));

RUN;

**%mend**;

%***BC***(sample = &num);

**%macro** ***Shapley***;

/\*Count No of Independent Variable\*/

%local nb word;

%let nb=1;

%let word=%qscan(&indepvar,&nb,%str( ));

%do %while(&word ne);

%let nb=%eval(&nb+1);

%let word=%qscan(&indepvar,&nb,%str( ));

%end;

%let nb=%eval(&nb-1);

/\*%put &=nvar.;

%put &=word.;\*/

DATA \_temp\_;

SET BootCasesn&num;

KEEP &indepvar &depvar;

RUN;

/\*Rename Variables\*/

DATA \_temp\_;

SET \_temp\_;

RENAME

%do i=**1** %to &nb;

%let var&i=%sysfunc(scan(&indepvar,&i,' '));

&&var&i =x&i %end ;;

RUN;

/\*Reg with selection=rsquare\*/

ods graphics on;

PROC REG DATA=\_temp\_ corr all;

ods select Corr NObs ANOVA ParameterEstimates SubsetSelSummary CollinDiag;

ods output SubsetSelSummary=RsquareValues Corr=Corr;

MODEL &depvar=x1-x&nb. / tol vif collin SELECTION=RSQUARE;

RUN;

ods graphics off;

DATA Corr (RENAME=(Satisfaction = Ryj));

SET Corr;

KEEP Variable Satisfaction;

IF Variable = "Satisfaction" then delete;

RUN;

DATA RsquareValues;

SET RsquareValues;

DROP Dependent Model control Modelindex;

RUN;

DATA Null\_Model;

NumInModel =**0**;

Rsquare=**0**;

VarsInModel='x0';

RUN;

PROC APPEND BASE=RsquareValues DATA=null\_model;

RUN;

PROC SORT DATA=RsquareValues;

BY NumInModel VarsInModel;

RUN;

/\*Shapley Value computation\*/

%do i=**1** %to &nb;

DATA R2\_with\_&i.( RENAME=(RSquare=Rsquare\_with VarsInModel=x\_with))

R2\_without\_&i. (DROP =NumInModel RENAME=(RSquare=Rsquare\_without VarsInModel=x\_without));

SET RsquareValues;

IF FIND(VarsInModel,"x&i") GT **0** THEN OUTPUT R2\_with\_&i.;

ELSE OUTPUT R2\_without\_&i.;

RUN;

DATA R2\_delta\_&i.;

SET R2\_with\_&i.;

SET R2\_without\_&i. ;

RUN;

DATA R2\_delta\_&i.;

SET R2\_delta\_&i.;

KEEP x&i. NumInModel;

x&i.=Rsquare\_with-Rsquare\_without;

RUN;

%end;

DATA R2\_delta;

MERGE %do i=**1** %to &nb;

R2\_delta\_&i. %end; ;

RENAME

%do i=**1** %to &nb;

%let var&i=%sysfunc(SCAN(&indepvar,&i,' '));

x&i=&&var&i %end ; ;

RUN;

PROC IML;

USE R2\_delta;

READ ALL VAR {&indepvar} INTO indvar;

READ ALL VAR {NumInModel} INTO NIM;

p=ncol(indvar); /\*p = n\*/

q=nrow(indvar);

sv=j(q,p,**0**);

do j=**1** to p;

do i=**1** to q;

do m=**1** to NIM[i,];

gamma=**0**;

w=fact(NIM[i,]-**1**)#(fact(p-(NIM[i,]-**1**)-**1**)); /\*# = multiplication \*/

/\*fact(NIM[i,]-1) = k! p = n\*/

w=w/fact(p);

gamma=gamma+w; /\*ve dau\*/

sv[i,j]=gamma#indvar[i,j]; /\*delta la ve sau\*/

end;

end;

end;

sv=SV[+,]`; /\*[+,] = reduction des lignes par la somme; ` = transpose\*/

RSquare=sv[+,];

R2Share = sv/RSquare\***100**;

Question={&indepvar};

Cname={SV Var};

Share={"Share(%)"};

Total = "100%";

PRINT SV [rowname=Question colname=Cname] ;

PRINT RSquare ;

PRINT R2Share [rowname=Question colname=Share] ;

PRINT Total;

CREATE SV var {Question sv};

APPEND;

CLOSE SV;

QUIT;

PROC DATASETS lib=work

nolist kill;

QUIT;

RUN;

**%MEND** Shapley;

%***Shapley***;

ods html close ;