

**CONSIDERATION OF CUSTOMER SATISFACTION AT SAN FRANCISCO INTERNATIONAL AIRPORT AND ITS IMPLICATIONS**

**Mingzhen Xiong**  
Bentley University  
Waltham, MA 02452  
U.S.A

**Paul D. Berger**  
Bentley University  
Waltham, MA 02452  
U.S.A

Corresponding author: Dr. Berger

**Abstract**

This paper examines which services at the San Francisco Airport (SFO) appear to be drivers of how often a traveler flies out of SFO. We use the latest available data from SFO and different variables and build upon previous satisfaction studies of SFO and other airports. We use descriptive statistics, t-tests for differences in mean, simple and multiple stepwise regression analyses. We perform separate analyses for leisure travelers and business travelers.

**Keywords:** Satisfaction study, Airport services, San Francisco International Airport, Stepwise regression analysis, Descriptive statistics

**Introduction**

Does the quality of service in an airport influence people's choices to go to an airport? We might conclude that it has no impact, since customers of the airport have almost no choice if it is a large international airport; people must go there if they decide to travel by plane. However, is this completely true? Does the service of airport really not matter? What if there are other airports that, at least for some destinations, are "competitors?" It is interesting to study services in an airport to determine whether selected services receive higher satisfaction rates, and whether these satisfaction rates have any influence at all on people's airport choices. We analyze the above issue using customer satisfaction survey data from an airport. Fortunately, San Francisco International Airport (SFO) publishes their customer survey data based on more than 2000 customers every year. We use the most up-to-date data available from SFO. We use descriptive statistics, selected results of t-tests for differences in means, simple and multiple-stepwise linear regression analysis, to address these issues.

## **LITERATURE REVIEW**

Harvey (1987) conducted the first study at the San Francisco Airport (SFO). He focused on choice of 3 airports in the San Francisco Bay area including SFO. His main findings were that ground access time and frequency of direct flights to destinations were the key factors in airport choice. The next study of SFO consumers was by Pelse, Nijkamp, and Rietveld (2001); they built models to analyze the influence of choosing an airport and concluded that the access time to the airport was among the most influential determinants of the choice of airport and airline. They also presented an expanded study in Pelse et al. (2003). Also, other researchers, such as Fodness (2007), as well as Curtis, Rhoades & Waguespack (2012), found that high-quality airport service and good airline experience are very important for an airport to attract more customers. During recent years, the expectation and satisfaction of customers became more and more important. In Kamarudin's article published in *Airports Council International World Reports* (2015), it was shown that satisfied customers are likely to spend more money in the airport than those who are not satisfied with airport services.

Several other studies have analyzed satisfaction, loyalty, and general evaluations of airports by consumers of specific airports. Some of these have used data from past surveys (i.e., prior to the latest-available 2016 survey) at SFO. These would include Wang, Hong & Berger (2015) and Wang, Hong, and Berger (2016); these studies address different independent variables than the current study. Other more recent studies of customer satisfaction or attitude or choice at airports other than SFO include Mattazo et al. (2012), Suki (2014), Al Refaie (2014), Chang (2013), and Baker (2013), Ishii et al. (2009), and Hess & Polak (2005). These studies considered data from airports in a variety of locations, including Taiwan, Malaysia, and Jordan. These are discussed in the aforementioned Wang et al. (2016). As we noted, our study uses different independent variables than previous studies, and business travelers and leisure travelers are analyzed separately. Harvey's work (1987) also treated business and leisure travelers separately.

## **METHODOLOGY**

The dataset used in this research is from San Francisco International Airport official website – 2016 Customer Survey Dataset (SFO Customer Survey Data, 2016). SPSS was used to analyze the dataset. Our analysis was divided into two parts. Part one uses the means of each question's score to see which services received the highest and lowest satisfaction rates, as well as selected t-tests for differences in means. Part two of our analysis uses simple and multiple/stepwise linear regression to examine the relationship between airport service satisfaction rates and customer attendance frequency.

Q2 on the database states the main purpose of the trip. To analyze business and leisure travelers separately, we classified the following options as "Business Travel," and we recoded each category as a "1":

“Business/Work/Job Interview”  
“School/school event”  
“Conference/convention”  
“Military” and

“Volunteer/Political/Religious”

Then, the other selections, as listed below, were considered as "Leisure Travel," and we recoded each of these categories as a "0":

“Pleasure/Vacation/Recreation”

“Visit friends or relatives”

“Wedding/funeral/graduation/ reunion”

“Escorting others (children/elderly)/personal errands/medical purpose”

“Moving/immigration/traveling between homes”

“Other (specify)” and

“Blank/non-response”

The “Split File” function in SPSS was used to facilitate separate treatment of the two types of travelers.

The Q7 series contained 15 questions, each measuring the assessed quality of a different airport’s service; one of them "Q7ALL," was a measure of all of the services, in aggregate. These are listed in Table 1.

Q7ART: Artwork and exhibitions
Q7FOOD: Restaurants
Q7STORE: Retail shops and concessions
Q7SIGN: Signs and directions inside SFO
Q7WALKWAYS: Escalators/Elevators/Moving walkways
Q7SCREENS: Information on screens/monitors
Q7INFODOWN: Information booths (lower level - near baggage claim)
Q7INFOUP: Information booths (upper level - departure area)
Q7WIFI: Accessing and using free WiFi at SFO
Q7ROADS: Signs and directions on SFO airport roadways
Q7PARK: Airport parking facilities
Q7AIRTRAIN: AirTrain
Q7LTPARKING: Long term parking lot shuttle (bus ride)
Q7RENTAL: Airport Rental Car Center
Q7ALL: SFO Airport as a whole

Table 1: A list of the Q7 set of airport services quality/satisfaction assessments

The Q9 series contained 6 questions, each measuring an assessment of the cleanliness of different airport locations; as in the Q7 series, one of the questions, "Q9ALL," considered cleanliness overall. These are listed in Table 2.

Q9BOARDING: Boarding areas  
Q9AIRTRAIN: Airtrain  
Q9RENTAL: Airport Rental Car Center  
Q9FOOD: Airport restaurants  
Q9RESTROOM: Restrooms  
Q9ALL: Overall cleanliness

Table 2: A list of the Q9 set of airport (CLEANLINESS) quality/satisfaction assessments

These 21 questions are the set of independent variables (X's) in our regression analyses. The questions in Q7 series were evaluated by an interval scale, which defines 5 as “Outstanding” and 1 as “Unacceptable”<sup>1</sup>. The original data also used 6 as “Have never used or visited / Not applicable” and 0 as “Blank”. The questions in Q9 series were evaluated by a similar interval scale, which defines 5 as “Clean”, and 1 as “Dirty”<sup>2</sup>. The 0 and 6 have same meanings as in the Q7 series. The “6's” and “0's” were dropped from the data set.

In all of the regression analyses, the dependent variable (Y) was Q5TIMESFLOWN, which represents the number of times a traveler flew out of SFO in the past 12 months. In this question, 1 represents 1 time, 2 represents 2 times, 3 represents 3-6 times, 4 represents 7-12 times, 5 represents 13-24 times, 6 represents "More than 24 times," 0 represents Blank/Multiple responses. Sensibly, all "0's" were dropped from the analysis. This dependent variable measure, now ranging from 1-5, was treated as an interval-scale variable, with the recognition that this was not ideal, but, in our view, the best available action. After all responders with "0's" and "6's" were dropped from the database, as described above, we had a sample size of 333 travelers remaining.

### MEANS - DISCUSSION AND ANALYSIS

We first ran the “Descriptives” function to find the mean of all 21 X's and made various comparisons via means testing. Prior to examining the means, our initial hypotheses, based on some common sense and the previous literature, were that, for both business or leisure travelers, the services they care most about in the Q7 series would be signs and directions, screens with airplane information, and escalators, elevators and moving walkways. Therefore, based on the assumption that airport management knows this, and would likely pay more attention to these services, we hypothesized that travelers are more likely to give higher satisfaction ratings for Q7SIGN, Q7ROADS, Q7SCREENS, and Q7WALKWAYS. For the Q9 series, With a similar assumption about the knowledge of airport management, we hypothesized that the highest ratings would be for Q9RESTROOMS and boarding areas (Q9BOARDING). It was not clear what the results would be for "Q7ALL" and "Q9ALL."

Table 3 displays the mean and standard deviation for **Leisure Travelers** and Table 4 similarly for **Business Travelers**.

<sup>1</sup> Technically, one may argue that the scale is "merely" ordinal. However, such scales are routinely analyzed as if representing an interval-scale.

<sup>2</sup> Here again, the same comment as noted in footnote 1 applies here,

**Descriptive Statistics**

	N	Minimum	Maximum	Mean	Std. Deviation
Q7ART	216	1	5	4.08	.939
Q7FOOD	216	1	5	3.97	.917
Q7STORE	216	1	5	3.98	.920
Q7SIGN	216	2	5	4.22	.811
Q7WALKWAYS	216	2	5	4.24	.752
Q7SCREENS	216	2	5	4.25	.785
Q7INFODOWN	216	2	5	4.19	.826
Q7INFOUP	216	2	5	4.19	.826
Q7WIFI	216	1	5	4.20	.887
Q7ROADS	216	1	5	4.14	.875
Q7PARK	216	2	5	4.11	.866
Q7AIRTRAIN	216	1	5	4.13	.860
Q7LTPARKING	216	1	5	4.07	.900
Q7RENTAL	216	2	5	4.06	.903
Q7ALL	216	1	5	4.22	.764
Q9BOARDING	216	2	5	4.40	.789
Q9AIRTRAIN	216	1	5	4.36	.835
Q9RENTAL	216	1	5	4.33	.829
Q9FOOD	216	2	5	4.38	.804
Q9RESTROOM	216	2	5	4.29	.848
Q9ALL	216	2	5	4.40	.771
Valid N (listwise)	216				

Table 3: Means and Standard deviations for **Leisure Travelers**

**Descriptive Statistics**

	N	Minimum	Maximum	Mean	Std. Deviation
Q7ART	117	1	5	3.94	.967
Q7FOOD	117	1	5	3.65	1.085
Q7STORE	117	1	5	3.78	1.026
Q7SIGN	117	1	5	3.98	.965
Q7WALKWAYS	117	1	5	3.99	.942
Q7SCREENS	117	1	5	3.99	.969
Q7INFODOWN	117	1	5	3.80	1.069
Q7INFOUP	117	1	5	3.85	1.047
Q7WIFI	117	1	5	3.91	1.055
Q7ROADS	117	1	5	3.94	.922
Q7PARK	117	1	5	3.80	.958
Q7AIRTRAIN	117	1	5	4.00	.910
Q7LTPARKING	117	1	5	3.82	.970
Q7RENTAL	117	1	5	3.84	.909
Q7ALL	117	1	5	3.92	.902
Q9BOARDING	117	1	5	4.22	.930
Q9AIRTRAIN	117	1	5	4.16	.909
Q9RENTAL	117	1	5	4.09	.965
Q9FOOD	117	1	5	4.15	.958
Q9RESTROOM	117	1	5	4.01	.951
Q9ALL	117	1	5	4.13	.886
Valid N (listwise)	117				

Table 4: Means and Standard deviations for **Business Travelers**

From Table 3, we can see that the highest four means in the Q7 series are Q7SCREENS (4.25), Q7WALKWAYS (4.24), Q7SIGN (4.22, tied with Q7ALL). The highest three means in the Q9 series are Q9BOARDING (4.40), Q9ALL (4.40) and Q9FOOD (4.38). The result is close to what we "postulated" - except for Q9FOOD replacing Q9RESTROOM. However, there are no statistically-significant differences among these high-ranking services. In fact, there are only a few statistically-significant differences among all of the Q7 services at  $\alpha = .05$ ; the only two groups that are statistically-significant at  $\alpha = .05$  are (Q7SCREENS, Q7WALKWAYS, Q7SIGN, Q7WIFI) and (Q7FOOD, Q7STORES). For the Q9 series, there are no statistically-significant differences at all at  $\alpha = .05$ .

From Table 4, we can see that the highest four means in the Q7 series are Q7AIRTRAIN (4.00), Q7SCREENS (3.99, tied with Q7WALKWAYS), and Q7SIGN (3.98). The highest three means in the Q9 series are Q9BOARDING (4.22), Q9AIRTRAIN (4.16) and Q9FOOD (4.15). The rank-order results at the top are almost the same as what we got for leisure travelers. The only difference is that Q7AIRTRAIN was in the top group business travelers. Still, the results are similar to what was "postulated." The only statistically-significant difference at  $\alpha = .05$  is for the Q7 series between Q7FOOD and 9 of the other 14 items. Again, there are no statistically-significant differences at  $\alpha = .05$  among any of the Q9 items.

What is also worth noting is that for *every single Q7 and Q9 item, the satisfaction level is higher for the Leisure Traveler than for the Business Traveler*. And, for the Q7 items, 11 out of the 15 items are statistically-significantly different at  $\alpha = .05$  (all except for Q7ART, Q7ROADS, Q7AIRTRAIN, and Q7RENTAL). For the Q9 items, 4 of the 6 items are statistically-significantly different at  $\alpha = .05$  (all except for Q9BOARDING and Q9AIRTRAIN).

## **REGRESSIONS - DISCUSSION AND ANALYSIS**

### ***Leisure Travelers - Q7 series***

We performed a stepwise regression analysis including, as eligible, the 15 independent variables in the Q7 series. Recall that the dependent variable, Y, is Q5TIMESFLOWN, the number of times the responder has flown out of SFO in the past year. The results are displayed in Figures 1a and 1b:

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.161 <sup>a</sup>	.026	.021	1.234
2	.250 <sup>b</sup>	.063	.054	1.213
3	.279 <sup>c</sup>	.078	.065	1.206
4	.303 <sup>d</sup>	.092	.075	1.200

- a. Predictors: (Constant), Q7LTPARKING
- b. Predictors: (Constant), Q7LTPARKING, Q7INFOUP
- c. Predictors: (Constant), Q7LTPARKING, Q7INFOUP, Q7STORE
- d. Predictors: (Constant), Q7LTPARKING, Q7INFOUP, Q7STORE, Q7FOOD

Figure 1a: **Model Summary** for Q7 series of independent variables, leisure travelers

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
4	(Constant)	2.575	0.463		5.568	0.000
	Q7LTPARKING	-0.373	0.128	-0.269	-2.915	0.004
	Q7INFOUP	0.422	0.137	0.280	3.088	0.002
	Q7STORE	-0.441	0.169	-0.325	-2.608	0.010
	Q7FOOD	0.298	0.162	0.219	1.833	0.068

a. Dependent Variable: Q5TIMESFLOWN

Figure 1b: **Coefficients** for Q7 series of independent variables, leisure travelers

We can note that 4 of the 15 variables were included in the final model of the stepwise regression analysis. We used a "p-to-enter" of .10. The R-squared value of the final model, as seen in Figure 1a, is .092, or 9.2%. This indicates that these 4 variables are estimated to explain about 9.2% of the variability in Y (Q5TIMESFLOWN). We are not "shocked" that the R-squared value is not higher, since, as we noted in the introduction, in many cases, there is really little choice available to the customer - he/she wants to, or needs to, fly to wherever is necessary/ desired, and quite often, there are often no options other than the specific airport available -

regardless of the quality of the services at the airport. Still, we should note that the regression analysis results are highly statistically significant and, indeed, 3 of the 4 variables in the final model are statistically significant at  $p \leq .01$ . This "lack-of-shock" statement applies also to subsequent results, but we will not repeat this sentiment each time. Still, we must admit that some of the results seem to not be logical, except under a narrow, and non-verifiable assumption noted below (and sadly, this statement also applies for some of the subsequent results). Here are the interpretations of the coefficients and corresponding discussion; all discussion of being more vs. less satisfied, and more vs. less often, pertain specifically to SFO.

Q7LTPARKING: The coefficient is -0.373, indicating that Q5TIMESFLOWN will increase with a decrease in Q7LTPARKING. This implies that, assuming that other variables in the model are held constant, if a leisure traveler is less satisfied with the long-term parking lot shuttle (bus ride), he/she will fly more often. In a simple regression of Q5TIMESFLOWN and Q7LTPARKING, the coefficient was also negative<sup>3</sup>. The only explanation we can think of that would account for this, and several other "seemingly backwards" results is to reverse the cause-and-effect mode of thinking, and conclude that people who fly more often perhaps have the type of disposition or personality that they are more often "less satisfied with life" and rate satisfaction with most things as "low," whereas people who fly less often have the type of disposition or personality that they are more easily satisfied with things in general and tend to rate most things more highly. These days, in most people's opinion, traveling is much less enjoyable than it used to be, and is often viewed as the most unpleasant part of a trip, and that could lead to the result that those who fly more often are less happy people (in general, or with respect to flying), and thus, more likely to not rate satisfaction as high for most things. We will not repeat this entire explanation every time there is a "seemingly backwards" statistically-significant result, but just refer to it as the LHP ("less happy people") result.

Q7INFOUP: The coefficient is 0.422, indicating that Q5TIMESFLOWN will increase with an increase in Q7INFOUP. This implies that, assuming other variables in the model are held constant, if a leisure traveler is more satisfied with information booths (upper level – departure area), he/she will fly more often. In a simple regression of Q5TIMESFLOWN and Q7INFOUP, the coefficient was also positive.

Q7STORE: The coefficient is -0.441, indicating that Q5TIMESFLOWN will increase with a decrease in Q7STORE. This implies that, assuming other variables in the model are held constant, if a leisure traveler is less satisfied with the retail shops and concessions in the airport,

---

<sup>3</sup> It must be noted that the interpretations are not identical. The multiple (stepwise) regression results indicates that there is a significant difference in the stated direction, holding the other variables in the model constant. The simple regression result is indicating that there is a significant difference in the stated direction without controlling for any other variables. *It is NOT pre-ordained that both results must match, either in terms of significance or in terms of direction.* The spirit of this footnote holds true for all of the other simple regression results we discuss.

he/she will fly more often. Another LHP result? ( $p = .01!!$ ) In a simple regression of Q5TIMESFLOWN and Q7STORE, the coefficient was also negative.

**Q7FOOD:** The coefficient is 0.298, indicating that Q5TIMESFLOWN will increase with an increase in Q7FOOD. This implies that, assuming other variables in the model are held constant, if a leisure traveler is more satisfied with airport restaurants, he/she will fly more often.

None of the other Q7 services are statistically significant at  $\alpha = .05$  when performing a simple regression between each and Q5TIMESFLOWN. However, the simple regression between Q7RENTAL and Q5TIMESFLOWN has  $p = .069$ ; apparently, Q7RENTAL, which did not make the final stepwise regression model, is redundant with one or more of the four variables that did make the final model.

***Leisure Travelers - Q9 series***

We next performed a stepwise regression analysis including, as eligible, the 6 independent variables in the Q9 series. the results are displayed in Figures 2a and 2b:

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.131 <sup>a</sup>	.017	.013	1.239
2	.212 <sup>b</sup>	.045	.036	1.225
3	.240 <sup>c</sup>	.057	.044	1.220

a. Predictors: (Constant), Q9RESTROOM  
 b. Predictors: (Constant), Q9RESTROOM, Q9BOARDING  
 c. Predictors: (Constant), Q9RESTROOM, Q9BOARDING, Q9RENTAL

Figure 2a: **Model Summary** for Q7 series of independent variables, leisure travelers

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.076	.436		7.054	.000
	Q9RESTROOM	-.192	.100	-.131	-1.930	.055
2	(Constant)	2.545	.481		5.291	.000
	Q9RESTROOM	-.515	.163	-.350	-3.161	.002
	Q9BOARDING	.435	.175	.275	2.485	.014
3	(Constant)	2.624	.481		5.453	.000
	Q9RESTROOM	-.421	.171	-.287	-2.459	.015
	Q9BOARDING	.666	.222	.421	3.004	.003
	Q9RENTAL	-.346	.205	-.230	-1.686	.093

a. Dependent Variable: Q5TIMESFLOWN

Figure 2b: **Coefficients** for Q9 series of independent variables, leisure travelers

We can note that 3 of the 6 variables were included in the final model of the stepwise regression analysis. We again used a "p-to-enter" of .10. The R-squared value of the final model, as seen in Figure 2a, is .057, or 5.7%. This indicates that these 3 variables are estimated to explain about 5.7% of the variability in Y (Q5TIMESFLOWN). We should note that some of the regression analysis results are highly statistically significant and, indeed, 2 of the 3 variables in the final model are statistically significant at  $p \leq .015$ . Here are the interpretations of the coefficients and corresponding discussion:

**Q9RESTROOM:** The coefficient is -0.421, indicating that Q5TIMESFLOWN will increase with a decrease in Q9RESTROOM. This implies that, assuming that other variables in the model are held constant, if a leisure traveler is less satisfied with the cleanliness of the restroom, he/she will fly more often. Another LHP result? ( $p = .015!!$ ) In a simple regression of Q5TIMESFLOWN and Q9RESTROOM, the coefficient was also negative.

**Q9BOARDING:** The coefficient is 0.666, indicating that Q5TIMESFLOWN will increase with an increase in Q9BOARDING. This implies that, assuming that other variables in the model are held constant, if a leisure traveler is more satisfied with the cleanliness of the boarding area, he/she will fly more often. This makes sense; indeed, all flyers are exposed to the boarding area, while not all flyers use the restroom. In a simple regression of Q5TIMESFLOWN and Q9BOARDING, the coefficient was actually negative and had a high, very non-significant p-

value. It was difficult to reconcile and interpret this simple-regression result in light of the stepwise regression results.

**Q9RENTAL:** The coefficient is -0.356, indicating that Q5TIMESFLOWN will increase with a decrease in Q9RENTAL. This implies that, assuming that other variables in the model are held constant, if a leisure traveler is less satisfied with the cleanliness of the rental-car area, he/she will fly more often. Another LHP result? In a simple regression of Q5TIMESFLOWN and Q9RENTAL, the coefficient was also negative.

None of the other Q9 variables are statistically significant at  $\alpha = .05$  when performing the simple regressions between each and Q5TIMESFLOWN.

**Business Travelers - Q7 series**

We next performed a stepwise regression analysis including, as eligible, the 15 independent variables in the Q7 series, but now for the business travelers. The results are displayed in Figures 3a and 3b:

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.370 <sup>a</sup>	.137	.130	1.407

a. Predictors: (Constant), Q7LTPARKING

Figure 3a: **Model Summary** for Q7 series of independent variables, business travelers

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4.872	.530		9.187	.000
	Q7LTPARKING	-.575	.135	-.370	-4.273	.000

a. Dependent Variable: Q5TIMESFLOWN

Figure 3b: **Coefficients** for Q7 series of independent variables, business travelers

We can see from Figures 3a and 3b that only 1 of the 15 variables was included in the final model of the stepwise regression analysis: Q7LTPARKING. We again used a "p-to-enter" of .10. The R-squared value of the final model, as seen in Figure 3a, is .137, or 13.7%. This is

nearly 50% higher than the corresponding R-squared for leisure travelers (earlier noted as .092 or 9.2%), and indicates that this 1 variables is estimated to explain about 13.7% of the variability in Y (Q5TIMESFLOWN). However, unfortunately, the sign of the coefficient does not seem to be logical, unless we "invoke" the LHP conclusion. For business travelers, it is easier to envision an LHP scenario, since these travelers likely have less choice in which airport to use and how often to fly. Here is the interpretations of the coefficient and corresponding discussion:

**Q7LTPARKING:** The coefficient is -0.575, which means the Q5TIMESFLOWN will increase with Q7LTPARKING's decrease. Given that there is only one variable in the final stepwise regression model, the results is identical to the simple regression result, and indicates that if a business traveler is less satisfied with the long-term parking lot shuttle (bus ride), he/she will fly more often.

Strangely, for business travelers, virtually all of the Q7 variables were statistically significant ( $p < .05$ ) in simple regressions with Y (Q5TIMESFLOWN) - indeed, all except for Q7ART. We use the word "strangely," since only Q7LTPARKING enters the stepwise regression model. Apparently, once Q7LTPARKING enters the model, all of the other variables that are statistically significant in a simple regression are redundant with Q7LTPARKING - this would be the only reason none of these other variables would enter the stepwise regression.

**Business Travelers - Q9 series**

We lastly performed a stepwise regression analysis including, as eligible, the 6 independent variables in the Q9 series, this time for the business travelers. The results are displayed in Figures 4a and 4b:

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.265 <sup>a</sup>	.070	.062	1.460

a. Predictors: (Constant), Q9ALL

Figure 4a: **Model Summary** for Q9 series of independent variables, business travelers

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4.539	.646		7.030	.000
	Q9ALL	-.452	.153	-.265	-2.952	.004

a. Dependent Variable: Q5TIMESFLOWN

Figure 4b: **Coefficients** for Q9 series of independent variables, business travelers

We can see from Figures 4a and 4b that only 1 of the 6 variables was included in the final model of the stepwise regression analysis: Q9ALL. We again used a "p-to-enter" of .10. The R-squared value of the final model, as seen in Figure 4a, is .070, or 7.0%. This is slightly higher than the corresponding R-squared for leisure travelers (earlier noted as .057 or 5.7%), and indicates that this 1 variable is estimated to explain about 7.0% of the variability in Y (Q5TIMESFLOWN). However, unfortunately, *again*, we need to invoke the LHP proposition. Here is the interpretation of the coefficient and corresponding discussion:

Q9ALL: The coefficient is -0.452, which means the Q5TIMESFLOWN will increase with Q9ALL's decrease. Given that there is only one variable in the final stepwise regression model, the results is identical to the simple regression result, and indicates that if a business traveler is less satisfied with overall cleanliness, he/she will fly more often. *Again*, the authors cannot find a logical explanation for this result.

Four of the other five Q9 variables were statistically significant ( $p < .05$ ) in simple regressions with Y (Q5TIMESFLOWN) - indeed, all except for Q9AIRTRAIN. This is the same "strange" phenomenon that we indicated with the Q7 series of variables. Apparently, once Q9ALL enters the model, all of the other variables that are statistically significant in a simple regression are redundant with Q9ALL.

### **SOME FINAL COMMENTS**

Q7LTPARKING appears to be the most reliable variable to predict how often travelers fly during the past 12 months, but the direction of the relationship seems backwards. That and many of the statistically-significant results are in directions of relationships that seem counter-intuitive, and for which we postulated the "LHP" possibility. For leisure travelers, it would make sense to see that the travelers who were more satisfied with information booths (upper level – departure area) and restaurants would fly more often than those less satisfied with those two services. For business travelers, only the Q7LTPARKING variable survived after the stepwise regression, but all variables except Q7ART are statistically significant in a simple regression. This must be due to a high level of multi-collinearity among the independent variables in analyzing the business travelers' dataset.

### **LIMITATIONS AND DIRECTIONS FOR FUTURE RESEARCH**

This research uses the data from only one airport, SFO, which may not be representative of all airports in the US. The ideal research study on this topic should be based on data from several airports, to see if results are very general across different airports in the U.S., or whether there are large differences from airport to airport in the U.S. Also, ultimately, it might be useful

to distribute the same questionnaire across airports in different countries, to examine global conclusions.

The  $R^2$  values in all of the models are somewhat low in absolute terms. The largest still explains only about one-eighth (12.7%) of the variability in how often one flies. There are clearly many other factors which affect how often one files, and these can be sought. Also, we did not explore non-linear relationships between the dependent variable and independent variables; nor did we examine interaction effects among the independent variables. Future works may wish to consider these modifications.

We repeat again that several results seem to be not make sense unless we adopt the "LHP" hypothesis. There may be other explanations for these "seemingly backwards" results that the authors have not considered. There would seem to be too many of these to ascribe them to simple type I errors, or to the bias for significant results that sometimes occurs when employing stepwise regression (e.g., Fritz & Berger, 2016). SFO is the not only airport where the travelers have no choice about attending, whether they are satisfied or unsatisfied with various airport services, and a "better airport," is, in the large majority of cases, unavailable. Indeed, the vast majority of airports are in this category. It might be interesting to study a situation in which there are two, or even three, airports among which the potential flyer has a legitimate choice. We are not certain how often this situation exists, although it would seem that New York city, with Kennedy and LaGuardia airports, might qualify for this situation, at least for many of the locations to which people wish to fly.

## REFERENCES

- Al-Rafaie, Abbas, Bata, Nour, and Issam Jalham (2014), "Examining Factors That Affect Passengers' Overall Satisfaction and Loyalty: Evidence from Jordan Airport," *Jordan Journal of Mechanical and Industrial Engineering*, 8(2), p. 94-101.
- Baker, David (2013), "Service quality and Customer Satisfaction in the Airline Industry: A Comparison Between Legacy Airlines and Low-Cost Airlines," *American Journal of Tourism Research*, 2(1), p. 67-77.
- Chang, Yu-Chun (2013), "Factors Affecting Airport Access Mode-choice for Elderly Air Passengers," [\*Transportation Research Part E: Logistics and Transportation Review\*](#), 57, p.105-112.
- Curtis, T., Rhoades, D. L., & Waguespack, B. P. (2012). "Satisfaction with Airline Service Quality: Familiarity Breeds Contempt", *International Journal of Aviation Management*, 1(4).
- Fodness, D. and Murray, B. (2007), "Passengers' expectations of airport service quality", *Journal of Services Marketing*, Vol. 21 Issue: 7, pp.492-506.
- Fritz, M. and Berger, P. D. (2015), "Improving the User Experience Through Practical Data Analysis," Morgan-Kaufmann-Elsevier, Waltham, MA, p. 259-260.
- Harvey, Greig (1987), "Airport Choice in a Multiple Airport Region," *Transportation Research Part A: General*, 21(6), p. 439-449.
- Hess, Stephane, and John W. Polak (2005), "Mixed Logit Modeling Of Airport Choice In Multi-Airport Regions," [\*Journal of Air Transport Management\*](#), 11(2), p. 59-68.

- Ishii, Jun, [Sunyoung, Jun](#), and [Kurt Van Dender](#) (2009), "Air Travel Choices In Multi-Airport Markets," *Journal of Urban Economics*, 65(2), p. 216-227.
- Kamarudin, R. (2015) "Managing Customer Expectation for Passenger Service at Airport", *ACI World Report*.
- Mattozo, Teofilo C., Silva, Gutemberg S., Costa, Jose A., and Fernandes-Neto, Andre P. (2012), "Logistic Regression Applied to Airport Customer Satisfaction Using Hierarchical Quality Model," *Intelligent Data Engineering and Automated Learning - IDEAL 2012*, Volume 74- 35, p. 558-567.
- Pels, E., Nijkamp, P. and RietveldI, P. (2001), "Airport and airline choice in a multiple airport region: an empirical analysis for the San Francisco Bay area", *Reg. Studies* 35, 1–9.
- Pels, E., Nijkamp, P. and RietveldI, P. (2003), "Access to and competition between airports: a case study for the San Francisco Bay area," *Transportation Research Part A: Policy and Practice*, 37(1), p. 71-83.
- Suki, Norazah, M. (2014), "Passenger Satisfaction With Airline Service Quality in Malaysia: A Structural Equation Modeling Approach," *Research in Transportation Business and Management*, 10(1), p. 26-32.
- San Francisco International Airport Customer Survey Data <https://www.flysfo.com/media/customer-survey-data>, Accessed Nov 22, 2017.
- Wang, Xiao, Mengnan Hong, & Paul D. Berger (2015), " Customer-Satisfaction Analysis At San Francisco International Airport," *International journal of Management Studies*, 2(1), p. 1-12.
- Wang, Xiao, Mengnan Hong, & Paul D. Berger (2016),"Determining Key Factors in Consumer Evaluation of an Airport, *Journal of Marketing Management*, 4(1), p. 1-14.