JOURNAL OF APPLIED QUANTITATIVE METHODS

ISSN 1842-4562
Member of DOAJ

Journal Home > Volume 6, Issue 2, June 30, 2011

Measuring Passenger Satisfaction: A Strategy Based on Rasch Analysis and the ANOM

Pdf Version

Michele GALLO

Keywords

customer satisfaction, rating scale model, analysis of means, metro system

Abstract

Measuring passenger satisfaction presents several difficulties since customer satisfaction in the public transport sector is subject to different conditions which are different than those that affect other sectors. In this work, a strategy based on Rasch analysis and the Analysis of Means (ANOM) is proposed. This study is based on the idea that the Rasch rating scale model gives 'sufficient statistic' for an underlying unidimensional latent trait such as the satisfaction generated by local transport operators. Furthermore, the ability of passengers, measured by the rating scale model, is studied by means of ANOM decision charts to verify if there are different levels of satisfaction between the different groups of passengers.

(top)
(back to issue)

MEASURING PASSENGER SATISFACTION: A STRATEGY BASED ON RASCH ANALYSIS AND THE ANOM

Michele GALLO
Department of Human and Social Science,
University of Naples L'Orientale, Naples, Italy

E-mail: mgallo@unior.it

Abstract: *Measuring passenger satisfaction presents several difficulties since customer satisfaction in the public transport sector is subject to different conditions which are different than those that affect other sectors. In this work, a strategy based on Rasch analysis and the Analysis of Means (ANOM) is proposed. This study is based on the idea that the Rasch rating scale model gives 'sufficient statistic' for an underlying unidimensional latent trait such as the satisfaction generated by local transport operators. Furthermore, the ability of passengers, measured by the rating scale model, is studied by means of ANOM decision charts to verify if there are different levels of satisfaction between the different groups of passengers.*

Key words: *customer satisfaction; rating scale model; analysis of means; metro system*

1. Introduction

Nowadays, people are more mobile and expect efficient, high quality public transportation services. In order to meet the increasing mobility demand, public transport companies have to tailor their services they supply to the wants and needs of their current or potential customers. An important source of information on quality assurance is the customer satisfaction survey, where customer satisfaction in the public transport sector is subject to conditions different than those concerning other sectors. In fact, satisfaction is not the only factor influencing the users' behavior since it is also influenced by 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 from different means of transportation (public or private), it is presumed, and the customer satisfaction becomes a vital concern for companies and organizations in their efforts to improve service quality, and retain the passenger's loyalty.

In the last three decades, more conceptual customer satisfaction models have been proposed in statistical 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 as the multidimensional nature of customer satisfaction and the different nature of the data (Gallo, 2007). However, the multi-item scale needs to measure only customer satisfaction. No more attributes or behavior need to be measured together but only one latent variable.

Many latent trait models could be used to measure customer satisfaction, but the Rasch models are distinguished from others by a fundamental statistical characteristic, viz., the subject's sum score is a 'sufficient statistic' for the underlying unidimensional latent trait (Wright and Linacre, 1989). The model is based on the simple idea that the passengers who have a high total score on an item are more satisfied than those 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 the items which are more difficult to generate satisfaction and the passengers who are more satisfied than others.

Generally, a customer satisfaction survey should be designed to collect the data in a less intrusive and idiosyncratic way, as much as possible. In a public transport sector a good way to submit a customer satisfaction questionnaire is on the platform or on the train. This is only possible when the questionnaire has not many items to measure customer satisfaction and a few additional items regarding the characteristics of passengers are given. These latter items could be used to identify the different levels of satisfaction within various groups of passengers.

To 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') are used to measure the passenger satisfaction, where each item has four different levels (Likert scale). Other items (sex, age, profession, purpose of travel, day of interview, number of travel frequency in a week, intermodal transportation service used) would give additional information on the passengers.

The purpose of this work is to determine whether the questionnaire used is adequate to give a measure along the continuum of the underlying passenger satisfaction. Therefore, rating scale model is applied to improve the measurement tool. When a valid measure of passenger satisfaction is given, a graphical procedure like the Analysis of Means (ANOM) is used in order to understand the different levels of satisfaction between 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 will not vary from item to item. For these kinds of questionnaires the rating scale (Andrich, 1978; Wright and Masters, 1982) is the more appropriate version of the Rasch models.

Rating scale model - within a probabilistic framework - converts ordinal raw-score data 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 as:

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

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

The logits measures are given by $\ln\left(P_{ij(m)} / \left(1-P_{ij(m)}\right)\right)$. For passenger, the logit indicates whether one passenger is more able than another to get satisfaction. For item, logit measures indicate whether one item is more difficult than another to generate satisfaction. And for rating scale categories, logit measures indicate whether one rating scale category is greater or less than another in degree (for example: does the 'satisfaction' category represents less satisfaction than the 'strong satisfaction' category).

This method is more flexible and it is independent from specific passenger and item distributional forms. Moreover the logit measure $\ln\left(P_{ij(m)} / \left(1-P_{ij(m)}\right)\right)$, 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 satisfaction are valid and reliable, they can be employed in a model that needs linear and normal distributed data like ANOM.

2.2. Analysis of Means

The phrase "analysis of means" was used for the first time by Ott (1967). And based on Bonferroni inequalities, he proposed ANOM as a multiple comparison procedure that could be used instead of, or as a follow up to, analysis of variance (ANOVA). However, after 1982 exact critical value for the main effects of ANOM in balanced designs were obtained (Nelson, 1982). Nowadays ANOM is proposed in many cases for experimental or non experimental data related to normally, binomial and Poisson distributed data (Nelson et al., 2005). In this paper, ANOM is useful when the desired outcome is to identify differences between groups and, in case of observational data, when a different number of observations is generally given for each group (one-factor unbalanced ANOM).

Let n_k be the number of observations into group *k* (*k* = 1,…,*K*) with μ_k being the mean for a *k*th group, the hypothesis to test is $H_0 : \mu_1 = \ldots = \mu_k = \ldots = \mu_K$ versus the alternative one that is different. Similar to the ANOVA, 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

$$\bar{y}_k = \frac{1}{n_k}\sum_{i=1}^{n_k} y_{ik}$$

and sample variances $s_k^2 = \frac{1}{n_k - 1}\sum_{i=1}^{n_k}\left(y_{ik} - \bar{y}_k\right)^2$ of each group can be used.

ANOM procedures for studies with unequal samples require to consider the decision lines

$$\bar{y} \pm cv\left(\alpha, K, n-K\right)\sqrt{MS_e}\sqrt{\frac{n-n_k}{nn_k}} \qquad (2)$$

where $\bar{y} = \frac{1}{K}\sum_{k=1}^{K}\bar{y}_k$ is the overall mean, $MS_e = \frac{1}{K}\sum_{k=1}^{K}s_k^2$ is the mean square error, and $cv(\alpha, K, n-K)$ 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. Full theory behind ANOM (as multivariate negatively correlated singular *t* distribution, power curve etc.) was showed by Nelson (1985).

3. Application

MetroNapoli S.p.A., which manages all of the rail transport in the city of Naples, conducted a survey analysis on 2,473 passengers, according to stratified random sampling, to measure the passenger satisfaction of medium sized metro systems. The questionnaires were submitted by 10 different interviewers in the second week of October. Each item had four ordinal scales viz., 'strong dissatisfaction' / 'dissatisfaction' / 'satisfaction' / 'strong satisfaction'.

The analysis consisted of two parts. Firstly, as stated above, Rasch analysis focuses on the psychometric properties of the items, passengers, and rating scale categories. The WINSTEPS program (Linacre and Wright, 2000) was used in order to obtain Rasch measurements from these data. Secondly, with the goal of investigating whether the passenger's satisfaction was influenced by personal characteristics (profession, age, sex, purpose of travel), the logit measure of the passenger satisfaction were used into ANOM analysis. Moreover, other aspects of service (the day on which interview took place, weekly travel frequency, intermodal transportation service) were investigated. The ANOM was done by a simple routine developed in R software, but it is included as a standard option in more statistical software (included SPSS, SAS and MINITAB).

3.1. Rating scale model results

The WINSTEPS computer program was used to perform the rating scale model and the compatibility of the raw data with the Rasch measurement model, which was verified by more fit statistics. A statistics summary of the WINSTEPS program shows the reliability and separation for items and passengers. In this case, the reliability index observed is 0.99 for items and 0.78 for passengers, where the values range between 0 and 1. The estimates for items show the replicability placement of items across other passengers measuring the same construct index. Separation index, that is alternative to reliability, estimates the spread of items (15.15) and the spread of passengers (1.89) on the underlying latent trait.

The results for the rating scale analysis of the passenger satisfaction are shown in Figure 1. The vertical line represents the variable passenger satisfaction into a log-odds scale. Passengers are aligned to the left and represented by "#". The more satisfied are at the top. Items are aligned to the right, and the more difficult to generate satisfaction are at the top. It is verified that the distribution of passenger is normal and displayed in a higher position than item distribution. Therefore, passengers have more probability to get satisfaction from MetroNapolis' service.

```
          Persons - MAP - Items
               <more>|<rare>
 5          .######  +
                     |
                     |
                     |
                     |
                 .#  |
 4                   +
                     |
                     |
                     |
                .##  |
                  .  |
 3                . T+
                .##  |
                     |
                  .  |
               .###  |
                  .  |
 2             .###  +
                    S|
              .####  |
                     |
              .####  |
              .####  |
 1                .  +
      .############  |T
            .###### M|   Passenger information
                  .  |   Staff behavior
           .#######  |S
           .#######  |   Passenger comfort           Train cleanness
 0          .######  +M  Personal and financial sec  Station cleanness
              .####  |
                     |S  Frequency of service        Regularity of service
               .### S|
                .##  |   Security
                 ##  |T
-1               .#  +
                 .#  |
                     |
                  .  |
                  . T|
                     |
-2                .  +
                  .  |
                     |
                  .  |
                     |
                     |
-3                .  +
                     |
                     |
                     |
                     |
                     |
-4                .  +
               <less>|<frequ>
```

Figure 1. Person-item map for passenger satisfaction

Note: Each "#" is 27 passengers, which are aligned to the left of the corresponding log-odds measure of satisfaction. The items are aligned to the right of the corresponding log-odds measure of difficulty to generate satisfaction.

More details for item measure are given in Tables 1. This table lists items in order of measure. 'Passenger information' is the attribute of service that has more difficulty to generate satisfaction followed by 'Staff behavior' and 'Train cleanliness'. The attributes that have less difficulty to generate satisfaction are 'Security' and 'Regularity of service'. Two types of fit statistics are given for each item. Ideally, for rating scale model the infit and outfit mean-square will be 1.0, but values included between 0.6 and 1.4 indicate that the deviation from expectation is acceptable (Bond e Fox, 2001). In particular, the 1.16 infit mean-square statistic for the item 'Passenger information' is the highest variation between observed data and the Rasch model predicted (16% more variation). The 'Train cleanliness' and 'Station cleanliness' items have 18% less variation in the observed response than the value that had been modeled. Similarly, outfit mean-square for the item 'Passenger information' has the highest variation (20%) and 'Station cleanliness' has 17% less variation in the observed data than the value modeled. Finally, the point-measure correlation is, for each item, a positive value that is included between 0.58 and 0.68, and these values show the absence of mis-scoring and anormal polarity.

Table 1. Items statistics

Item	Model		Infit	Outfit	Ptmea	Exact	Match
	Measure	S.E.	MnSq	MnSq	Corr.	Obs%	Exp%
Passenger information	.63	.03	1.16	1.20	.63	49.0	50.9
Staff behavior	.45	.03	1.08	1.10	.65	52.4	52.5
Train cleanliness	.17	.03	.82	.84	.68	60.2	54.3
Passenger comfort	.10	.03	.89	.93	.64	61.2	55.3
Station cleanliness	.08	.03	.82	.83	.67	61.6	55.3
Personal and financial security	-0.2	.03	1.11	1.10	.62	54.9	55.6
Frequency of service	-.34	.03	1.05	1.09	.59	56.8	57.7
Regularity of service	-.39	.03	1.07	1.07	.58	56.6	57.8
Security	-.67	.03	.99	.96	.59	61.8	58.8

Note: 'Measure' is the estimate for the item difficulty to generate satisfaction. 'S.E.' is the standard error of the estimate. 'Infit MnSq' and 'Outfit MnSq' are infit and outfit mean-square statistic, respectively. 'Ptmea Corr' is the point measure correlation.

Category frequency counts and percentage for the rating scale is shown in Table 2. Similarly to the mean-square infit and outfit, these fit statistics have only small deviation from expectation. The highest deviation for infit and outfit mean-square is given from category 1 with 1.5 and 1.09 respectively.

Table 2. Summary of category structure

Rating scale category (Score)	Category			Infit MnSq	Outfit MnSq
	Count	Percentage	Measure		
Strong dissatisfaction (1)	1,444	7%	(-2.83)	1.05	1.09
Dissatisfaction (2)	4,621	22%	-1.02	1.02	1.06
Satisfaction (3)	9,736	47%	.85	.93	.92
Strong satisfaction (4)	4,838	23%	(3.11)	.98	.99

Note: 'Measure' is the estimate for each category. 'Infit MnSq' and 'Outfit MnSq' are infit and outfit mean-square statistic respectively.

MetroNapoli S.p.A. is interested to investigate whether the satisfaction depends on the personal characteristics of passengers (profession, age, purpose of travel, and sex). With this goal, ANOM decision chart was built. They have a centerline, located at the overall mean, and upper and lower decision limit (see equation 2). The group means are plotted, and those that fall beyond the decision lines are significantly different from the overall mean. In Figure 2.a the passengers were stratified into eight professional categories (viz., 'freelancer', 'businessman', 'worker', 'employee', 'unemployed', 'housewife', 'student', 'others') and the mean of passenger satisfaction (in logit measure) for each group was plotted in decision chart. Similarly, the passengers were stratified into age groups: 'less than 25', '25 – 40', '40 – 60, 'more than 60' (Figure 2.b), and according to the purpose of travel - 'work', 'school', 'college/university', 'personal practice', 'tourism', 'others' – (Figure 2.c). Finally, the sex of passengers was considered (Figure 2.d). At a 5% level of significance, only two professional categories have a level of satisfaction which is statistically different from the overall mean satisfaction ('student' and 'housewife'). The first can be satisfied with more difficulty, while to satisfy the housewives is easier. In the Figure 2.b it is possible to observe that there is a trend between passenger satisfaction and age groups, when the age of passengers increases is more simply observe passenger satisfied. Regarding the passengers that use MetroNapoli's service, the 'tourism' category is the most satisfied of the overall passengers. Significantly more satisfied of the overall mean are the passengers that use this service for 'personal practice', while the passengers that would use MetroNapoli's service to go to 'college/university' are significantly less satisfied. Finally, there is no difference in term of satisfaction between the 'male' and 'female' passengers.

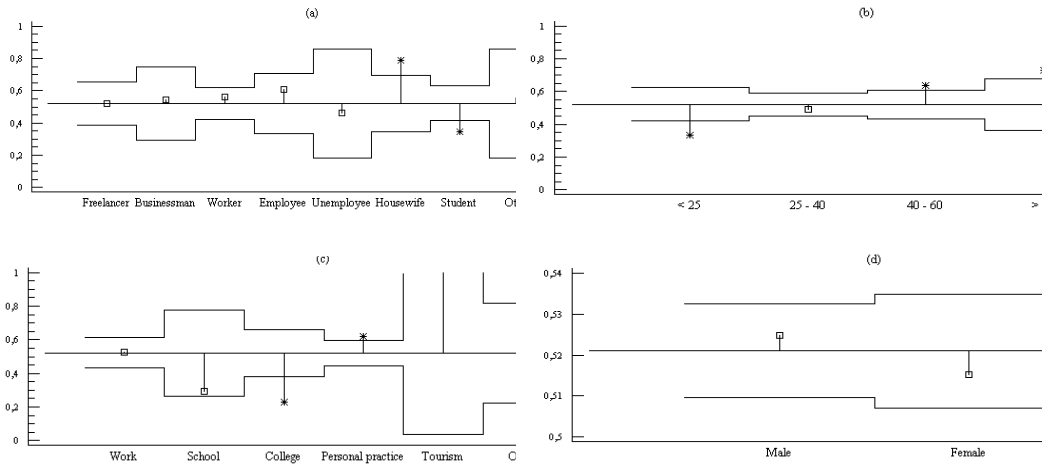

Figure 2. ANOM decision chart for personal characteristics of passengers – Profession (a), Age (b), Reason of travel (c), and Sex (d)

The decision chart helps to verify whether some characteristics of the travel influence the passenger satisfaction. For this reason, in Figure 3.a - 3.c passenger satisfaction considers the day when the interview was carried out, the frequency of travel per week ('3 or less', '4 – 6', '6 – 10', 'more than 10'), and modes of transportation that the passengers had used before arriving at the station where interview was held (intermodal passenger transportation) - 'None', 'Line 1', 'Line 6', 'Cable railway', 'Line 2', 'Railway', 'Bus', 'Cumana', 'MetroCampania', 'Circumvesuviana', and 'Car' are the categories considered. ANOM decision chart shows how, at a 5% level of significance, only Mondays and Fridays have a level of passenger satisfaction different from the overall mean. Respectively, Monday is the day of week which is harder to generate satisfaction, while Friday is easier (Figure 3.a). Regarding the relationship between the passenger satisfaction and the weekly frequency of travel, it is possible to observe that only when the number of times the users travel weekly is 'more than 10' the passenger satisfaction is statistically lower than the overall mean (Figure 3.b). Figure 3.c shows only a category that gives a level of satisfaction different from the overall mean. Passengers who had used the MetroNapoli's service after the Circunvesiana's service are statistically more satisfied than others.

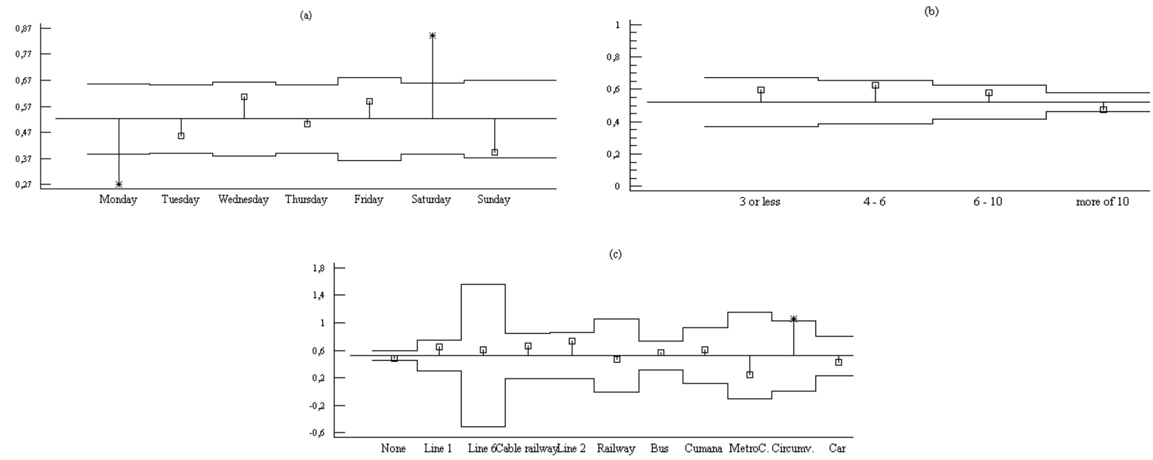

Figure 3. ANOM decision chart for characteristics of travel – Day of interview (a), Frequency of travel per week (b), and Intermodal passenger transportation (c).

4. Conclusions and discussion

The principal purpose of this paper has been to show how the Rasch analysis could be linked with other statistical methods to extract more information of data. Here, Rasch analysis was used to measure the passenger satisfaction of MetroNapoli S.p.A., and then, to get more fine information, the statistical relationship between the linear measurement of passenger satisfaction and some personal passenger information was studied. In this way ANOM decision chart is a tool that well integrated Rasch analysis.

The second aim of the work is show how, in the same case, ANOM is a good alternative to the most popular ANOVA to compare a group of means. In fact, it offers two clear advantages over ANOVA. First of all, it is more intuitive and provides an easily understood graphical result, which clearly indicates the means that are different from the overall mean. Finally, it sheds light on the nature of the differences among the groups. Moreover, in the same case, ANOM is able to give evidences on differences between the groups which can be seen in the ANOVA table. In fact, Figure 3.c shows that the passengers who had used the Circumvesuviana's service were more satisfied than the overall mean. In this case, and generally, when only one (or few) category is very different from the overall mean, ANOVA does not reject the null hypothesis (Table 3).

Table 3. ANOVA

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Between groups	20,3728	10	2,03728	1,97	0,0331
Within groups	2323,68	2243	1,03597		
Total (Corr.)	2344,05	2253			

References

1. Andrich, D. **A rating formulation for ordered response categories,** Psychometrika, 43, 1978, pp. 561-574
2. Bond, T. G. and Fox, C. M. **Applying the Rasch model: Fundamental measurement in the human sciences,** London: Erlbaum, 2001
3. Gallo, M. **The scaling problems in service quality evaluation,** Metodološki Zvezki, 2; 2007, pp. 165-176
4. Linacre, J. M. and Wright, B. D. **A user's guide to WINSTEPS: Rasch model computer program,** Chicago: MESA Press, 2000
5. Nelson, P. R. **Exact critical point for the analysis of means,** Comunications in Statistics: Theory and Methods, 11(6), 1982, pp. 699-709
6. Nelson, P. R., Wludyka, P. S. and Copeland, K. A. F. **The Analysis of Means: A graphical method for comparing means, rates, and proportions,** Philadelphia: Siam, 2005
7. Nelson, P. R. **Power curves for the Analysis of Means,** Technometrics, 27, 1985, pp. 65-73
8. Ott, E. R. **Analysis of Means – A Graphical Procedure,** Industrial Quality Control, 24, 1967, pp. 101-109. Reprinted in Journal of Quality Technology, 15, 1983, pp. 10-18
9. Wright, B. D. and Masters, G. N. **Rating scale analysis,** Chicago: MESA Press, 1982
10. Wright, B.D. and Linacre, J.M. **Observations are always ordinal; measurements, however, must be interval,** Chicago, IL: MESA Psychometric Laboratory, 1989