We are IntechOpen, the world's leading publisher of Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Identification of Key Drivers of Net Promoter Score Using a Statistical Classification Model

Daniel R. Jeske[1], Terrance P. Callanan[2] and Li Guo[3]
[1]Professor and Chair, Department of Statistics,
University of California, Riverside, CA
[2]Customer Experience Manager, Six Sigma Champion,
Carestream Health, Rochester, NY
[3]Graduate Student, Department of Statistics,
University of California, Riverside, CA
USA

1. Introduction

Net Promoter Score (NPS) is a popular metric used in a variety of industries for measuring customer advocacy. Introduced by Reichheld (2003), NPS measures the likelihood that an existing customer will recommend a company to another prospective customer. NPS is derived from a single question that may be included as part of a larger customer survey. The single question asks the customer to use a scale of 0 to 10 to rate their willingness and intention to recommend the company to another person. Ratings of 9 and 10 are used to characterize so-called 'promoters,' ratings of 0 through 6 characterize 'detractors,' and ratings of 7 and 8 characterize 'passives.' The NPS is calculated as the percentage of respondents that are promoters minus the percentage of respondents that are detractors.

The idea behind the labels given to customers is as follows. Promoters are thought to be extremely satisfied customers that see little to no room for improvement, and consequently would offer persuasive recommendations that could lead to new revenue. The passive ratings, on the other hand, begin to hint at room for improvement and consequently the effectiveness of a recommendation from a Passive may be muted by explicit or implied caveats. Ratings at the low end are thought to be associated with negative experiences that might cloud a recommendation and likely scare off prospective new customers. Additional discussion on the long history of NPS can be found in Hayes (2008).

Some implementations of NPS methodology use reduced 5-point or 7-point scales that align with traditional Likert scales. However it is implemented, the hope is that movements in NPS are positively correlated with revenue growth for the company. While Reichheld's research presented some evidence of that, other findings are not as corroborative (Kenningham et al., 2007). Regardless of whether there is a predictive relationship between NPS and revenue growth, implementing policies and programs within a company that improve NPS is an intuitively sensible thing to do [see, for example, Vavra (1997)]. A difficult and important question, however, is how to identify key drivers of NPS. *Calculating* NPS alone does not do this.

This chapter is an illustrative tutorial that demonstrates how a statistical classification model can be used to identify key drivers of NPS. Our premise is that the classification model, the data it operates on, and the analyses it provides could usefully form components of a Decision Support System that can not only provide both snapshot and longitudinal analyses of NPS performance, but also enable analyses that can help suggest company initiatives aimed toward lifting the NPS.

We assume that the NPS question was asked as part of larger survey that also probed customer satisfaction levels with respect to various dimensions of the company's services. We develop a predictive classification model for customer advocacy (promoter, passive or detractor) as a function of these service dimensions. A novelty associated with our classification model is the optional use of constraints on the parameter estimates to enforce a monotonic property. We provide a detailed explanation of how to fit the model using the SAS software package and show how the fitted model can be used to develop company policies that have promise for improving the NPS. Our primary objective is to teach an interested practitioner how to use customer survey data together with a statistical classifier to identify key drivers of NPS. We present a case study that is based on a real-life data collection and analysis project to illustrate the step-by-step process of building the linkage between customer satisfaction data and NPS.

2. Logistic regression

In this section we provide a brief review of logistic and multinomial regression. Allen and Rao (2000) is a good reference that contains more detail than we provide, and additionally has example applications pertaining to customer satisfaction modeling.

2.1 Binomial logistic regression

The binomial logistic regression model assumes that the response variable is binary (0/1). This could be the case, for example, if a customer is simply asked the question "Would you recommend us to a friend?" Let $\{Y_i\}_{i=1}^{n}$ denote the responses from n customers, assigning a "1" for Yes and "0" for No. Suppose a number of other data items (covariates) are polled from the customer on the same survey instrument. These items might measure the satisfaction of the customer across a wide variety of service dimensions and might be measured on a traditional Likert scale. We let $\underset{\sim}{x}_i$ denote the vector of covariates for the i-th sampled customer and note that it reflects the use of dummy variable coding for covariates that are categorical scale. For example, if the first covariate is measured on a 5-point Likert scale, its value is encoded into $\underset{\sim}{x}_i$ by using five dummy variables $\{x_{1j}\}_{j=1}^{5}$, where $x_{1j} = 1$ if and only if the Likert response is j.

The binomial logistic regression model posits that Y_i is a Bernoulli random variable (equivalently, a binomial random variable with trial size equal to one) with success probability p_i, and further, that the success probability is tied to the covariates through the so-called link function $p_i = \exp(\alpha + \underset{\sim}{\beta}' \underset{\sim}{x}_i) / [1 + \exp(\alpha + \underset{\sim}{\beta}' \underset{\sim}{x}_i)]$, where $\underset{\sim}{\beta}$ is a vector of model parameters (slopes). Continuing with the 5-point Likert scale example above, there would be five slopes $\{\beta_{1j}\}_{j=1}^{5}$ associated with the five dummy variables $\{x_{1j}\}_{j=1}^{5}$ used to code the first covariate.

Model fitting for the binomial logistic regression model entails estimating the parameters α and $\underset{\sim}{\beta}$ via maximum likelihood. The likelihood function for this model is

$$L(\alpha, \underset{\sim}{\beta}) = \prod_{i=1}^{n} \frac{\exp[(\alpha + \underset{\sim}{\beta}'\underset{\sim}{x}_i)Y_i]}{1 + \exp(\alpha + \underset{\sim}{\beta}'\underset{\sim}{x}_i)} \tag{1}$$

and the maximum likelihood estimate (MLE) of (α, β') is the value, say $(\hat{\alpha}, \hat{\beta}')$, that maximizes this function. Once the MLE is available, the influence of the covariates can be assessed by the magnitude (relative to their standard error) of the individual slopes. In particular, it can be ascertained which attributes of customer service have a substantial effect on making the probability of a 'Yes' response high.

2.2 Multinomial logistic regression

Suppose now that the outcome variable has more than two categories. For example, suppose the responses $\{Y_i\}_{i=1}^{n}$ are measured on the 11-point NPS scale. The multinomial logistic model seeks to represent the probability of *each* response as a function of the covariates. Since each response is an ordinal categorical variable taking on values in the set $\{0, 1, \ldots, 10\}$, we consider a multinomial logistic regression model with 10 cumulative link functions:

$$\Pr(Y_i \le j) = \frac{\exp(\alpha_j + \underset{\sim}{\beta}'\underset{\sim}{x}_i)}{1 + \exp(\alpha_j + \underset{\sim}{\beta}'\underset{\sim}{x}_i)} \quad , \quad j = 0, 1, \ldots, 9 \tag{2}$$

where $\{\alpha_j\}_{j=0}^{9}$ are intercept parameters and $\underset{\sim}{\beta}$ is again a vector of slope parameters. In order to affect the required non-decreasing behavior (relative to j) of the right hand sides of the link functions, the constraint $\alpha_0 \le \alpha_2 \cdots \le \alpha_9$ is imposed on the intercept parameters. Starting with $\Pr(Y_i = 0) = \Pr(Y_i \le 0)$, and then differencing the above expressions in (2) as in $\Pr(Y_i = j) = \Pr(Y_i \le j) - \Pr(Y_i \le j-1)$, $j = 1, \ldots, 10$, we obtain expressions for the individual probabilities of the response as a function of the intercept and slope parameters. Defining the responses $\{Y_{ij}\}_{j=0}^{10}$ to be one (zero) if and only if $Y_i = j$, the likelihood function for this model is

$$L(\underset{\sim}{\alpha}, \underset{\sim}{\beta}) = \prod_{i=1}^{n} \prod_{j=0}^{10} P(Y_i = j)^{Y_{ij}} \tag{3}$$

and the MLE of $(\underset{\sim}{\alpha}, \beta')$ is the value, say $(\hat{\underset{\sim}{\alpha}}, \hat{\beta}')$, that maximizes this function. Once the MLE is available, the magnitudes of the slope estimates (relative to their standard errors) can be used to identify the covariates that push the distribution of the response towards 9s and 10s. We note that the constraint on the intercepts is a standard constraint. In the next section, we will discuss an additional and novel constraint that can optionally be imposed on the slope parameters.

3. Case study

Carestream Health, Inc. (CSH) was formed in 2007 when Onex Corporation of Toronto, Canada purchased Eastman Kodak Company's Health Group and renamed the business as Carestream Health. They are an annual \$2.5B company and a world leader in medical imaging (digital and film), healthcare information systems, dental imaging and dental practice management software, molecular imaging and non-destructive testing. Their

customers include medical and dental doctors and staff and healthcare IT professionals in small offices and clinics to large hospitals and regional and national healthcare programs. A major company initiative is to create a sustainable competitive advantage by delivering the absolute best customer experience in the industry. Customer recommendations are key to growth in the digital medical space and no one has been able to do it consistently well. The foundation for taking advantage of this opportunity is to understand what's important to customers, measure their satisfaction and likelihood to recommend based on their experiences, and drive improvement.

While descriptive statistics such as trend charts, bar charts, averages and listings of customer verbatim comments are helpful in identifying improvement opportunities to improve the Net Promoter Score (NPS), they are limited in their power. First, they lack quantitative measurements of correlation between elements of event satisfaction and NPS. As a consequence, it is not clear what impact a given process improvement will have on a customer's likelihood to recommend. Second, they lack the ability to view multi-dimensional relationships – they are limited to single factor inferences which may not sufficiently describe the complex relationships between elements of a customer's experience and their likelihood to recommend.

This section summarizes the use of multinomial logistic regression analyses that were applied to 5056 independent customer experience surveys from Jan 2009 – Jan 2010. Each survey included a question that measured (on a 5-point Likert scale) how likely it would be for the customer to recommend colleagues to purchase imaging solutions from CSH. Five other questions measured the satisfaction level (on a 7-point Likert scale) of the customer with CSH services obtained in response to an equipment or software problem. Key NPS drivers are revealed through the multinomial logistic regression analyses, and improvement scenarios for specific geographic and business combinations are mapped out. The ability to develop a quantitative model to measure the impact on NPS of potential process improvements significantly enhances the value of the survey data.

3.1 CSH Customer survey data

The 5-point Likert response to the question about willingness to recommend is summarized in Table 1 below. CSH calculates a unique net promoter score from responses on this variable using the formula $\text{NPS} = \sum_{i=1}^{5} w_i \hat{p}_i$, where $\underset{\sim}{w} = (-1.25, -0.875, -0.25, 0.75, 1.0)'$ is a vector of weights and where $\hat{p}_i$ is the estimated proportion of customers whose recommendation score is i. Two interesting characteristics of the weight vector are, first, the penalty for a 1 (2) exceeds the benefit of a 5 (4), and second, the negative weight for a neutral score is meant to drive policies toward delighting customers.

Recommendation	Interpretation
1	Without being asked, I will advise others NOT to purchase from you
2	Only if asked, I will advise others NOT to purchase from you
3	I am neutral
4	Only if asked, I will recommend others TO purchase from you
5	Without being asked, I will recommend others TO purchase from you

Table 1. Meaning of Each Level of Recommendation Score

Let $Y \in \{1,2,3,4,5\}$ be a random variable denoting the willingness of a particular customer to recommend CSH. The values $\{p_i\}_{i=1}^5$ represent the theoretical probabilities of the possible values for Y. That is, $p_i = \Pr(Y = i)$. The multinomial logistic regression model treats the customer demographic variables and the customer satisfaction ratings as the covariates $\underline{x}_i$, linking their values to the probability distribution of Y such that their influence on the values $\{p_i\}_{i=1}^5$ can be ascertained. Since the expected value of NPS is a function of these $\{p_i\}_{i=1}^5$ values, the end result is a model that links NPS to what customers perceive to be important. This linkage can then be exploited to determine targeted pathways to improve NPS via improvement plans that are customer-driven.

The demographic covariates include the (global) region code, country code, business code and the customer job title. The demographic covariates are coded using the standard dummy variable coding technique. For example, region code utilizes 7 binary variables $\{RC_i\}_{i=1}^7$ where

$$RC_i = \begin{cases} 1 & \text{if customer falls in i-th global region} \\ 0 & \text{otherwise}. \end{cases}$$

Country code utilizes similar dummy variables, but because countries are nested within regions we use the notation $\{CC_{j(i)}\}_{i=1}^7 \, _{j=1}^{n_i}$, where n_i is the number of country codes within the i-th region and where

$$CC_{j(i)} = \begin{cases} 1 & \text{if customer falls in j-th country within } i\text{-th region} \\ 0 & \text{otherwise}. \end{cases}$$

In the data set we have $\{n_i\}_{i=1}^7 = \{3,1,1,7,5,4,2\}$. Business code utilizes two dummy variables $\{BC_i\}_{i=1}^2$ where

$$BC_i = \begin{cases} 1 & \text{if customer is aligned with } i\text{-th business code} \\ 0 & \text{otherwise}. \end{cases}$$

Finally, job title utilizes 10 dummy variables $\{JT_i\}_{i=1}^{10}$ where

$$JT_i = \begin{cases} 1 & \text{if customer has } i\text{-th job title} \\ 0 & \text{otherwise}. \end{cases}$$

The customer satisfaction covariates are also coded using the dummy variable scheme. The data on these covariates are the responses to the survey questions identified as q79, q82a, q82b, q82d and q82f. These questions survey the customer satisfaction on 'Overall satisfaction with the service event,' 'Satisfaction with CSH knowledge of customer business and operations,' 'Satisfaction with meeting customer service response time requirements,' 'Satisfaction with overall service communications,' and 'Satisfaction with skills of CSH employees,' respectively.

Survey questions q82c and q82e, which survey satisfaction with 'Time it took to resolve the problem once work was started,' and 'Attitude of CSH employees' were also considered as covariates, but they did not show themselves to be statistically significant in the model. Their absence from the model does not necessarily imply they are not important drivers of

their overall satisfaction with CSH, but more likely that their influence is correlated with the other dimensions of overall satisfaction that are in the model. Each customer satisfaction covariate is scored by customers using a 7-point Likert scale (where '1' indicates the customer is "extremely dissatisfied" and '7' indicates "extremely satisfied"), and thus each utilizes 7 dummy variables in the coding scheme. We denote these dummy variables as $\{q79_i\}_{i=1}^7$, $\{q82a_i\}_{i=1}^7$, $\{q82b_i\}_{i=1}^7$, $\{q82d_i\}_{i=1}^7$, and $\{q82f_i\}_{i=1}^7$, respectively, and they are defined as follows:

$$q79_i = \begin{cases} 1 & \text{if customer response to q79 is } i \\ 0 & \text{otherwise}, \end{cases}$$

$$q82a_i = \begin{cases} 1 & \text{if customer response to q82a is } i \\ 0 & \text{otherwise}, \end{cases}$$

$$q82b_i = \begin{cases} 1 & \text{if customer response to q92b is } i \\ 0 & \text{otherwise}, \end{cases}$$

$$q82d_i = \begin{cases} 1 & \text{if customer response to q82d is } i \\ 0 & \text{otherwise}, \end{cases}$$

$$q82f_i = \begin{cases} 1 & \text{if customer response to q82f is } i \\ 0 & \text{otherwise}. \end{cases}$$

Assembling all of the covariates together, we then have a total of 77 covariates in $\underset{\sim}{x}$. Thus, the vector of slopes $\underset{\sim}{\beta}$ in the link equations has dimension 77×1. Combined with the 4 intercept parameters $\{\alpha_i\}_{i=1}^4$, the model we have developed has a total of 81 parameters. We note it is conceivable that interactions between the defined covariates could be important contributors to the model. However, interaction effects based on the current data set were difficult to assess because of confounding issues. As the data set gets larger over time, it is conceivable the confounding issues could be resolved and interaction effects could be tested for statistical significance.

3.2 Model fitting and interpretation

The SAS code for obtaining maximum likelihood estimates (MLEs) for the model parameters $\{\alpha_i\}_{i=1}^4$ and $\underset{\sim}{\beta}$ is shown in Appendix A. Lines 1-4 are used to read in the data that is stored as a space delimited text file 'indata.txt' that is located in the indicated directory. All of the input variables on the file are coded as integer values. The PROC LOGISTIC section of the code (lines 5-10) directs the fitting of the multinomial logistic regression model. The class statement is used to specify that all of the covariate variables are categorical in nature, and the param=glm option specifies to use the dummy variable coding scheme that was defined in the previous section. Table 2 summarizes the portion of the SAS output that reports the maximum likelihood estimates for $\{\alpha_i\}_{i=1}^4$ and $\underset{\sim}{\beta}$. Note that the zero for the slope of the last level of each covariate is a structural zero resulting from the non-full rank dummy variable coding used when fitting the model.

Parm.	Est.	Parm.	Est.	Parm.	Est.	Parm.	Est.
α_1	-7.80	$\beta_{18}\ (CC_{6(4)})$	-.41	$\beta_{39}\ (JT_7)$	-.15	$\beta_{60}\ (q82b_4)$	.53
α_2	-5.69	$\beta_{19}\ (CC_{7(4)})$	0	$\beta_{40}\ (JT_8)$	-.28	$\beta_{61}\ (q82b_5)$	.14
α_3	-2.34	$\beta_{20}\ (CC_{1(5)})$	-.11	$\beta_{41}\ (JT_9)$	-.62	$\beta_{62}\ (q82b_6)$	.23
α_4	-.045	$\beta_{21}\ (CC_{2(5)})$	.092	$\beta_{42}\ (JT_{10})$	0	$\beta_{63}\ (q82b_7)$	0
$\beta_1\ (RC_1)$	.11	$\beta_{22}\ (CC_{3(5)})$	-1.63	$\beta_{43}\ (q79a_1)$	1.92	$\beta_{64}\ (q82d_1)$	1.27
$\beta_2\ (RC_2)$	.59	$\beta_{23}\ (CC_{4(5)})$	.11	$\beta_{44}\ (q79a_2)$	2.09	$\beta_{65}\ (q82d_2)$	.92
$\beta_3\ (RC_3)$	1.23	$\beta_{24}\ (CC_{5(5)})$	0	$\beta_{45}\ (q79a_3)$	1.43	$\beta_{66}\ (q82d_3)$	.67
$\beta_4\ (RC_4)$	-.13	$\beta_{25}\ (CC_{1(6)})$	-.13	$\beta_{46}\ (q79a_4)$	.84	$\beta_{67}\ (q82d_4)$	.77
$\beta_5\ (RC_5)$	.59	$\beta_{26}\ (CC_{2(6)})$	.34	$\beta_{47}\ (q79a_5)$	.58	$\beta_{68}\ (q82d_5)$	.44
$\beta_6\ (RC_6)$	.40	$\beta_{27}\ (CC_{3(6)})$	-.23	$\beta_{48}\ (q79a_6)$	.19	$\beta_{69}\ (q82d_6)$	.19
$\beta_7\ (RC_7)$	0	$\beta_{28}\ (CC_{4(6)})$	0	$\beta_{49}\ (q79a_7)$	0	$\beta_{70}\ (q82d_7)$	0
$\beta_8\ (CC_{1(1)})$	-.45	$\beta_{29}\ (CC_{1(7)})$	.42	$\beta_{50}\ (q82a_1)$	2.68	$\beta_{71}\ (q82f_1)$	.85
$\beta_9\ (CC_{2(1)})$	-.60	$\beta_{30}\ (CC_{2(7)})$	0	$\beta_{51}\ (q82a_2)$	.71	$\beta_{72}\ (q82f_2)$	1.69
$\beta_{10}\ (CC_{3(1)})$	0	$\beta_{31}\ (BC_1)$	-.21	$\beta_{52}\ (q82a_3)$	1.05	$\beta_{73}\ (q82f_3)$	1.08
$\beta_{11}\ (CC_{1(2)})$	0	$\beta_{32}\ (BC_2)$	0	$\beta_{53}\ (q82a_4)$	.89	$\beta_{74}\ (q82f_4)$	.69
$\beta_{12}\ (CC_{1(3)})$	0	$\beta_{33}\ (JT_1)$	-.097	$\beta_{54}\ (q82a_5)$	.31	$\beta_{75}\ (q82f_5)$	.59
$\beta_{13}\ (CC_{1(4)})$	.78	$\beta_{34}\ (JT_2)$	-.25	$\beta_{55}\ (q82a_6)$	.14	$\beta_{76}\ (q82f_6)$	.16
$\beta_{14}\ (CC_{2(4)})$	-.53	$\beta_{35}\ (JT_3)$	-.20	$\beta_{56}\ (q82a_7)$	0	$\beta_{77}\ (q82f_7)$	0
$\beta_{14}\ (CC_{3(4)})$	.83	$\beta_{36}\ (JT_4)$	-.48	$\beta_{57}\ (q82b_1)$	1.31		
$\beta_{16}\ (CC_{4(4)})$	-.050	$\beta_{37}\ (JT_5)$	.11	$\beta_{58}\ (q82b_2)$	.81		
$\beta_{17}\ (CC_{5(4)})$	.24	$\beta_{38}\ (JT_6)$	-.47	$\beta_{59}\ (q82b_3)$	.75		

Table 2. Maximum Likelihood Estimates of $\{\alpha_i\}_{i=1}^4$ and $\underset{\sim}{\beta}$

The section of the PROC LOGISTIC output entitled 'Type-3 Analysis of Effects' characterizes the statistical significance of the covariates through p-values obtained by referencing a Wald chi-square test statistic to a corresponding null chi-square distribution. Table 3 shows the chi-square tests and the corresponding p-values, and it is seen that all covariate groups are highly significant contributors in the model.

One way to assess model adequacy for multinomial logistic regression is to use the model to predict Y and then examine how well the predicted values match the true values of Y. Since the output of the model for each customer is an estimated probability distribution for Y, a natural predictor of Y is the mode of this distribution. We note that this predictor considers equal cost for all forms of prediction errors. More elaborate predictors could be derived by assuming a more complex cost model where, for example, the cost of predicting 5 when the actual value is 1 is higher than the cost of predicting 5 when the actual value is 4. Table 4, the so-called confusion matrix of the predictions, displays the cross classification of all 5056 customers based on their actual value of Y and the model-predicted value of Y.

Covariate Group	Degrees of Freedom	Wald Statistic	p-value
RC	6	41.2	< .0001
CC	16	40.9	< .01
BC	1	7.9	< .01
JT	9	43.7	< .0001
q79	6	84.8	< .0001
q82a	6	56.5	< .0001
q82b	6	34.4	< .0001
q82d	6	34.8	< .0001
q82f	6	39.9	< .0001

Table 3. Statistical Significance of Covariate Groups

Actual Y	Predicted Y					
	1	2	3	4	5	Total
1	3	2	7	4	4	20
2	3	8	48	22	7	88
3	2	3	342	486	133	966
4	0	0	126	1233	723	2082
5	0	0	39	705	1156	1900
Total	8	13	562	2450	2023	5056

Table 4. Confusion Matrix of Multinomial Logistic Regression Model

A perfect model would have a confusion matrix that is diagonal indicating the predicted value for each customer coincided identically with the true value. Consider the rows of Table 4 corresponding to Y=4 and Y=5. These two rows account for almost 80% of the customers in the sample. It can be seen that in both cases, the predicted value coincides with the actual value about 60% of the time. Neither of these two cases predicts Y=1 or Y=2, and only 4% of the time is Y=3 predicted. The mean values of the predicted Y when Y=4 and Y=5 are 4.28 and 4.59, respectively. The 7% positive bias for the case Y=4 is roughly offset by the 11.8% negative bias for the case Y=5.

Looking at the row of Table 4 corresponding to Y=3, we see that 86% of the time the predicted Y is within 1 of the actual Y. The mean value of the predicted Y is 3.77, indicating a 26% positive bias. Considering the rows corresponding to Y=1 and Y=2, where only about 2% of the customers reside, we see the model struggles to make accurate predictions, often over-estimating the actual value of Y. A hint as to the explanation for the noticeable over-estimation associated with the Y=1, Y=2 and Y=3 customers is revealed by examining their responses to the covariate questions. As just one example, the respective mean scores on question q79 ("Overall satisfaction with the service event") are 3.8, 4.1 and 5.2. It seems a relatively large number of customers that give a low response to Y are inclined to simultaneously give favorable responses to the covariate questions on the survey. Although this might be unexpected, it can possibly be explained by the fact that the covariate questions are relevant to the most recent service event whereas Y is based on a customer's cumulative experience.

Overall, Table 4 reflects significant lift afforded by the multinomial logistic regression model for predicting Y. For example, a model that utilized no covariate information would have a

confusion matrix whose rows were constant, summing to the row total. In sum, we feel the accuracy of the model is sufficient to learn something about what drives customers to give high responses to Y, though perhaps not sufficient to learn as much about what drives customers to give low responses to Y.

Figure 1 is a graphical display of the slopes for each of the customer satisfaction covariates. The larger the coefficient value, the more detrimental the response level is to NPS. The y-axis is therefore labeled as 'demerits.'
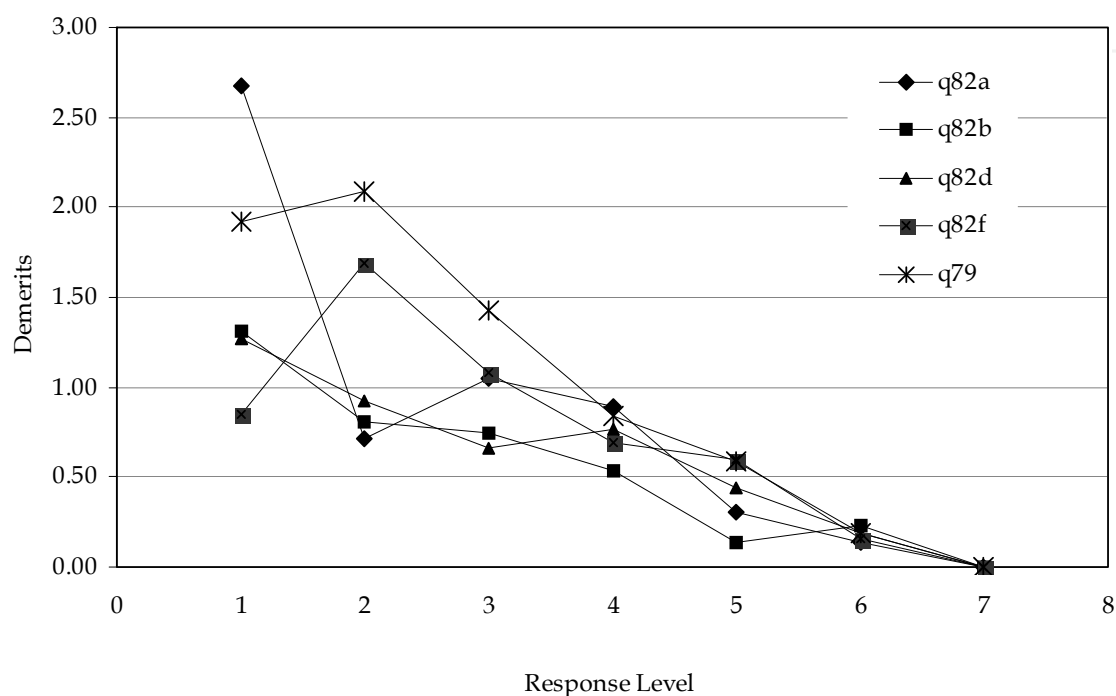

Fig. 1. MLEs of Slopes for 7-Point Likert Scale Customer Satisfaction Covariates

In view of the ordinal nature of the customer satisfaction covariates, the slopes, which represent the effect of the Likert scale levels, should decrease monotonically. That is, the penalty for a 'satisfied' covariate value should be less than or equal to that of a 'dissatisfied' covariate value. As such, it would be logical to have the estimated values of the slopes display the monotone decreasing trend as the response level of the covariates ascends. Figure 1 shows that the unconstrained MLEs for the slopes associated with the customer satisfaction covariates nearly satisfy the desired monotone property, but not exactly. The aberrations are due to data deficiencies or minor model inadequacies and can be resolved by using a constrained logistic regression model introduced in the next section.

3.3 Constrained logistic regression

Consider the situation where the i-th covariate is ordinal in nature, perhaps because it is measured on a k-point Likert scale. The CSH data is a good illustration of this situation, since all the customer satisfaction covariates are ordinal variables measured on 7-point Likert scale. Let the corresponding group of k slopes for this covariate be denoted by $\{\beta_{ij}\}_{j=1}^{k}$. In order to reflect the information that the covariates are ordered, it is quite natural

to impose the monotone constraint $\beta_{i1} \geq \beta_{i2} \cdots \geq \beta_{ik}$ onto the parameter space. Adding these constraints when finding the MLEs complicates the required maximization of the likelihoods in (1) and (3). In this section, however, we will show how this can be done using SAS with PROC NLP.

In order to simplify our use of PROC NLP, it is convenient to work with a full-rank parameterization of the logistic regression model. Because countries are nested within regions, a linear dependency exists between the dummy variables corresponding to regions and countries within regions. We can eliminate the linear dependency by removing region from the model and specifying country to be non-nested factor. The result of this model reparameterization is that instead of 6 degrees of freedom in the model for regions and 16 degrees of freedom for countries nested within regions, we equivalently have 22 degrees of freedom for countries. For the same purpose, we also redefine the dummy variable coding used for other categorical and ordinal covariates by using a full rank parameterization scheme. In particular, we use k-1 dummy variables (rather than k) to represent a k-level categorical or ordinal variable. With the full rank parameterization, the highest level of customer satisfaction has a slope parameter that is fixed to be 0. Lines 3-10 in the SAS code shown in Appendix B are used to set up the full rank parameterization of the logistic regression model.
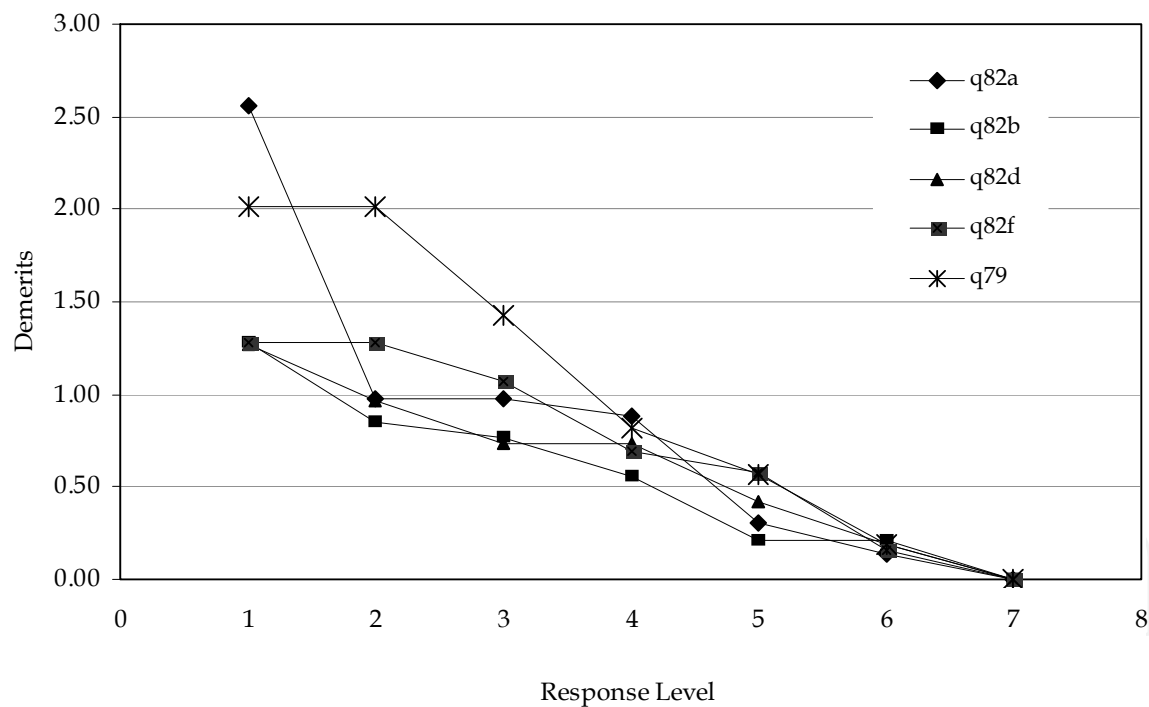

Fig. 2. Constrained MLEs of Slopes for 7-Point Likert Scale Customer Satisfaction Covariates

Beginning with line 12 in the SAS code, PROC NLP is used to derive the MLEs of the parameters under the constrained parameter space. The 'max' statement (line 13) indicates the objective function is the log-likelihood function of the model and that it is to be maximized. The maximization is carried out using a Newton-Raphson algorithm, and the 'parms' statement (line 14) specifies initial values for the intercept and slope parameters. The SAS variables bqj, baj, bbj, bdj and bfj are used to symbolize the slope parameters corresponding to the j-th response level of the customer satisfaction covariates q79, q82a,

q82b, q82d and q82f. Similarly, bcc$j$, bbc$j$, and bj are used to denote the slopes associated with different countries, business codes and job titles. The 'bounds' and 'lincon' statements (lines 15-21) jointly specify the monotone constraints associated with the intercept parameters and the slopes of the customer satisfaction covariates. Lines 22-29 define the log likelihood for each customer which, for the i-th customer, is given by

$$\text{loglik}_i(\underset{\sim}{\alpha}, \underset{\sim}{\beta}) = \sum_{j=1}^{5} Y_{ij} \log P(Y_i = j)$$

Covariate	MLE of Slope		Covariate	MLE of Slope	
	Unconstrained	Constrained		Unconstrained	Constrained
$q79a_1$	1.92	2.01	$q82d_1$	1.27	1.27
$q79a_2$	2.09	2.01	$q82d_2$	.92	.96
$q79a_3$	1.43	1.43	$q82d_3$	.67	.73
$q79a_4$	.84	.82	$q82d_4$	.77	.73
$q79a_5$	.58	.57	$q82d_5$	.44	.42
$q79a_6$	.19	.19	$q82d_6$	.19	.19
$q79a_7$	Structural 0	Implied 0	$q82d_7$	Structural 0	Implied 0
$q82a_1$	2.68	2.56	$q82f_1$	.85	1.28
$q82a_2$	.71	.98	$q82f_2$	1.69	1.28
$q82a_3$	1.05	.98	$q82f_3$	1.08	1.07
$q82a_4$	.89	.88	$q82f_4$	.69	.69
$q82a_5$	.31	.30	$q82f_5$	.59	.58
$q82a_6$	.14	.14	$q82f_6$	.16	.16
$q82a_7$	Structural 0	Implied 0	$q82f_7$	Structural 0	Implied 0
$q82b_1$	1.31	1.28			
$q82b_2$	.81	.85			
$q82b_3$	.75	.77			
$q82b_4$	.53	.56			
$q82b_5$	.14	.21			
$q82b_6$	.23	.21			
$q82b_7$	Structural 0	Implied 0			

Table 5. Unconstrained and Constrained Slope MLEs of Customer Satisfaction Covariates

Table 5 provides a side-by-side comparison of the constrained and unconstrained MLEs for the slopes of the customer satisfaction covariates, and Figure 2 is a plot that shows the monotone behavior of the constrained estimates. There is very little difference between the unconstrained and constrained MLEs for the demographic covariates. Recall that for the unconstrained MLEs, the zero for the slope of the last level of each covariate is a structural zero resulting from the non-full rank dummy variable coding used when fitting the model.

In the case of the constrained MLEs, the slopes of the last levels of the covariates are implied zeros resulting from the full-rank dummy variable coding used when fitting the model. Table 5 shows that incorporating the constraints do not lead to a substantial change in the estimated slopes. In an indirect way, this provides a sanity check of the proposed model. We will use the constrained estimates for the remainder of the case study.

3.4 Model utility
3.4.1 NPS for individual customers
The estimated coefficients of the model can be used to predict the distribution of Y for a customer with a given set of covariates as follows. Suppose a customer is from the second Country Code within the first Region Code, has the third Job Title and is associated with the second Business Code. These demographic covariates are coded as follows:

$$RC=(1,0,0,0,0,0,0)$$
$$CC=(0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)$$
$$JT=(0,0,1,0,0,0,0,0,0,0)$$
$$BC=(0,1)$$

Suppose further that the customer gives sub-element scores for q79, q82a, q82b, q82d and q82f of 6, 6, 7, 5 and 6, respectively. These sub-element covariates are coded as follows:

$$q79=(0,0,0,0,0,1,0)$$
$$q82a=(0,0,0,0,0,1,0)$$
$$q82b=(0,0,0,0,0,0,1)$$
$$q82d=(0,0,0,0,1,0,0)$$
$$q82f=(0,0,0,0,0,1,0)$$

Combining these covariates into $\underset{\sim}{x}$, using the estimates of $\{\alpha_i\}_{i=1}^4$ and β given in Table 2, and evaluating the equations in (2) gives the probability distribution for $\tilde{Y}$ for this customer profile as:

$$Y \sim \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ .00052 & .00377 & .10437 & .43860 & .45274 \end{pmatrix}$$

from which it follows that customers with this profile $\underset{\sim}{x}$ will have an (expected) NPS of 75.2%.

3.4.2 NPS for a population of customers
Consider now the sub-population of customers in the first Region Code. For this sub-population, the relative frequencies of the three Country Codes are 81.18%, 14.12% and 4.7%, respectively. The relative frequencies of the ten job titles are 11.76%, 17.65%, 12.94%, 18.82%, 27.06%, 3.53%, 0%, 4.71%, 2.35% and 1.18%. The relative frequencies of the two Business Codes are 82.35% and 17.65%, respectively. Thus, the demographic covariate vectors for this sub-population are:

$$RC = (1,0,0,0,0,0,0)$$
$$CC = (.8118,.1412,.0047,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)$$
$$JT = (.1176,.1765,.1294,.1882,.2706,.0353,0,.0471,.0235,.0118)$$
$$BC = (.8235,.1765)\,.$$

Additionally, the distributions for the sub-element scores within the first Region Code define the sub-element covariate vectors as follows:

$$q79=(.0118,0,.0235,.0471,.3176,.2824,.3176)$$
$$q82a=(.0118,0,.0118,.0471,.2471,.4000,.2824)$$
$$q82b=(.0118,.0118,.0235,.1176,.2235,.3529,.2588)$$
$$q82d=(.0118,.0118,.0118,.0588,.2353,.3765,.2941)$$
$$q82f=(.0118,0,.0235,.0235,.1882,.4471,.3059)$$

Combining these covariates into $\underset{\sim}{x}$, using the estimates of $\{\alpha_i\}_{i=1}^4$ and $\underset{\sim}{\beta}$ given in Table 2, and evaluating the equations in (2) gives the probability distribution for Y for this population profile as:

$$Y \sim \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ .00070 & .00503 & .13452 & .47771 & .38204 \end{pmatrix}$$

from which it follows that customers with this profile $\underset{\sim}{x}$ will have an (expected) NPS of 70.1%.

3.4.3 Targeted Pathways for Improving NPS

A purely empirical way to compute NPS is to use the observed distribution (based on all 5,056 survey responses) of Y for $\underset{\sim}{p}$ in the formula $\text{NPS} = \sum_{i=1}^5 w_i p_i$, and this yields 61.7%. Consider now filling out the covariate vector $\underset{\sim}{x}$ with the sample frequencies for the observed demographic covariates and with the observed sample distributions for the sub-element covariates. Using this $\underset{\sim}{x}$ with the model yields a predicted NPS of 65.7%. The close agreement between the data-based and model-based NPS scores is additional evidence that the model fits the data well, and it also instills confidence in using the model to explore "What If?" scenarios as outlined in Figure 3. Figure 3 defines sixteen "What If?" scenarios, labels them with brief descriptions, and then shows expected NPS score if the scenario is implemented. Table 6 contains a longer description of how each scenario was implemented. Each scenario can be evaluated on the basis of how much boost it gives to the expected NPS as well as the feasibility of establishing a company program that could make the hypothetical scenario real.

We illustrated potential pathways to improve the overall NPS score, but this can also be done with specific sub-populations in mind. For example, if the first region was under study, then one could simply adjust the demographic covariates as illustrated in section 3.4.2 before implementing scenarios adjustments.
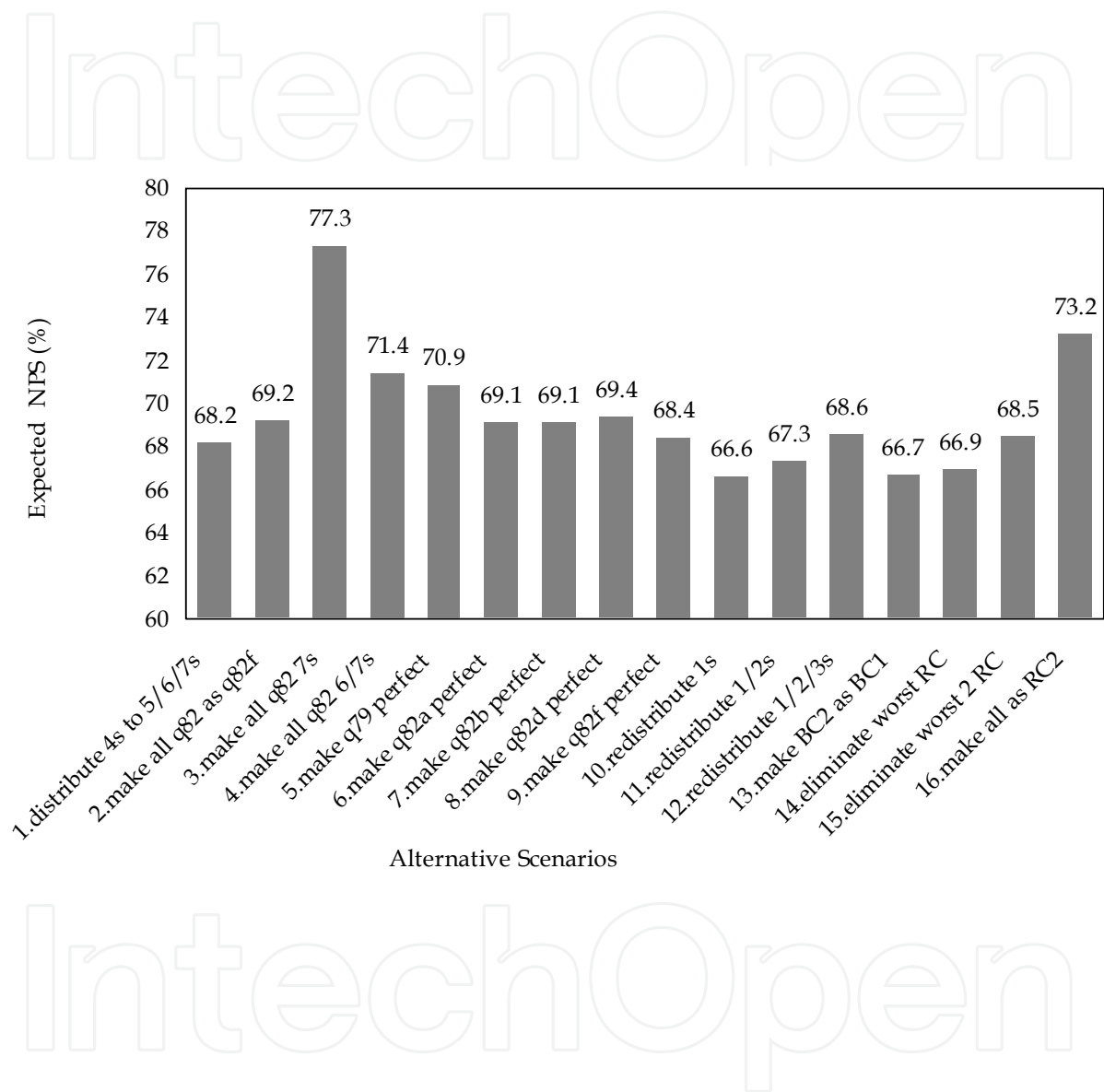
Fig. 3. Predicted NPS for Different Scenarios

Scenario	Brief Description
1	For each of q79, q82a, q82b, q82d and q82f, alter the distributions of their responses by reassigning the probability of a neutral response (4) equally to the probability of responses (5), (6) and (7)
2	Replace the response distribution for sub-elements q82a, q82b, and q82d with what was observed for q82f (which was the sub-element that had the most favorable response distribution)
3	Make the response distribution for each of q82a, q82b, q82d and q82f perfect by placing all the probability on response (7)
4	Improve the response distribution for each of q82a, q82b, q82d and q82f by placing all the probability equally on responses (6) and (7)
5	Improve the response distribution for q79 by placing all the probability on response (7)
6	Improve the response distribution for q82a by placing all the probability on response (7)
7	Improve the response distribution for q82b by placing all the probability on response (7)
8	Improve the response distribution for q82d by placing all the probability on response (7)
9	Improve the response distribution for q82f by placing all the probability on response (7)
10	Improve the response distribution for each of q79, q82a, q82b, q82d and q82f by distributing the probability of response (1) equally on responses (2)-(7)
11	Improve the response distribution for each of q79, q82a, q82b, q82d and q82f by distributing the sum of the probability of responses (1) and (2) equally on responses (3)-(7)
12	Improve the response distribution for each of q79, q82a, q82b, q82d and q82f by distributing the sum of the probability of responses (1), (2) and (3) equally on responses (4)-(7)
13	Simulate making Business Code 2 as good as Business Code 1 by setting BC=(1,0)
14	Improve the response distributions of q79, q82a, q82b, q82d, and q82f by replacing them by the average across the different Region Codes, excluding the worst Region Code
15	Improve the response distributions of q79, q82a, q82b, q82d, and q82f by replacing them by the average across the different Region Codes, excluding the two worst Region Codes
16	Improve the response distributions of q79, q82a, q82b, q82d, and q82f by replacing them all by the observed, respective, distributions for Region Code 2 (which was the region that had the most favorable response distribution)

Table 6. Implementation Detail for Each Scenario

4. Discussion

Alternative measures to NPS of customer advocacy include customer satisfaction (CSAT) and Customer Effort Score (CES) (Dixon et al., 2010). CES is measured on a 5-point scale and is intended to capture the effort required by a customer to resolve an issue through a contact-center or self-service channel. (Dixon et al., 2010) compared the predictive power of CSAT, NPS and CES on service customers' intention to do repeat business, increase their spending, and speak positively about the company. They concluded that CSAT was a relatively poor predictor, while CES was the strongest. NPS ranked in the middle.

The choice of which customer advocacy measure to use depends on many factors such as the type of company-to-customer relationship, the degree to which recommendations (for or against a company) influence a purchase decision, and whether the measures will be complemented by other customer feedback. To gain an in-depth understanding of customers' experiences and how to improve them may require multiple indicators. In the end, it is the action taken to drive improvements that customers value that is most critical.

Our case study validates the feasibility for using a multinomial logistic regression model as a means to identify key drivers of NPS, though it is clear that the same methodology could be employed with alternative measures of customer advocacy. Improvement teams at CSH have used this model to prioritize projects relative to their expected impacts on NPS. A novel aspect of our model development was the implementation of monotone constraints on the slope parameters of the ordinal covariates. Our illustrative SAS code showing how to impose the constraints on the maximum likelihood estimates should be of significant help to practitioners interested in doing the same thing.

5. Appendix A

```
1.  data indata;
2.  infile 'C:\CarestreamHealth\indata.txt';
3.  input RC CC BC JT Y q79 q82a q82b q82d q82f;
4.  run;
```

```
1.  proc logistic data=indata;
2.  class RC CC BC JT
3.      q79 q82a q82b q82d q82f/param=glm;
4.  model Y = RC CC(RC) BC JT
5.      q70 q79 q82a q82b q82d q82f;
6.  run;
```

6. Appendix B

```
1.  data indata;
2.  set indata;
3.  array cc{23} cc1-cc23; do i=1 to 23; if CC=i then cc{i}=1; else cc{i}=0;end;
4.  if BC=1 then bc1=1;else bc1=0;
5.  array jt{9} jt1-jt9; do i=1 to 9; if JT=i then jt{i}=1; else jt{i}=0;end;
6.  array q{6} q1-q6; do i=1 to 6; if q79=i then q{i}=1; else q{i}=0;end;
7.  array a{6} a1-a6;  do i=1 to 6; if q82a=i then a{i}=1; else a{i}=0;end;
```

8.   array b{6} b1-b6; do i=1 to 6; if q82b=i then b{i}=1; else b{i}=0;end;
9.   array d{6} d1-d6; do i=1 to 6; if q82d=i then d{i}=1; else d{i}=0;end;
10. array f{6} f1-f6; do i=1 to 6; if q82f=i then f{i}=1; else f{i}=0;end;
11. run;

12. proc nlp data=indata;
13. max loglik;
14. parms alp1=-7, alp2=-5, alp3=-2, alp4=-1, bcc1-bcc10=0, bcc12-bcc23=0, bbc1=0, bj1-bj9=0,   bq1-bq6=0, ba1-ba6=0, bb1-bb6=0,bd1-bd6=0, bf1-bf6=0;
15. bounds 0 <= bq6,0 <= ba6,0 <= bb6,0 <= bd6,0 <= bf6;
16. lincon 0<=alp4-alp3, 0<=alp3-alp2, 0<=alp2-alp1;
17. lincon 0<= bq5-bq6, 0<= bq4-bq5, 0<= bq3-bq4, 0<= bq2-bq3, 0<= bq1-bq2;
18. lincon 0<= ba5-ba6, 0<= ba4-ba5, 0<= ba3-ba4, 0<= ba2-ba3, 0<= ba1-ba2;
19. lincon 0<= bb5-bb6, 0<= bb4-bb5, 0<= bb3-bb4, 0<= bb2-bb3, 0<= bb1-bb2;
20. lincon 0<= bd5-bd6, 0<= bd4-bd5, 0<= bd3-bd4, 0<= bd2-bd3, 0<= bd1-bd2;
21. lincon 0<= bf5-bf6, 0<= bf4-bf5, 0<= bf3-bf4, 0<= bf2-bf3, 0<= bf1-bf2;
22. tp=cc1*bcc1+cc2*bcc2+cc3*bcc3+cc4*bcc4+cc5*bcc5+cc6*bcc6+cc7*bcc7+cc8*bcc8+cc9*bcc9+cc10*bcc10+cc12*bcc12+cc13*bcc13+cc14*bcc14+cc15*bcc15+cc16*bcc16+cc17*bcc17+cc18*bcc18+cc19*bcc19+cc20*bcc20+cc21*bcc21+cc22*bcc22+cc23*bcc23+bc1*bbc1+jt1*bj1+jt2*bj2+jt3*bj3+jt4*bj4+jt5*bj5+jt6*bj6+jt7*bj7+jt8*bj8+jt9*bj9+q1*bq1+q2*bq2+q3*bq3+q4*bq4+q5*bq5+q6*bq6+a1*ba1+a2*ba2+a3*ba3+a4*ba4+a5*ba5+a6*ba6+b1*bb1+b2*bb2+b3*bb3+b4*bb4+b5*bb5+b6*bb6+d1*bd1+d2*bd2+d3*bd3+d4*bd4+d5*bd5+d6*bd6+f1*bf1++f2*bf2+f3*bf3+f4*bf4+f5*bf5+f6*bf6;
23. pi1=exp(alp1+tp)/(1+exp(alp1+tp));pi2=exp(alp2+tp)/(1+exp(alp2+tp));
24. pi3=exp(alp3+tp)/(1+exp(alp3+tp));pi4=exp(alp4+tp)/(1+exp(alp4+tp));
25. if Y=1 then loglik=log(pi1);
26. if Y=2 then loglik=log(pi2-pi1);
27. if Y=3 then loglik=log(pi3-pi2);
28. if Y=4 then loglik=log(pi4-pi3);
29. if Y=5 then loglik=log(1-pi4);
30. run;

7. References

Allen, D. R. & Rao, T. R. N. (2000). *Analysis of Customer Satisfaction Data*, ASQ Quality Press, 978-0873894531, Milwaukee.

Dixon, M., Freeman, K. and Toman, N. (2010). *Stop Trying to Delight Your Customers*, *Harvard Business Review*, July-August, http://hbr.org/magazine.

Hayes, B. E. (2008). *Measuring Customer Satisfaction and Loyalty, 3rd edition*, ASQ Quality Press, 978-0873897433, Milwaukee.

Keiningham, T.; Cooil, B.; Andreassen, T. & Aksoy, L. (2007). A Longitudinal Examination of Net Promoter and Firm Revenue Growth, *Journal of Marketing*, Vol.71, No.3, (July 2007) 39-51, 0022-2429.

Reichheld, F. (2003). The One Number You Need to Grow, *Harvard Business Review*, Vol. 81, No.12, (December 2003) 46-54.

Vavra, T. G. (1997), *Improving Your Measurement of Customer Satisfaction: Your Guide to Creating, Conducting, Analyzing and Reporting Customer Satisfaction Measurement Programs*, ASQ Quality Press, 978-0873894050, Milwaukee.
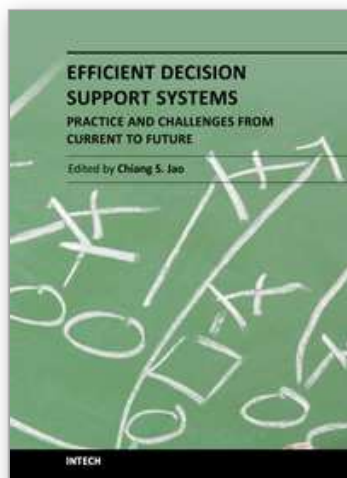
Efficient Decision Support Systems - Practice and Challenges From Current to Future

Edited by Prof. Chiang Jao

This series is directed to diverse managerial professionals who are leading the transformation of individual domains by using expert information and domain knowledge to drive decision support systems (DSSs). The series offers a broad range of subjects addressed in specific areas such as health care, business management, banking, agriculture, environmental improvement, natural resource and spatial management, aviation administration, and hybrid applications of information technology aimed to interdisciplinary issues. This book series is composed of three volumes: Volume 1 consists of general concepts and methodology of DSSs; Volume 2 consists of applications of DSSs in the biomedical domain; Volume 3 consists of hybrid applications of DSSs in multidisciplinary domains. The book is shaped upon decision support strategies in the new infrastructure that assists the readers in full use of the creative technology to manipulate input data and to transform information into useful decisions for decision makers.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Daniel R. Jeske, Terrance P. Callanan and Li Guo (2011). Identification of Key Drivers of Net Promoter Score Using a Statistical Classification Model, Efficient Decision Support Systems - Practice and Challenges From Current to Future, Prof. Chiang Jao (Ed.), ISBN: 978-953-307-326-2, InTech, Available from: http://www.intechopen.com/books/efficient-decision-support-systems-practice-and-challenges-from-current-to-future/identification-of-key-drivers-of-net-promoter-score-using-a-statistical-classification-model

INTECH
open science | open minds