A strategy to evaluate the satisfaction of different passenger groups

Michele Gallo and Davide Fiorenzo

Abstract The Measure of the passenger satisfaction presents more difficulties like the measure of human performance, attitude. In this work a strategy based on Rasch analysis and Analysis of Means (ANOM) is proposed. This is based on the idea that the Rasch rating scale model gives 'sufficient statistic' for an underlying unidimensional latent trait as the satisfaction generated by local transport operators. Furthermore, the ability of passengers, measured by the rating scale model, are studied by ANOM decision charts to verify if different groups have different level of satisfaction.

1 Introduction

Today people are more mobile, and expect higher-level performance and quality of service in public transportation. To meet the increasing mobility demand, the public transport companies have to tailor their services supply to the wants and needs of their actualor potential customers. An important source of information for quality assurance is the customer satisfaction survey, where the customer satisfaction in the public transport sector is subject to different conditions than in other sectors. In fact, satisfaction is not the only factor influencing the users' behavior but also a range of other factors, such as the accessibility to a certain model in a certain situation. Moreover when local transport is considered to be the freedom of an individual to choose between different means of transportation (public or private), it is preserved, and the customer satisfaction becomes a vital concern for companies and organizations in their efforts to improve service quality, and maintainance of passenger's loyalty.

Michele Gallo
Department of Social Science, University of Naples "L'Orientale", Italy e-mail: mgallo@unior.it
Davide Fiorenzo
e-mail: davide.fiorenzo@gmail.com

In the last three decades, more conceptual customer satisfaction models have been proposed in literature. Customer satisfaction is a result of a latent complex information process summarized in a multiple-items questionnaire in which one set of alternative responses are used for estimating probabilities of responses. For this reason the analysis of multi-item data should be considered the multidimensional nature of customer satisfaction. However, the multi-item scale needs to measure only customer's satisfaction. No more attributes or behavior need to be measured together but only one latent variable.

More latent trait models could be used to measure customer's satisfaction, but Rasch models are distinguished from others by a fundamental statistical characteristic - subject's sum score is a 'sufficient statistic' for the underlying unidimensional latent trait (Masters, 1984). The model is based on the simple idea that passengers who have a high total score on an item are more satisfied overall than passengers with low scores. Likewise, items that receive lower ratings are more difficult to endorse than items that receive higher ratings. This way, on a single continuum of interest, it is possible to clearly identify which items are more difficult to generate satisfaction and which passengers are more satisfied than others.

Generally a customer satisfaction questionnaire has additional items regarding the characteristics of passengers. These items could be used to identify the different levels of satisfaction within various groups of passengers. In this case a graphical procedure like the Analysis of Means (ANOM; Ott, 1967) is a good choice in order to understand if there are different levels of satisfaction between different kinds of passengers. The purpose of this work is to use Rasch measurement theory to examine: (1) whether the customer satisfaction model, used for measuring the passenger's satisfaction meets a valid set of measures, and (2) whether ANOM procedure gives additional information on the satisfaction level of the different groups of passengers.

2 Theory

2.1 Rating scale model

When all items present the same set of alternatives, it seems reasonable to expect that the relative difficulties of the steps between categories should not vary from item to item. For these kinds of questionnaires the rating scale (Wright *et al.*, 1982) is the more appropriate version of Rasch models.

Rating scale model - within a probabilistic framework - converts ordinal raw-score data, such as the scale strong dissatisfaction / dissatisfaction / satisfaction / strong satisfaction, into an interval-based measure, the log-odd metric or logit. Let $P_{ij(m)}$ be passenger i's probability of scoring m on item j, the rating scale model can be written:

$$P_{ij(m)} = \frac{\exp(\xi_i - \delta_j - \gamma_m)}{[1 + \exp(\xi_i - \delta_j - \gamma_m]} \tag{1}$$

where δ_j is the difficulty for item j to generate satisfaction, ξ_i is the attitude of ith the passenger to be satisfied, and γ_m is the threshold parameter associated with the transition between response categories *m-1* to *m*.

To estimate these parameters the "Joint Maximum Likelihood Estimation" algorithm is used in this work (Wright *et al.*, 1969). This method is more flexible and it is independent from specific passenger and item distributional forms. Moreover the logits measure $\ln(P_{ij(m)}/P_{ij(m-1)})$, of the items, passengers and rating scale categories, convert ordinal raw scores into linear interval measures.

When the diagnostic analysis assures that the measures of passenger's attitude are valid and reliable, they can be employed in a model than needs linear and normal distributed data like ANOM.

2.2 Analysis of Means

The ANOM is a technique proposed for analyzing experimental or non experimental data related to normally, binomial and Poisson distributed data (Nelson *et al.*, 2005). ANOM is mostly useful when the desired outcome is to identify differences between groups or treatments and, in case of observational data, when a different number of observations are generally given for each group. Let n_k be the number of observations into group k $(k = 1,...,K)$ with μ_k the mean for kth group, the hypothesis to test is $\mathrm{H}_0 : \mu_1 = ... = \mu_k = ... = \mu_K$ versus the alternative one is different. Similar to the analysis of variance, ANOM tests whether there are differences among the groups, but dissimilar to analysis of variance, when there are differences, it also indicates how groups differ by a decision chart.

If data is least approximately normally distributed and all the different groups have the same variance for obtaining upper and lower decision lines, the sample means $\bar{y}_k = \sum_{i=1}^{n_k} y_{ik}/n_k$ and sample variances $\mathrm{s}_k^2 = \sum_{i=1}^{n_k} (y_{ik} - \bar{y}_k)^2/(n_k - 1)$ of each group can be used. Nelson (1989) shows that for studies with unequal samples the decision lines are

$$\bar{y}_k \pm cv(\alpha, K, n)\sqrt{MS_e}\sqrt{\frac{n - n_k}{n * n_k}} \tag{2}$$

with $\bar{y} = \sum_{k=1}^{K} \bar{y}_k/K$ is the overall mean, $\mathrm{n}=\sum_{i=1}^{K} n_k$ is the number of observations, $\mathrm{MS}_e = \sum_{k=1}^{K} \bar{y}_k/K$ is the mean square error, and $cv(\alpha, K, n)$ is a critical value that depends on the level of significance desired α, the number of groups K, the degree of freedom for MS_e.

When the sample means for each groups are plotted between the decision lines given by (2) then there are not differences between groups on the level of significance α.

3 Application and conclusion

To show the goodness of the strategy proposed, a data set with 2.473 interviews are used. The passenger satisfaction is measured by 10 items ('station cleanness', 'train cleanness', 'passenger comfort', 'regularity of service', 'frequency of service', 'staff behavior', 'passenger information', 'safety', 'personal and financial security', 'escalators and elevators working') where each item has four levels (Likert scale). While others items (gender, reason of travel, kind of ticket, ...) are used to study the different level of satisfaction between the groups of passengers.

The results of analysis will be given during the presentation.

To analyze this situation, in literature is proposed the Many-Facet model (Linacre, 1989). It preserves the Rasch framework but one or more extra components of the measurement are introduced. An example is to consider the stations generally used from passengers. In this case the equation (1) could be written:

$$P_{ijh(m)} = \frac{\exp(\xi_i - \delta_j - \tau_h - \gamma_m)}{[1 + \exp(\xi_i - \delta_j - \tau_h - \gamma_m]} \tag{3}$$

where δ_j, ξ_i, and γ_m are been defined in section 2.1, τ_h is the difficulty for hth station to generate satisfaction, and $P_{ijh(m)}$ is the passenger i's probability of scoring m on item jth, when use the station hth.

Unfortunately the Many-Facet model does not satisfy the requirement of 'statistical sufficiency' and 'separability of parameters' because adding other parameters to estimates might improve the fit of the model, but in this case, the measure generated would be deteriorated into a qualitative description of data set observed and hardly replicable in future samples.

References

1. Linacre, J.M.: Many-facet Rasch measurement. MESA Press, Chicago (1989)
2. Masters, G.N.: Constructing an item bank using Partial Credit scoring. Journal of Educational Measurement **21 (1)**, 19–32 (1984)
3. Nelson, P.R.: Multiple Comparisons of Means Using Simultaneous Confidence Intervals. Journal of Quality Technology **21**, 232–241 (1989)
4. Nelson, P.R., Wludyka, P.S., Copeland, K.A.F.: The Analysis of Means: A Graphical Method for Comparing Means, Rates, And Proportions. Society for Industrial Applied, Philadelphia (2005)
5. Ott, E.R.: Analysis of Means - A Graphics Procedure. Industrial Quality Control **24**, 101–109 (1967)
6. Wright, B.D., Panchapakesan, N.: A procedure for sample-free item analysis. Educational and Psychological Measurement **29**, 23–48 (1969)
7. Wright, B.D., Masters, G.N.: Rating Scale Analysis: Rasch Measurement. MESA Press, Chicago (1982)