Finding Hospital Patient Satisfaction Drivers Using Sagata® Regression

© 2005 Theodore T. Allen
Senior Partner
Sagata Ltd.
1580 Cardiff Road
Columbus, Ohio 43221
tallen@sagata.com

Karthik Venkataraman, M.S.
Industrial Engineer
Healthcare Management
1522 Brookforest Dr.
Columbus, OH 43204
karthikbang@yahoo.com

Executive Summary

Regression analysis is a powerful tool for interpreting patient satisfaction surveys. Such analyses can help staff prioritize improvement efforts to increase satisfaction and revenues generated by keeping patients in the system for appropriate care. This paper illustrates several possible regression applications and the special abilities of Sagata® software (www.sagata.com) to analyze the effects of categorical factors such as patient illness type and to generate informative 3D surface plots.

Table of Contents

1. Introduction

Many hospitals use lengthy surveys collected when the patients discharge to better understand the quality of their own care. The response rates for these surveys are generally good compared with many other types of surveys. Understanding the implications of the survey data can lead to new strategies that increase demand for services in part by reducing inappropriate instances when patients leave the system.

Often, surveys are analyzed by deriving the sample correlations between performance related to specific subsystems, e.g., nursing stations, and the overall score. These sample correlations are equivalent to first order regression models. In this paper, we illustrate how Sagata® Regression can be applied to achieve benefits beyond simple correlation approaches.

Specifically, two types of benefits are explored: (1) investigation of the incremental effects of factors not in the survey but under study, e.g., time until the patient sees a doctor and (2) interpret interactions or the combined effects of subsystem choices, e.g., do recent efforts by teams compliment redesign efforts to shorten the times before patients see medical experts.

In Section 2, the application of regression modeling and Sagata® Regression Professional software to analyze customer satisfaction survey data is described. Section 3 includes an example of how survey data can be supplemented with additional data to help staffers better evaluate the effects of possible system improvements. In Section 4, a more in-depth analysis of the previous data is given that accounts for the combined effects of factors on satisfaction. Section 5 summarizes possible roles for regression in satisfaction analysis.

2. Basic Analysis Using Regression

Table 1 shows hypothetical data from questions answered by 25 patients. This data is similar to real data at a local hospital but changed partly to protect patients' and hospitals' confidentiality. Many large hospitals generate thousands of similar rows of data every month. Statements made from analysis of this data could, of course, only be relevant to decisions pertinent to patients similar to the 25. Yet, regression analysis and software such as Sagata® Regression can handle thousands of patients or data points.

A natural first step in modeling is to start with a "1st Order" and "Least Squares" regression model. *Figure 1(a)* shows the dialog used in applying Sagata® software. The subsystem related rankings ("X-Factors" or inputs) are in the excel spreadsheet cells D1:L26 and the overall ratings ("Y-Response" or outputs) are in cells M1:M26. The fitted model including the residual plot in *Figure 1(b)* and the coefficient table in *Table 2* is outputted to a new "Simple Model" worksheet.

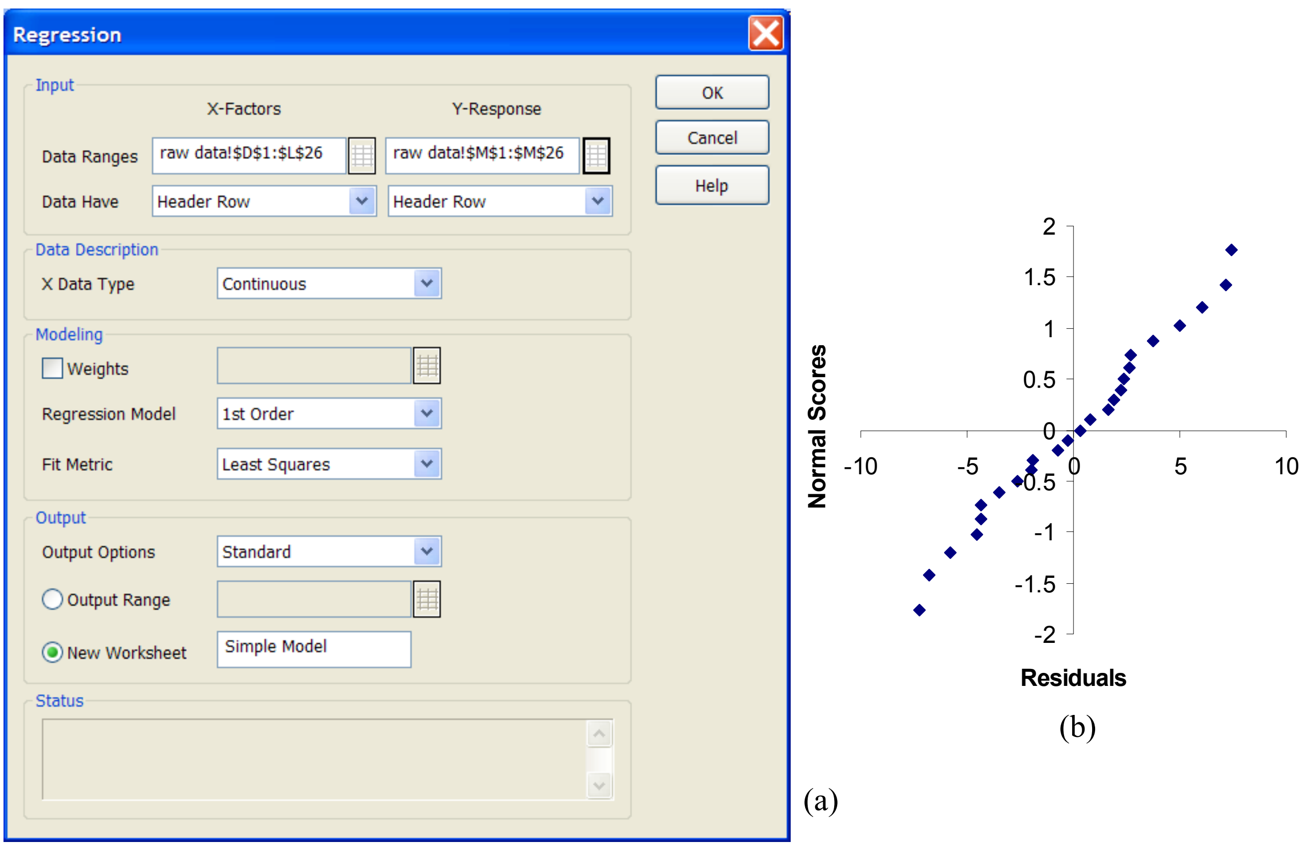


Figure 1. (a) Sagata® main dialog and (b) normal probability plot of residuals.
Table 1. Hypothetical survey data from 25 patients.

Patient #	Admissions	Meals	Rooms	Discharge	Physicians	Nurses	Visitors	Tests & Treatments	Personal	Overall
3002051	N22	22	80	78	86	91	85	93	74	74
3002052	O49	45	71	71	75	80	81	74	82	82
3002053	N1	12	83	77	91	84	84	82	94	89
3002054	N1	25	89	76	87	89	94	84	82	88
3002055	O2	55	64	80	77	83	79	85	83	95
3002056	N22	92	48	92	78	81	88	87	89	80
3002057	O49	12	88	81	80	83	85	87	85	80
3002058	N22	15	84	89	89	81	75	90	78	72
3002059	I1	28	91	92	91	85	85	87	84	79
3002060	N22	17	85	88	86	79	91	83	80	74
3002061	N22	22	82	89	87	81	91	76	87	84
3002062	O39	8	82	89	89	87	88	93	86	87
3002063	O39	3	92	85	90	84	81	90	82	87
3002064	N22	61	67	78	94	82	85	83	99	84
3002065	N1	22	82	87	83	78	91	83	78	88
3002066	O2	45	77	81	88	87	72	71	87	92
3002067	I1	15	90	93	86	85	94	84	81	89
3002068	I1	21	84	92	84	80	77	72	85	86
3002069	N22	39	78	92	85	79	78	85	89	72
3002070	O39	2	85	83	84	82	80	88	85	91
3002071	N22	39	78	86	84	78	86	85	88	81
3002072	O49	11	85	89	94	78	78	87	90	80
3002073	I1	31	80	87	80	80	87	85	82	85
3002074	N1	15	89	93	81	84	85	94	87	93
3002075	N22	22	78	75	90	89	76	91	84	73

Table 2. Coefficient estimates table from Sagata® software output.

Term	*Coefficients*	*Standard Error*	*t Stat*	*p-value*	*Lower 95%*	*Upper 95%*	*VIF*
Constant	-26.83	48.14	-0.56	0.586	-129.45	75.78	
Admissions	1.02	0.16	6.49	0.000	0.68	1.35	1.92
Meals	-0.10	0.21	-0.47	0.645	-0.55	0.35	1.58
Rooms	-0.16	0.33	-0.48	0.639	-0.86	0.55	2.32
Discharge	-0.64	0.41	-1.57	0.138	-1.50	0.23	1.86
Physicians	0.04	0.20	0.21	0.837	-0.39	0.48	1.24
Nurses	0.60	0.21	2.84	0.012	0.15	1.06	1.43
Visitors	0.04	0.31	0.13	0.901	-0.61	0.69	2.13
Tests & Treatments	0.31	0.20	1.54	0.144	-0.12	0.73	1.51
Personal	0.17	0.21	0.78	0.450	-0.29	0.62	1.34

Two advantages of the simple, first order model compared with sample correlations are the following. First, normal probability plots and adjusted R-squared values of residuals such as *Figure 1(b)* provide an indication of how trustworthy the results are. The importance of normal looking residuals in establishing evidence is described in standard regression textbooks such as Montgomery, Peck, and Vining (2001). The slightly nonlinear pattern in *Figure 1(b)* suggest that a simple linear relationship is might not be considered trustworthy.

The R-squared adjusted value of 0.78 for the simple model suggests that much (78%) of the variation in the patient overall satisfaction can be explained by the simple model. In other words, a weighted sum of individual, subsystem satisfaction scores can roughly predict overall satisfaction.

A second advantage of regression is that plots of regression predictions such as in *Figure 2* can give insight into potential causal relationships. For example, *Figure 2* seems to suggest that the success or failure of the admissions process to the hospital emergency room drives overall satisfaction. The high slope associated with satisfaction with nurses (and the high coefficient and low p-value in *Table 2* for the nurse term) suggest that nurses might be having a strong effect on overall satisfaction.

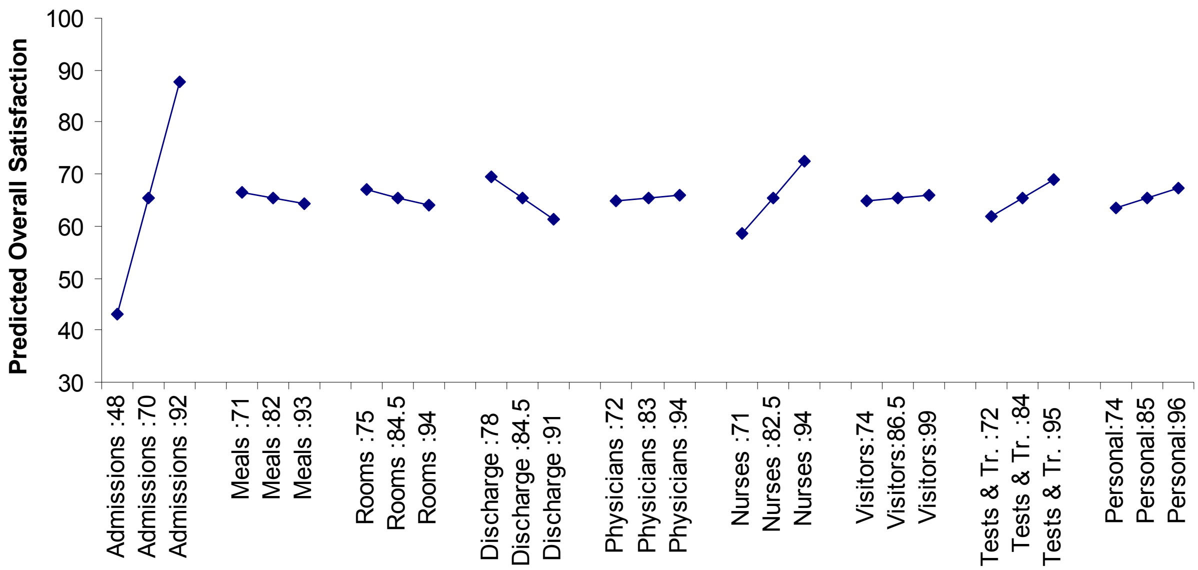


Figure 2. Simple, first order regression model predictions of overall satisfaction.

3. Supplementing with Data From Observational Studies

Many reasons can motivate staff to gather data on a manual, patient-by-patient basis. First, it might be of interest to evaluate the effects of possible changes. For example, it might be of interest to study possible effects from dramatically reducing the time between when patients arrive in emergency rooms and when they meet a doctor. Second, it might be a helpful way to generate a model with more predictive power, e.g., with higher adjusted R-squared and/or straighter lines on a normal probability plot of residuals.

Table 3 shows hypothetical data collected manually by staff to supplement survey data. The "Diagnosis" column describes the type of ailment that motivated the patient to come to the hospital as judged by the primary physician for that paper. Note that these ailments are generally categorically different. The "Physician Time" is the time between the arrival of the patients and their first meeting with a doctor.

The additional data plus regression model can help staff better understand what is driving overall satisfaction. In particular, the resulting analysis could help establish the focus on possible improvement project or policy. In the next section, a relatively in-depth analysis of the entire data taken together using Sagata® Regression software is presented.

Table 3. Hypothetical supplemental data for the same patients in the survey.

Patient #	Diagnosis	Physician Time
3002051	N22	22
3002052	O49	45
3002053	N1	12
3002054	N1	25
3002055	O2	55
3002056	N22	92
3002057	O49	12
3002058	N22	15
3002059	I1	28
3002060	N22	17
3002061	N22	8
3002062	O39	8
3002063	O39	3

Patient #	Diagnosis	Physician Time
3002064	N22	61
3002065	N1	22
3002066	O2	45
3002067	I1	15
3002068	I1	21
3002069	N22	39
3002070	O39	2
3002071	N22	39
3002072	O49	11
3002073	I1	31
3002074	N1	15
3002075	N22	22

4. In-Depth Analysis Using Categorical Factors and Interactions

In this section, a more thorough analysis of the entire data is considered. This analysis includes the combined effects factors on satisfaction. For example, it can answer questions such as whether prompt attention from either doctors or nurses is sufficient for keeping the average patient satisfied.

Sagata® software permits the analysis of X-factors that are either quantitative or categorical (e.g., with qualitatively different settings such as ailment types). Also, the software permits convenient evaluation of many possible causal relationships using watch statistics such as Adjusted R-squared. Sagata® software includes both stepwise regression and "MinPRESS" auto modeling. The latter often suggests models that are relatively trustworthy because they cross-validate relatively well. See the Sagata® help facility and www.sagata.com for additional details.

Figure 3 shows the Regression Model dialog used to generate the relatively desirable fitted model. The first step used was to change the number in the upper left of the dialog to permit the auto modeling feature to consider terms up to second order. Next, the auto modeling button was pressed generating many of the terms on the selected terms list in the interactive dialog. Finally, manual additions and subtractions of terms were entertained to help make the model more intuitively believable. In this case, no changes resulted from this step.

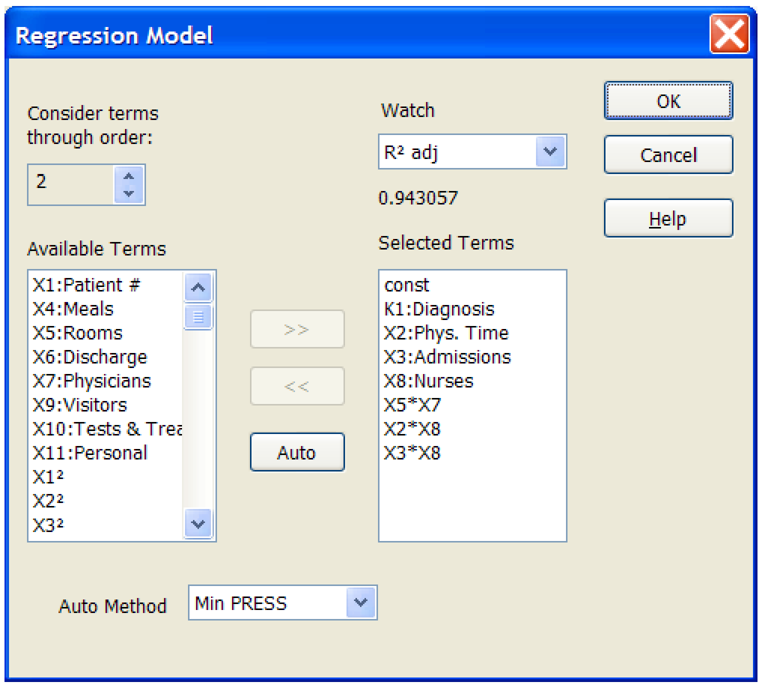


Figure 3. The regression model dialog used to generate the model used for analysis.

The predictions of the derived regression model are shown in *Figure 4*. These predictions focus solely on the effects of individual factors when changed singly. The results would indicate a different average satisfaction level associated with patients having different ailments which was modulated further by different experiences and perceptions of the various hospital subsystems. The associated prediction model can be written:

Predicted Overall Rating = 456.23 + 5.71 × Diagnosis_O49 + 8.29 × Diagnosis_N1 – 1.19 × Diagnosis_O2 + 3.31 × Diagnosis_I1 + 13.47 × Diagnosis_O39 – 4.63 × Phys. Time – 3.24 × Admissions – 4.87 × Nurses – 0.0017 × Rooms × Physicians + 0.0531 × Phys. Time × Nurses + 0.0452 × Admissions

Where Diagnosis_O49 is a 1 if the patient type in question has ailment O49 and zero otherwise and similarly for Diagnosis_N1,…Diagnosis_O39. If there were more data, it might be possible to study the interactive effects of patient ailment types and hospital subsystems. Such as study might indicated that additional policies for addressing different ailments might be warranted.

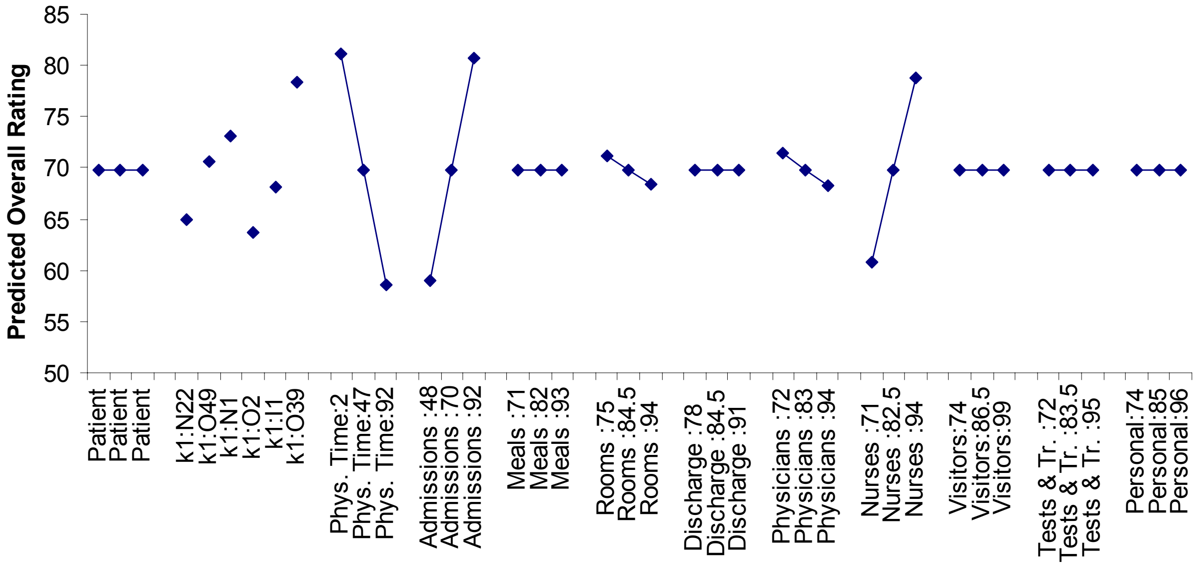
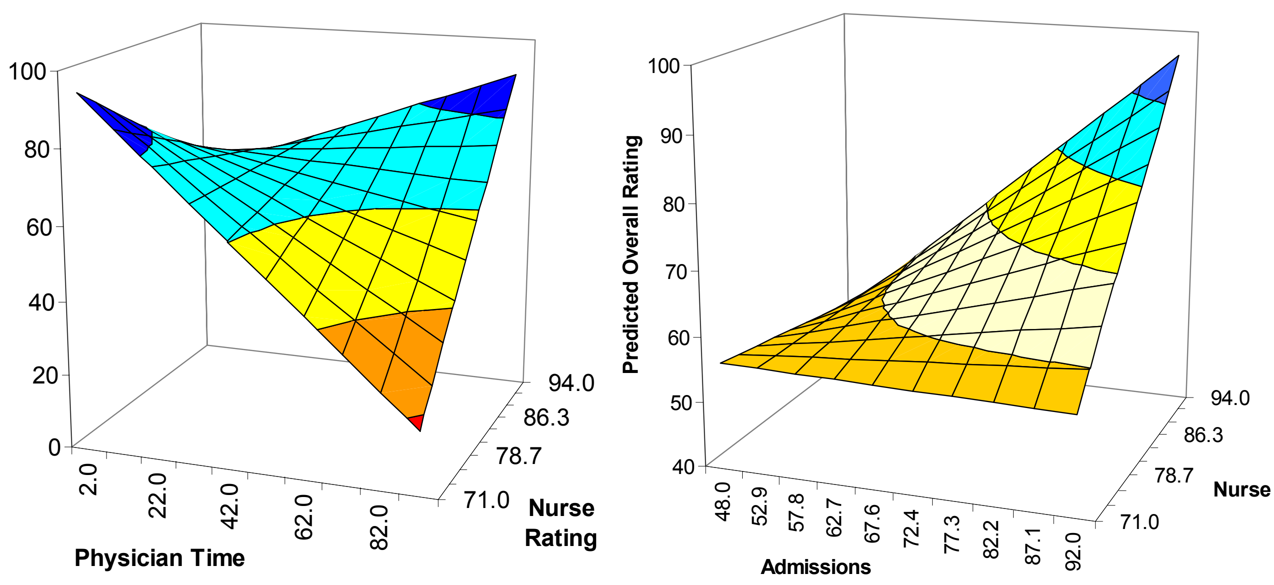


Figure 4. Relatively in-depth regression model predictions of overall satisfaction.

Figure 5(a) shows the interactive effects of doctor speed and nurse efforts on patient overall satisfaction. The results indicate that, at least for the 25 patients in the study, excellent nurse care can substitute for prompt physician attention in the attainment of excellent ratings. For fixed promptness of doctors, only an excellent admissions process combined with excellent attention from nurses can result in high satisfaction averages.



(a) (b)

Figure 5. 3D surface plots of model predictions as a function of (a) phys. time and nurse ratings and (b) admissions and nurse ratings.

5. Conclusions and Other Possible Roles For Regression

In the hypothetical example considered in this paper, the regression model developed would likely provide valuable information for hospital improvement teams. Often, these improvement teams are lead by so-called "six sigma black-belts" and experts in lean production (e.g., see Allen, 2005). These leaders could use the analysis described to analyze system design alternatives and to help determine whether specific strategies are appropriate to increase satisfaction and reduce patients leaving the system at inappropriate times.

A somewhat similar analysis was performed at a real hospital. The results lead an improvement team to restructure the arrival process in the emergency room. The redesigned process put medical experts (doctors and nurses) in contact with patients much more quickly after their arrival. The resulting increase to customer satisfaction was confirmed to dramatically reduce the "elopements" of patients (inappropriate departures) and to generate an estimated $2M in annual revenues.

It is widely believed that regression analysis can provide useful insights in many similar improvement projects. In applying regression, Sagata® software can help by providing key abilities like (1) analyzing the effects of categorical factors (e.g., ailment type) and (2) producing 3D surface plots to visualize the interactive effects of factors on important outcomes like overall patient satisfaction.

Acknowledgments

Jamie Bukovac and Beth West deserves immense credit for exposing us to important applications of regression in health care and for many useful insights. Also, we thank Emily Patterson and Mike Bernshteyn from Sagata Ltd. for many useful discussions and editing.

References

Allen, T. T. (2005), Introduction to Six Sigma For Scientists and Engineers, Springer Verlag.

Montgomery, D. C., E. A. Peck, and G. G. Vining (2001), *Introduction to Linear Regression Analysis*, Student Solutions Manual, Wiley.