San Francisco International Airport

2011 Airline Passenger Survey Analysis

This report analyses the passenger information and satisfaction related factors and attributes on overall passenger satisfaction with San Francisco international airport.

Parag Shah

MAT 460 Sample Survey Methods

San Francisco International Airport

2011 Airline Passenger Survey Analysis

This report presents analysis of an airport-wide customer survey conducted on behalf of San Francisco International Airport (SFO) by Corey, Canapary and Galanis. The San Francisco International Airport (SFO) conducts an annual customer survey to gauge customer satisfaction on number of issues, with particular emphasis on airport environment, safety and cleanliness.

The latest survey was taken in 2011, interviews were held at all terminals and boarding areas from 11th to 26th May 2011. The questionnaires were available in English, Japanese, Chinese and Spanish.

Respondents were first asked a series of travel related questions along with few demographic questions, followed by series of closed assessment questions. Assessment questions were related to rating different airport service areas and overall airport rating. Demographic and travel information is used as basis of comparison for rating question and to describe characteristics of respondents at travel information level.

The goal of this report is to estimate the passenger rating of San Francisco airport and determine important factors which affects the rating i.e. find linear dependence of rating with other factors, if it exists.

The complete data dictionary can be found at

http://www.flysfo.com/web/export/sites/default/download/about/news/pressres/stats/datasets/2011SFOCustomerSurvey.zip

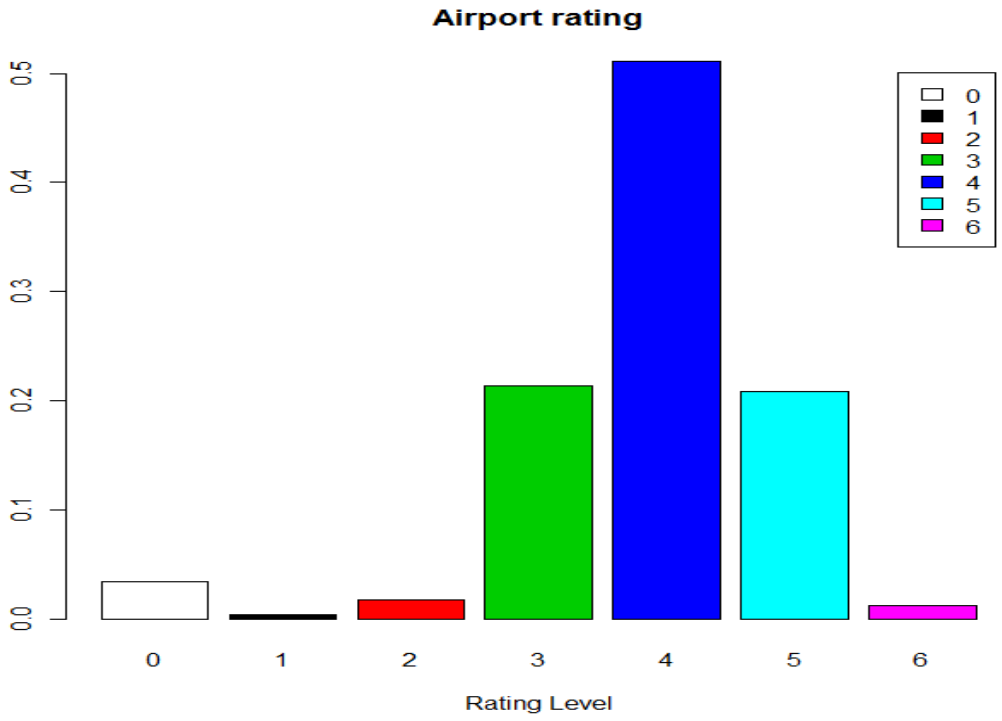
Survey Summary

There were 65 questions asked to passengers and numbers of respondents were 3872. Most of the questions were categorical with exception of Destination, time of arrival/departure, state, zip code and country of residence.

Respondents were asked to rate airport facilities, services and airport environment along with overall rating of the airport. Along with airport ratings, passenger travel information were also collected like main purpose of the trip, arrival and departure time, mode of transportation, residence information, age, gender and household income level. The airport ratings were divided in 5 levels, 1 being Unacceptable and 5 being Outstanding.

43% of people who were travelling on business trip rated SFO airport as whole being Unacceptable and 45% travelling on vacation/leisure rated outstanding. 38% of people who were travelling first time through SFO rated worst compared to 61% who travelled second time through SFO. 40% of people who gave good review of Art and exhibition in airport also gave good review of airport. 25% of passengers, who gave Unacceptable rating for restaurants, also gave bad ratings for airport whereas 30% who liked restaurants on SFO gave good rating for airport. Out of 70% of people who liked retail shops and Concessions, only 37% gave good rating for airport. Out of 70% of those who liked signs and direction on SFO, 58% gave favorable ratings. 50% of people in survey gave good rating for airport safety and those who gave good rating also gave good rating for airport. 60% of those who were coming from different region or had connecting flight gave outstanding rating for airport where as people in northern California or bay area gave overall poor rating.

The graphs below shows distribution of ratings. Overall airport ratings consists of 50% of good ratings. The mode of overall ratings was found and it was 4 and mean was 3.84.



Since there are 65 variables, to find linear relationship between overall airport rating and other variable, spreaman correlation coefficient was computed for continuous variables and chi-sq test or fishers test was conducted for categorical variables.

Based on above mentioned test, we found that Overall Airport rating was significantly correlated (at significance level 0.05) with most of the other rating except for country of residence, zip code, state and SFO website rating. Most significant (based on p-value and chi-sq score) variable affecting the overall rating was airports parking facility rating, airtrain rating and airport cleanliness factors.

Non-Response and Missing Data in Survey

Several rating questions had non-response and missing data coded as 0 or 6. Some of them can be left untreated since they are not used for analyses e.g. rating of SFO website has missing data but since we are not using those variables we can left it as it is. But if we have to than we can use random imputation since those variables are not correlated with response variable or design.

For other rating variables, since some of them are dependent on response variable these are estimated using post-stratification weight adjustments to reduce the overall bias.

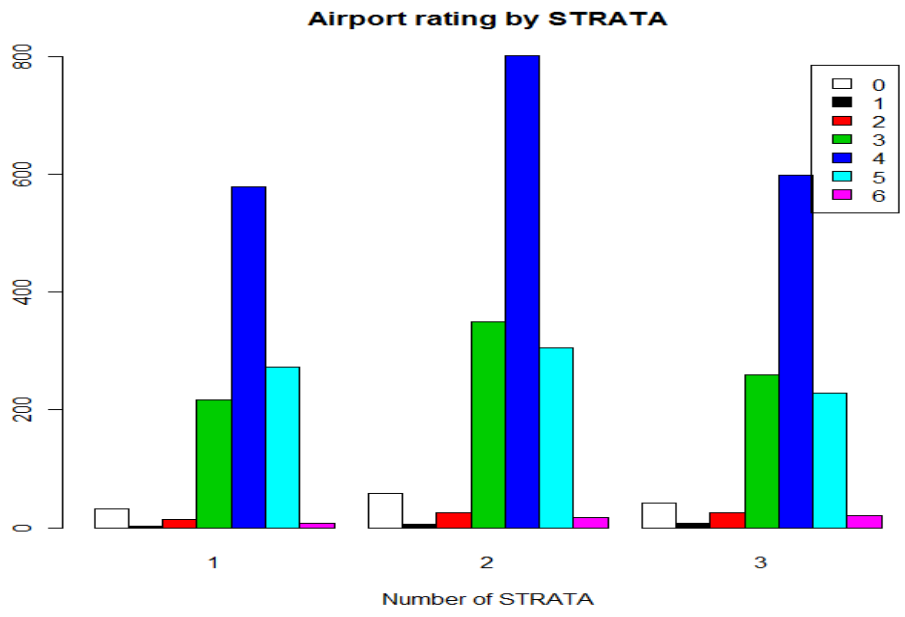
NOTE: Some of rating variables has levels which have not been defined in data dictionary, so these could be non-response or measurement error or simply omitted in dictionary. To rectify this levels which has not been defined are merged with closest level. E.g. restaurants ratings supposed to have levels 1 to 5 but in data we found some rows has 0 and 6 value. So 0 has been merged with 1 and 6 with 5. This is done for all the ratings which have these problems.

Survey Design

Design & Analysis was conducted using stratified random sampling. The stratum was chosen based on time of day respondent's flight was schedule to leave. The data was group into 3 stratum, strata 1 if departure was before 11 AM, strata 2 if departure was between 11 and 5 PM and strata 3 if departure was after 5 PM. This was done to reduce the variability of overall rating.

After dividing the data, stratum 1 had 1125 observations, stratum 2 had 1564 observations and stratum 3 had 1183 observations.

The rating distribution with respect to strata is shown below.



To estimate the overall rating, first proportional allocation was used with sampling rate of 30% within each stratum. Thus random sample of 337, 469, 354 observation was taken (SRSWOR) from each stratum respectively and mean and variance of overall rating was calculated.

This initial stratified design was conducted using strata function in sampling package, which calculates the inclusion probabilities using given sampling rate within each stratum. This was used to calculate the initial weights and population size which was used to create the initial survey design using survey package.

The mean and variance of overall rating was calculated after stratifying with 30% sampling rate:

Mean	Variance
3.83	1.15

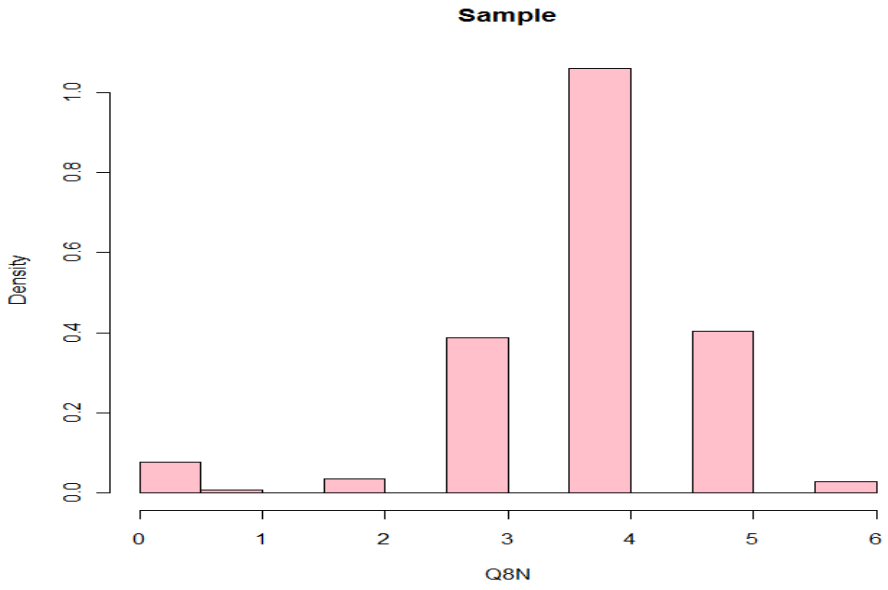
The mean in survey data was 3.84, whereas in sample data it is 3.83 and variance 1.15 which high. The mean and standard error estimates within strata are:

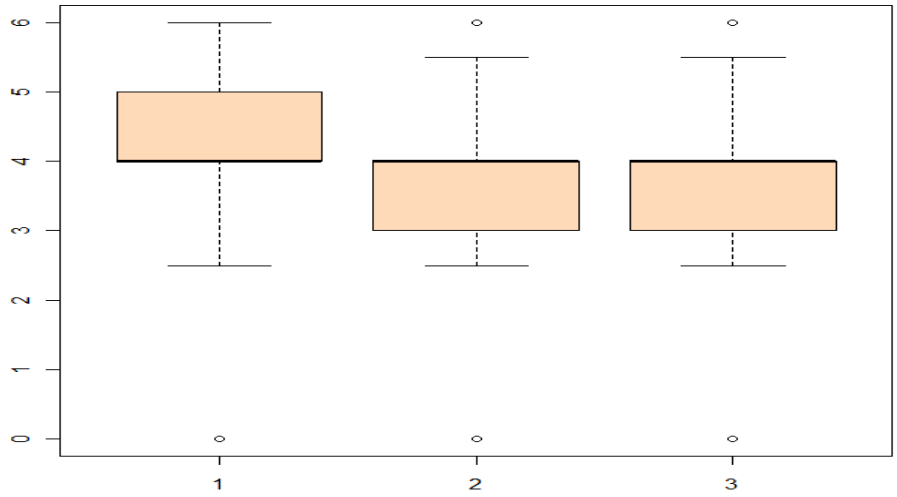
STRATA	Mean of Overall Rating	SE
1	3.96	0.028
2	3.76	0.027
3	3.78	0.031

Strata 2 and 3 have lower rating than strata 1 which is higher than population mean. The upper and lower bound on confidence interval is presented below:

Quintile	Lower Bound	Upper Bound
0.25	3	4
0.5	4	4
0.75	4	4

The Density of response variable and box plot using survey data is presented below.





As mentioned above variability between stratum 2 and 3 is small but between 1 and 2/3 is relatively large. The histogram of sample rating is similar to survey data.

The mean of overall rating in sample data is close to survey data but variance could have been smaller than 1.15. This suggests better sampling scheme is needed which adjusts the inclusion probability based on ratings within each stratum and add more sample to reduce the variability between stratum.

Disproportionate Allocation using Unequal Probability

To reduce the variability, the inclusion probability was adjusted to take into account number of observations associated with each level of overall airport rating within each stratum. The sampling rate was kept same as before i.e. 30% of observation within each stratum was sampled using poisson method since variable of interest was categorical. But sampling frame was different than before. Thus population size and weights were also readjusted based on inclusion probability and new survey design was made.

The mean and variance after readjustment of weights:

Mean	Variance
3.85	0.87

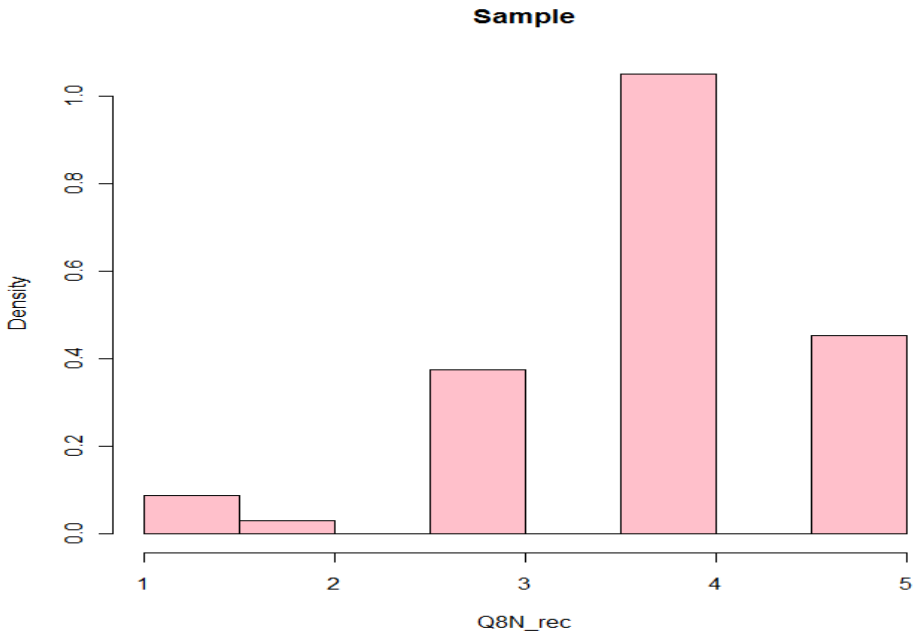
The mean in new sample data is same 3.87 and variance is reduced from 1.15 to 0.86 which is better than before. The mean and standard error estimates within strata are:

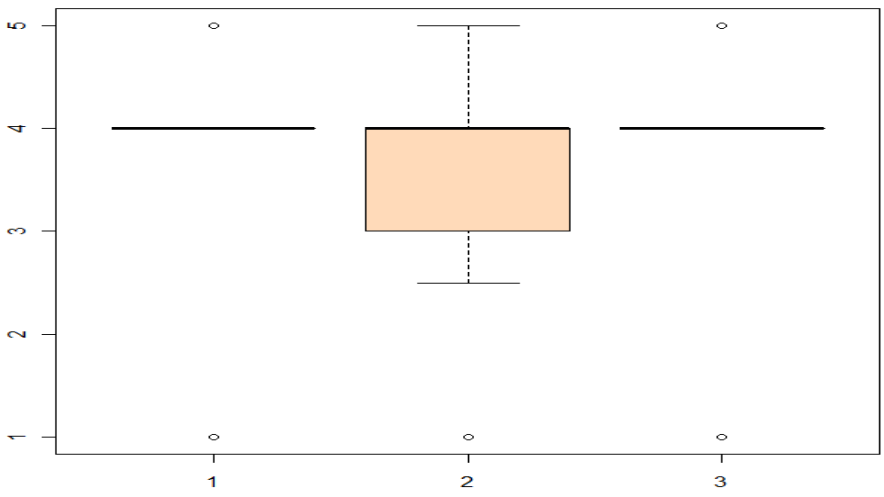
STRATA	Mean of Overall Rating	SE
1	3.95	0.039
2	3.82	0.038
3	3.83	0.040

Strata 2 have lower rating than strata 1 and 3. The upper and lower bound on confidence interval is presented below which is same as before:

Quintile	Lower Bound	Upper Bound
0.25	3	4
0.5	4	4
0.75	4	4

The Density of response variable and box plot using survey data is presented below.





As mentioned above variability between stratum 1 and 3 is small to none but between 2 and 1/3 is relatively large.

Further attempt was made to reduce the variance using post stratification and calibration by using auxiliary information available in survey data.

Mean and variance estimates using post stratification

Mean	Variance
3.85	0.86

The variance estimate was decreased slightly from 0.87 to 0.86 and mean of overall airport rating was same. The calibration method also gave similar estimation.

Model Based Approach to Stratified Sampling Using Unequal Weights

The rating was estimated using Generalized Linear Model and airport rating was predicted using STRATA and AirTrain. Mean and Variance estimates were much better compared design based approach with post-stratification.

Mean	Variance
3.84	0.11

The model was fitted using STRATA and AirTrain factors, using svyglm method and weights calculated in design based approach.

Only problem was residuals were not normally distributed.

Summary

Survey was sampled using stratified random sampling first using proportional allocation with equal weights and sampling rate of 30% within each stratum. The mean overall rating was close to survey data but variance was large around 1.15. To reduce the variance multiple techniques were used, merging of category and post stratification using unequal weights gave better result and variance was reduced to 0.87. In the end GLM model was used using unequal weights, this gave much better estimation and lower variance.

Conclusion

This survey was taken to identify the important factors linked to overall airport ranking and to accurately estimate the ratings. Using stratified survey design and other statistical techniques this goal can be achieved. Since SFO conducts survey with similar questionnaire every year, the estimates can be made better using two-stage sampling design using older survey or using some other data set which can give estimate of total population like number of passengers travel using SFO etc..