Outcome Variables and their Analysis in Survey Research

David W. Gerbing

In a large Dallas, TX mall there are two Mexican restaurants, one at either end of the mall: SFG or JSC. Management at the SFG wants to better understand their customers and compare them to the customers of the primary competitor. A market survey was conducted with customers interviewed at the mall.

405 customers of either one of the two restaurants responded to the survey:

    http://web.pdx.edu/~gerbing/data/SFG.csv

Below are some of the items, each item preceded by its name in the data table. These items can be organized into four classes (the full survey contains more).

Survey Structure

The *screening* items identify who is eligible to participate in the study and their classification.

```
---------
Screening
---------
```

x_s4: Which of the following restaurants have you eaten at most recently?

SFG
JSC

other: dropped from the survey

The *demographics* address characteristics of the respondent.

```
------------
Demographics
------------
```

x32. Gender: Male, Female

x34. Age (years)

x35. Household annual income

The *product attributes* in this situation are the characteristics of the restaurant.

```
------------------
Product Attributes
------------------

Strongly                        Strongly
Disagree                          Agree
    1    2    3    4    5    6    7

x12. has friendly employees

x13. is a fun place to eat

x14. has large size portions

x15. has fresh food

x16. has reasonable prices

x17. has an attractive interior

x18. has excellent food taste

x19. has knowledgeable employees

x20. serves food at the proper temperature

x21. has quick service
```

The *outcome* variables reflect the overall level of satisfaction the customer experienced with the restaurant.

```
-----------------
Outcome Variables
-----------------


x22. How satisfied are you with ____?

Not Satisfied                                Very
  At All                                  Satisfied
     1     2     3     4     5     6     7


x23. How likely are you to return to ____ in the future?

Definitely Will              Definitely Will
 Not Return                         Return
     1     2     3     4     5     6     7


x24. How likely are you to recommend ____ to a friend?

Definitely Will              Definitely Will
 Not Recommend                   Recommend

     1     2     3     4     5     6     7
```

One goal of such a project is to understand the level and variability of customer satisfaction. Are customers generally satisfied with our restaurant? How satisfied are our customers compared to the competition?

> **Primary goal of the analysis**: Explain the level of customer satisfaction in terms of the product attributes.

Use the customer evaluation of the product attributes to explain the overall level of customer satisfaction.

Data Analysis

The primary statistical technique to address the "why" of customer satisfaction is multiple regression. Take the outcome variable of interest and regress it on the product attributes, which are the predictor variables in the regression model. That is, the outcome variable is the response variable, generically referred to as Y. The product attributes are the predictor variables, each generically referred to as X_i, the i$^{\text{th}}$ predictor variable.

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \ldots$$

The goal of the regression analysis is to identify those products that meaningfully contribute to overall satisfaction with the product. First identify those attributes that are statistically related to the response variable. For these products identify those that have the largest slope coefficients, positive or negative.

Each slope coefficient, b_i, is the average change in the outcome response variable for a 1-unit increase in the corresponding product attribute predictor variable, holding all the other predictor variables constant. In this example all the product attributes are measured on the same 7-point scale, so the size of the coefficients can be directly compared to each other in addition to their respective p-values.