An Investigation of the Effects of Carrier Groups on Airline Quality Rating Components Using a Two-way Analysis of Variance

John H. Mott and Branden K. Avery
Purdue University

Abstract

The Airline Quality Rating is a quantitative determination of the quality of U.S. domestic air carriers based on parameters published by the U.S. Department of Transportation. The rating is unique in that it is of interval scale and is comparable across carriers and time periods (Bowen & Headley, 2012). In order to gain insight into the relationship of the AQR metric to the carriers to which it is applied, it is helpful to group the carriers and examine the effects of those groupings on the four individual factors that comprise the AQR. Such a methodology allows one to better understand the relationship of the AQR metric to each of the carrier groups and ultimately improve the predictability of the metric. The authors employ a two-way analysis of variance to determine differences between carrier group means for each of the four AQR factors while examining longitudinal effects, along with post-hoc difference testing. While the post-hoc test results indicate significant differences between some groupings, suggesting that separate econometric models for those groupings might be created with the goal of more accurately forecasting the metric, some of the assumptions upon which the ANOVA is predicated are violated. This article will examine those violations and suggest that further research using nontraditional methods (e.g., Bayesian analysis) is indicated.

Background

The Bureau of Transportation Statistics, a unit of the Research and Innovative Transportation Administration, which is itself an agency of the U. S. Department of Transportation, has collected operational performance data from U. S. domestic air carriers since 1987 (USDOT, 2013). Among this data are key metrics consisting of on-time arrival, denied boarding, mishandled baggage, and customer complaint frequencies. These metrics, which constitute a primary portion of the dataset, can have critical effects upon both airline customer loyalty and profitability.

The Airline Quality Rating, an annual analysis that provides an overall ranking of quality of service among air carriers with at least 1% of the domestic passenger volume in a given year, combines these four primary reporting metrics, which themselves consist of a total of 15 elements, into a formula that provides a consistent measure, thereby allowing comparisons both longitudinally and across carriers (Bowen and Headley, 1991).

The potential for predicting the AQR rankings using an econometric model seems considerable. Such predictions could be utilized by managers to better allocate resources

in an effort to improve the metrics that comprise the overall measure. For example, if certain economic conditions correlate with an increase in denied boardings, company revenue management policies related to no-show passengers (i.e., overbooking percentages for specific flights) could be adjusted in such a manner as to mitigate denied boarding consequences. This would presumably result in an increase in the perception of quality of the particular carrier among passengers.

It is reasonable to assume that AQR comparisons across carriers could be facilitated by partitioning those carriers into distinct service groups, as this would tend to increase correlations of the independent predictor variables with the overall AQR index. This assumption is intuitive, in that it is clear that carriers providing different levels of service utilize different operational procedures, while there is a high degree of procedural similarity among carriers providing service at the same level. This study attempts to determine, with the goal of econometric model development in mind, whether this assumption is correct; that is, whether separate econometric models are appropriate for different carrier groups.

A challenge in making longitudinal comparisons across the AQR dataset is that incumbent carriers are occasionally dropped from the rankings, either because they have merged with another carrier or because they fail to meet the required 1% domestic volume criterion for inclusion in the AQR. The carriers used in this study were selected in order to maintain consistency across the study period, as noted below.

Literature Review

The Airline Deregulation Act (P.L. 95-504) of 1978 did more than allow market forces to set prices and carriers to select routes based on traveler demand; deregulation also permitted the determination of the level of service quality provided to customers (Tiernan, Rhoades, & Waguespack, 2008). After the passage of the Act, the U. S. government began to collect various metrics of airline service quality; however, since these records were not publicly available, data-driven understanding of competitors' service and performance could not be used as a competitive advantage by carriers in order to gain customers. Airline passengers had little knowledge about which airlines performed better than their competitors or which provided a higher quality of service. Passenger knowledge consisted solely of personal experience or the latest information reported in the news media (Rhoades & Waguespack, 2008).

The publication of the first Air Travel Consumer Report (ATCR) in 1987 resulted in a significant impact on commercial air transportation as a result of increased awareness of carrier performance by the traveling public. The ATCR is a monthly product of the U. S. Department of Transportation that is "designed to assist consumers with information on the quality of services provided by the airlines" (USDOT, 2011, pg. 2). The ATCR allowed the public to view for the first time reported data for on-time performance, mishandled baggage, denied boarding, and customer complaints.

The ATCR has served as a base for researchers to explore quality in the U. S. airline industry. One group of researchers investigating quality in the U. S. airline industry includes Bowen and Headley, who published the first Airline Quality Rating (AQR) in 1991 (Bowen & Headley, 2012). The Airline Quality Rating (AQR) has for the past 23 years ranked U. S. air carriers that account for at least 1% of the domestic passenger volume. The AQR provides a month-by-month measure of quality using a weighted average of metrics representing on-time arrivals (OT), involuntary denied boardings (DB), mishandled baggage (MB), and a combination of 12 customer complaint categories (CC). The 12 customer complaint categories are flight problems; oversales; reservations, ticketing, and boarding; fares; refunds; baggage; customer service; disability; advertising; discrimination; animals; and other (Bowen & Headley, 2012).

The AQR quality measure is determined as follows:

$$Q = ((8.63 \times OT) + (-8.03 \times DB) + (-7.92 \times MB) + (-7.17 \times CC)) / ((8.63 + 8.03 + 7.92 + 7.17))$$

where on-time arrivals are reported monthly, involuntary denied boardings are reported quarterly per 10,000 passengers, mishandled baggage is reported monthly per 1,000 passengers, and customer complaints are reported monthly per 100,000 passengers (Bowen & Headley, 2012).

The original AQR quality metric was extended to include statistical process control concepts in an effort to provide an additional tool that industry managers could use to "monitor quality, identify problems, and provide timely feedback on the effectiveness of tactics to improve quality." (Bowen, Headley, & Lutte, 1993, p. 38). Headley and Bowen (1997) also offered development considerations for facilitating the adaptation of the AQR's weighted average approach to the international airline industry.

More recently, a second group of researchers at Embry-Riddle Aeronautical University employed the ATCR in a somewhat different manner. Waguespack and Rhoades separated safety, service, and financial performance and simply normalized safety and service data by total departures. Separating these metrics allowed the researchers to examine service quality and safety independently and explore the relationship between the parameters. Their Service Disquality Index (SDI) has been used both to provide a 20-year perspective on service quality performance at the major U. S. carriers (Rhoades & Waguespack, 2008), and to measure service quality issues between carriers in different industry segments (Rhoades & Waguespack, 2000a, 2000b).

Research Methodology

This study was conducted using the four initial metrics that were part of the original ATCR and which comprise the AQR measure (OT, DB, MB, and CC). The data spans a six-year period from 2006 to 2011 (Bowen & Headley, 2007 – 2012). As mentioned

previously, longitudinal studies using AQR data are somewhat difficult to conduct because of a lack of a consistent carrier base over the period. Because of these inconsistencies, the present research focused on the fourteen airlines that have been included in the annual AQR reports over the period. These fourteen carriers have been classified according to their business models into one of three service groups: legacy (or network) carriers, regional carriers, and low-cost carriers. The fourteen airlines and their groupings are presented in Table 1.

Table 1
Carrier Groupings

Legacy Carriers	Regional Carriers	Low-Cost Carriers
Alaska Airlines	American Eagle Airlines	Air Tran Airways
American Airlines	Atlantic Southeast Airlines	Frontier Airlines
Continental Airlines	Mesa Airlines	Jet Blue Airlines
Delta Airlines	Sky West Airlines	Southwest Airlines
United Airlines		
US Airways		

Legacy carriers operate large route networks primarily using a hub-and-spoke route model. Their networks include international destinations and medium-to-large domestic cities (Erstad, Jednachowski, Bowen, Meehan, & Bowen, 2013). In addition, legacy carriers operate diverse fleets of aircraft with approximate capacities of from 100 to over 300 passengers.

Regional carriers often operate under code-sharing agreements with legacy carriers. They operate flights primarily from smaller cities to their respective partners' hubs using smaller, more efficient aircraft carrying approximately 9 to 99 passengers (Forbes & Lederman, 2006). Regional carriers have the most dynamic scheduling and overhead cost models, due to their need to support service from their hubs to medium and large cities by their respective legacy carriers. This service is intended to complement the legacy carrier's routes by adding additional frequency during off-peak times to ensure higher load factors.

Low-cost carriers often employ both point-to-point and hub-and-spoke models. They typically serve medium-to-large cities. These carriers generally have lower load factors when compared with legacy carriers, and therefore have more dynamic scheduling needs as the passenger demand changes (Erstad, Jednachowski, Bowen, Meehan, & Bowen,

2013). Low-cost carriers operate larger aircraft equipped with approximately 125 to 175 seats.

As a classical relative-frequency approach to the research question was desired, a repeated measures analysis of variance (ANOVA) of the data was conducted using IBM SPSS statistical software. Such an analysis is used to determine whether there are any statistically significant differences between the population means of three or more related groups. The REPEATED statement in the SPSS General Linear Model provides automatic computation and analyses for several common choices of contrast variables (Lund & Lund, 2013). The data must be in multivariate form in order to perform this analysis.

Each of the four metrics was treated as a separate dependent variable, and the independent variables were the carrier service level groups and the year groupings. While the use of a one-way ANCOVA (versus the two-way repeated measures ANOVA) would allow the control for variance in the year groupings, since the primary interest in this study is the difference in the carrier service level group means, the two-way repeated measures ANOVA was chosen due to potential correlation between the independent variables that can create difficulty in interpretation of the results.

One studying the carrier groupings in Table 1 will realize that the proposed use of 14 carriers (with two additional legacy carriers) leads to an unbalanced analysis, which presents considerable complexity in the two-factor design, as the orthogonality property of main effects does not carry over to the unbalanced case, meaning that the F-ratios are dependent on the order in which the sources of variation are considered (Shaw & Mitchell-Olds, 1993). While this can be corrected using the Type III sum-of-squares in SPSS, such a correction can lead to biased results, as main effects may be distorted in the presence of significant interactions.

The hypotheses that were used for this study are as follows:

- H_o: There is no statistically significant difference between the means at the given α level.
- H_a: There is a statistically significant difference between the means at the given α level.

A significance level of α = 0.05 was applied in all tests.

Results

Four multivariate repeated measures analyses of variance were conducted to determine whether there were statistically significant differences between carrier group means for each of the four AQR factors (on-time arrivals, mishandled baggage, denied boardings, and customer complaints) over the course of six years. While the research

data included few, if any, outliers, a Shapiro-Wilk test (p < .05) indicated that the data was not normally distributed for any of the carrier groups. Fortunately, the repeated measures ANOVA is reasonably robust to violations of normality, meaning that minor violations of the normality assumption will still provide valid results. (Lund & Lund, 2013).

First, on-time performance was analyzed. A test of between-subjects effects for the carrier service level variable indicated a significant difference between the service levels, F (2,165) = 3.846, p = .023. A Tukey-Kramer post-hoc difference test indicated significant differences between the regional and low-cost carrier groups (p = .032). Levene's test of homoscedasticity indicates that the between-subjects variances are homogeneous for five out of the six years; 2006 was the only year in which significant heteroscedasticity was indicated (Table 2).

Mauchly's Test of Sphericity, χ^2 (14) = 38.67, p = .001, indicated that the data violated the sphericity assumption. As a result, a Huynh-Feldt correction was applied (ε = .961). With the correction, statistically significant changes in on-time arrivals were indicated over the six years, F (4.803, 792.575) = 60.027, p < .0005, partial η^2 = .267. The means and standard deviations over the six year period can be found below in Table 3.

Table 2
Levene's Test of Equality of Error Variances for On-Time Arrival

Year	F	df1	df2	p
2006	5.099	2	165	.007
2007	1.853	2	165	.160
2008	.135	2	165	.874
2009	1.955	2	165	.145
2010	.032	2	165	.968
2011	.936	2	165	.394

Secondly, mishandled baggage was analyzed. There was a significant difference between carrier service levels suggested by the between-subject effects test, F (2,165) = 190.984, p = .001. All six years were significant in the Levene's test, suggesting that the between-subjects variances are not homogeneous (Table 4). Tukey-Kramer post-hoc testing indicated significant differences between the legacy and regional (p = .001), legacy and low-cost (p = .001), and regional and low-cost carrier groups (p = .001).

Mauchly's Test of Sphericity, χ^2 (14) = 414.23, p = .001, suggested a sphericity violation. Because of this, a Greenhouse-Geisser correction was applied (ε = .549). With this correction, statistically significant changes in mishandled baggage were indicated

over the six years, $F (2.743, 452.617) = 354.635$, $p < .0005$, partial $\eta^2 = .682$. The means and standard deviations are presented in Table 5.

Table 3
Descriptive Statistics for On-Time Arrival

Year Carrier Group	Mean	Std. Deviation	N
2006	.7488	.04093	72
Legacy	.7197	.06512	48
Regional	.7707	.05831	48
Low-cost	.7468	.05691	168
Total			
2007	.7198	.06863	72
Legacy	.7086	.08174	48
Regional	.7620	.07811	48
Low-cost	.7287	.07790	168
Total			
2008	.7515	.07486	72
Legacy	.7486	.07911	48
Regional	.7736	.07995	48
Low-cost	.7570	.07782	168
Total			
2009	.7995	.05254	72
Legacy	.7751	.06608	48
Regional	.7857	.06387	48
Low-cost	.7886	.06047	168
Total			
2010	.8237	.05390	72
Legacy	.7968	.05578	48
Regional	.7989	.05954	48
Low-cost	.8089	.05721	168
Total			
2011	.8092	.05537	72
Legacy	.7873	.06154	48
Regional	.7907	.06683	48
Low-cost	.7976	.06106	168
Total			

.

Table 4

Levene's Test of Equality of Error Variances for Mishandled Baggage

Year	F	df1	df2	p
2006	28.273	2	165	<.001
2007	14.401	2	165	<.001
2008	16.588	2	165	<.001
2009	23.595	2	165	<.001
2010	26.078	2	165	<.001
2011	22.246	2	165	<.001

Next, denied boardings were analyzed. A test of between-subjects effects suggested a significant difference between the carrier service levels, $F(2,165) = 83.231$, $p = .001$. The Tukey-Kramer post-hoc test again revealed significant differences between the legacy and regional ($p = .001$), legacy and low-cost ($p = .001$), and regional and low-cost carrier groups ($p = .001$). Levene's test also indicated between-subjects heteroscedasticity for all six years of the study (Table 6).

Mauchly's Sphericity Test, $\chi^2(14) = 502.40$, $p = .001$, again indicated that the data violated the sphericity assumption. Therefore, a Greenhouse-Geisser correction was applied ($\varepsilon = .391$). With this correction, statistically significant changes in denied boardings over the six years were indicated, $F(1.957, 322.961) = 6.961$, $p < .0005$, partial $\eta^2 = .040$. Table 7 summarizes the relevant means and standard deviations.

Lastly, customer complaints were analyzed. A test of between-subjects effects for the carrier service level variable indicated a significant difference between the service levels, $F(2,165) = 52.983$, $p = .001$. Levene's test indicates a higher degree of homogeneity of between-subjects variances in four out of the six years of the independent variables; the two years that are significant in the test are 2007 and 2009. Tukey-Kramer post-hoc test indicated significant differences between the legacy and regional ($p = .001$), legacy and low-cost ($p = .001$), and regional and low-cost carrier groups ($p = .001$).

Mauchly's Test of Sphericity, $\chi^2(14) = 107.52$, $p = .001$, once again indicated that the data violated the sphericity assumption. As a result, a Huynh-Feldt correction was applied ($\varepsilon = .820$). With the correction, statistically significant changes in customer complaints were indicated over the six years, $F(4.099, 676.397) = 14.326$, $p < .0005$, partial $\eta^2 = .080$. The means and standard deviations over the six year period are displayed in Table 9.

Table 5
Descriptive Statistics for Mishandled Baggage

Year Group	Carrier	Mean	Std. Deviation	N
2006	Legacy	6.1697	1.73837	72
	Regional	13.2077	4.03767	48
	Low-cost	4.8398	1.28341	48
	Total	7.8006	4.29026	168
2007	Legacy	6.7678	1.84949	72
	Regional	11.5696	3.12272	48
	Low-cost	5.3283	1.65751	48
	Total	7.7285	3.35653	168
2008	Legacy	5.0108	1.49106	72
	Regional	8.8425	2.81798	48
	Low-cost	3.8394	1.28880	48
	Total	5.7709	2.77142	168
2009	Legacy	3.8288	1.11966	72
	Regional	6.5075	2.19484	48
	Low-cost	2.5379	.77098	48
	Total	4.2253	2.10695	168
2010	Legacy	3.1832	.79181	72
	Regional	5.6942	1.93822	48
	Low-cost	2.5392	.77069	48
	Total	3.7166	1.77093	168
2011	Legacy	3.0747	.59611	72
	Regional	5.4798	1.58115	48
	Low-cost	2.4242	.80225	48
	Total	3.5643	1.60600	168

Table 6

Levene's Test of Equality of Error Variance for Denied Boarding

Year	F	df1	df2	p
2006	29.911	2	165	<.001
2007	23.217	2	165	<.001
2008	57.505	2	165	<.001
2009	40.678	2	165	<.001
2010	29.736	2	165	<.001
2011	41.174	2	165	<.001

Table 7
Descriptive Statistics for Denied Boarding

Year	Carrier Group	Mean	Std. Deviation	N
2006		1.1862	.60402	72
Legacy		2.1313	1.59763	48
	Regional	.3869	.38905	48
	Low-cost	1.2279	1.16390	168
	Total			
2007		1.2217	.77840	72
Legacy		2.2844	1.42382	48
	Regional	.5606	.52627	48
	Low-cost	1.3364	1.15890	168
	Total			
2008		1.1442	.44564	72
Legacy		2.3781	1.22177	48
	Regional	.5850	.51981	48
	Low-cost	1.3370	1.03475	168
	Total			
2009		1.3392	.57502	72
Legacy		2.1375	1.31447	48
	Regional	.8406	.82083	48
	Low-cost	1.4248	1.03137	168
	Total			
2010		1.2092	.67706	72
Legacy		1.9706	1.52391	48
	Regional	.9756	1.03612	48
	Low-cost	1.3600	1.14458	168
	Total			
2011		.9171	.38187	72
Legacy		1.5394	.88332	48
	Regional	.5550	.37635	48
	Low-cost	.9914	.68182	168
	Total			

Table 8

Levene's Test of Equality of Error Variance for Customer Complaints

Year	F	df1	df2	p
2006	3.549	2	165	.031
2007	11.553	2	165	<.001
2008	2.533	2	165	.083
2009	10.267	2	165	<.001
2010	4.255	2	165	.016
2011	5.453	2	165	.005

Table 9
Descriptive Statistics for Customer Complaints

Year Group	Carrier	Mean	Std. Deviation	N
2006	Legacy	1.0406	.36774	72
	Regional	.9283	.52879	48
	Low-cost	.4196	.24783	48
	Total	.8311	.47305	168
2007	Legacy	1.7551	.97122	72
	Regional	.9631	.52858	48
	Low-cost	.6306	.40791	48
	Total	1.2075	.87669	168
2008	Legacy	1.4246	.63629	72
	Regional	.7910	.49795	48
	Low-cost	.7665	.52394	48
	Total	1.0555	.64963	168
2009	Legacy	1.1926	.52479	72
	Regional	.6121	.22220	48
	Low-cost	.7306	.39514	48
	Total	.8948	.49386	168
2010	Legacy	1.4137	.58790	72
	Regional	.6752	.37561	48
	Low-cost	.9179	.56351	48
	Total	1.0611	.61508	168
2011	Legacy	1.5103	.67401	72
	Regional	.9056	.55948	48
	Low-cost	.7217	.39974	48
	Total	1.1122	.67129	168

Discussion

There are a number of assumptions associated with the repeated measures analysis of variance that must be made by the researcher. Requirements that the dependent variable be continuous and the independent variables be categorical in nature are clearly met by the data being analyzed. Further assumptions that the differences between groups be free

of outliers and normally distributed and that the differences between all combinations of groups be homoscedastistic are met to varying degrees, as described below.

Normality of the differences in the dependent variable between groups can be evaluated using Q-Q plots and the Shapiro-Wilk test (Kinney, 2002). In this study, both the Shapiro-Wilk test ($p < .05$) and the Q-Q plots indicated that the data was not normally distributed for any of the carrier groups. Fortunately, as noted previously, repeated measures ANOVA is reasonably tolerant of violations of normality (Lund & Lund, 2013). Figure 1 below provides one example of a Q-Q plot that shows evidence of the data being not normally distributed.
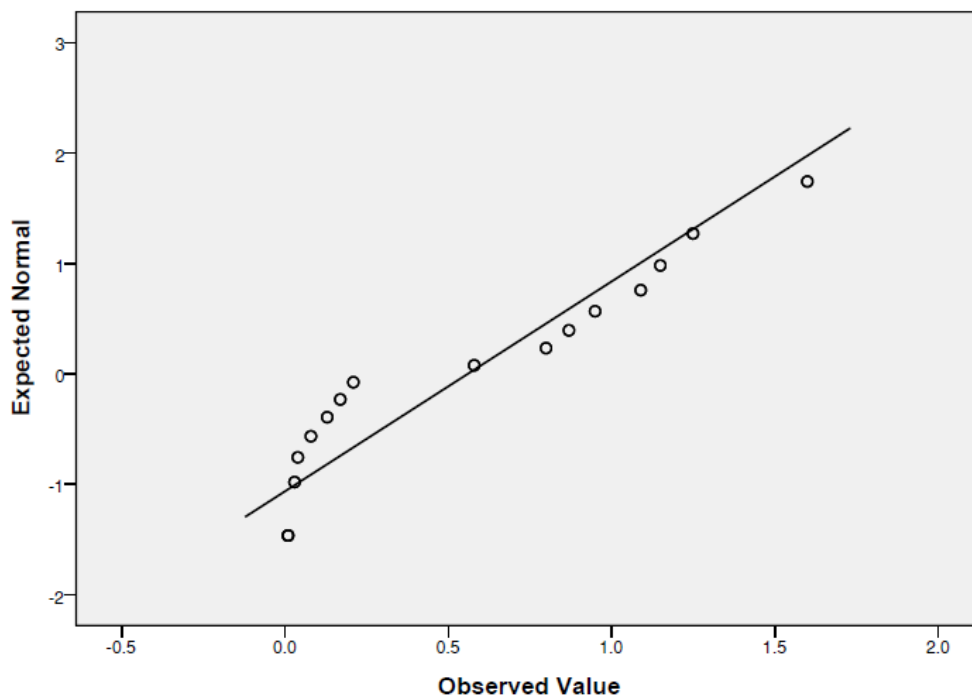
Figure 1. Normal Q-Q plot of low-cost carriers for 2007.

A Levene's test was conducted to determine the level of between-subjects homogeneity of variance for each of the four AQR components over the six year study period. This testing indicated varying degrees of between-subjects homoscedasticity for each of the components. Likewise, the assumption of sphericity was examined using Mauchly's test and found to be violated for each of the four components. While the repeated measures analysis of variance is much more susceptible to sphericity violations than it is to lack of normality, there are corrections that may be employed in such cases to produce a more valid critical F-value; among these are the Greenhouse-Geisser and the Huynh-Feldt corrections (Lund & Lund, 2013). Nonetheless, due to the lack of both

homoscedasticity and normality among the carrier groups, other statistical methods that do not require such assumptions are worthy of consideration for the analysis of this dataset.

Given that the analysis is reasonably tolerant to the lack of normality and using the appropriate corrections for lack of sphericity, the researchers proceeded to examine the differences between legacy, regional, and low-cost carriers to determine whether a significant difference between the groups exists relative to the four primary AQR components (OT, DB, MB, CC) over the six year study period. Based on the post-hoc test results, significant differences in on-time arrivals were not indicated between legacy and regional carriers or legacy and low-cost carriers; however, such differences were indicated between regional and low-cost carriers. It should be noted that the resulting p-value ($p = .064$) for the legacy-regional post-hoc comparison was quite close to the fixed α value of .05; therefore, further research using Bayesian methods is suggested.

With regard to mishandled baggage, denied boardings, and customer complaints, significant differences between legacy, regional, and low-cost carriers were indicated, implying rejection of the null hypothesis, H_o. This suggests that there is sufficient evidence to support the premise that separate econometric predictive models are needed by airline managers to facilitate quality of service improvements.

Managers within the different carrier groups will be able to utilize predictive modeling to better forecast the AQR components. In addition, the results presented herein also describe longitudinal changes in the marginal means of each of the components (Figures 2 through 5). For example, Figure 2 indicates that significant improvement has occurred among the legacy carriers over the study period with regard to on-time arrivals, while a lesser degree of improvement has occurred among low-cost carriers. The longitudinal changes in the slopes of the marginal means and the attendant interactions imply the difficulty of constructing accurate predictive models for these components; that difficulty is clearly compounded by the failure to partition the carriers into groups.

Conclusion and Further Research

In this Frequentist study of the four primary components of the Airline Quality Rating over a six-year period, two-way analyses of variance in conjunction with Tukey-Kramer post-hoc difference testing have shown that separate predictive models are appropriate, based on the differences between the carrier groups. Utilizing these predictive models, airline managers will be able to more accurately forecast the components and thus the overall quality rating, thereby allowing them to refine resource allocation methods in an effort to improve quality of service. It is suggested that future research be conducted using Bayesian statistical methods rather than the null hypothesis significance testing (NHST) methods used in this study. Bayesian models do not depend on corrections to ensure that test assumptions are met; instead, Bayesian methods rationally mitigate statistical error based on the data itself (Kruschke, 2010). In addition, when using

Bayesian hierarchical modeling, any concern regarding the robustness of the analysis technique is reduced.
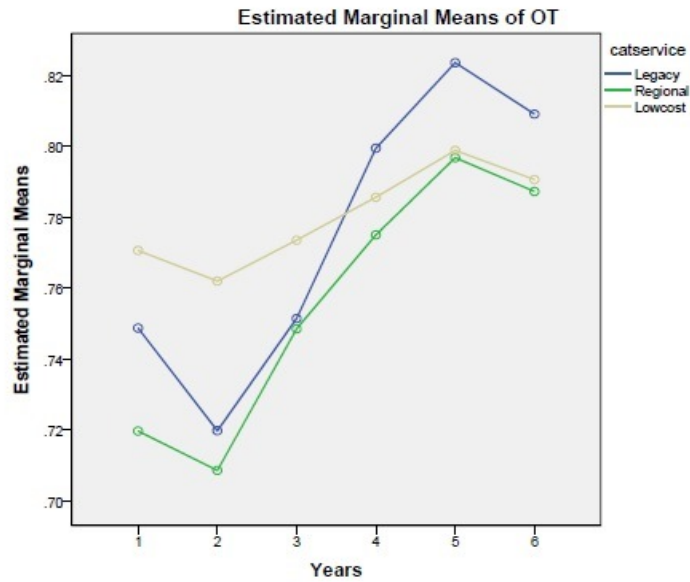

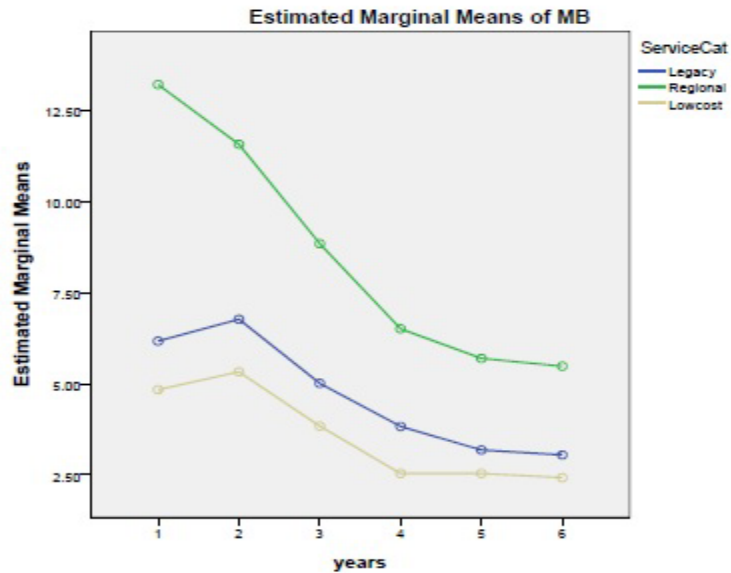*Figure 2.* Estimated marginal on-time arrival means plotted against study year.

Figure 3. Estimated marginal mishandled baggage means plotted against study year.
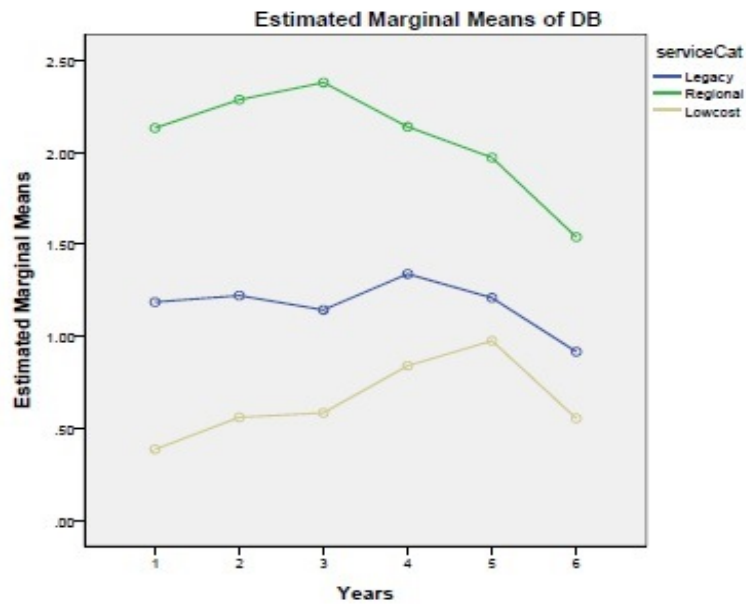
Figure 4. Estimated marginal denied boarding means plotted against study year.
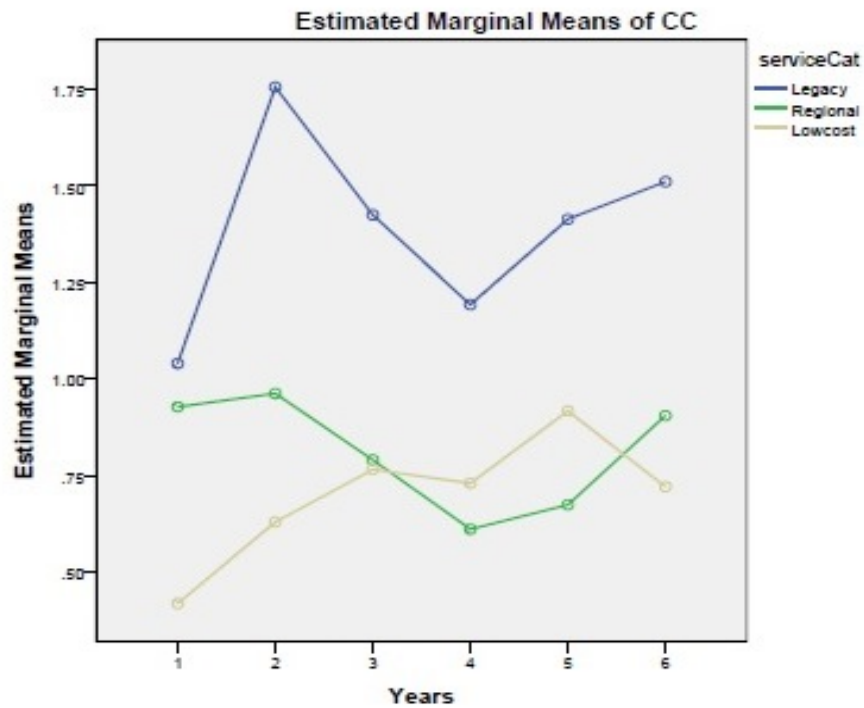

Figure 5. Estimated marginal customer complaint means plotted against study year.

References

Bowen, B. D., & Headley, D. E. (2012). *Airline quality rating 2012* (Rep.). West Lafayette, IN: Purdue e-Pubs.

Bowen, B. D., & Headley, D. E. (2011). *Airline quality rating 2011* (Rep.). West Lafayette, IN: Purdue e-Pubs.

Bowen, B. D., & Headley, D. E. (2010). *Airline quality rating 2010* (Rep.). West Lafayette, IN: Purdue e-Pubs.

Bowen, B. D., & Headley, D. E. (2009). *Airline quality rating 2009* (Rep.). West Lafayette, IN: Purdue e-Pubs.

Bowen, B. D., & Headley, D. E. (2008). *Airline quality rating 2008* (Rep.). West Lafayette, IN: Purdue e-Pubs.

Bowen, B. D., & Headley, D. E. (2007). *Airline quality rating 2007* (Rep.). West Lafayette, IN: Purdue e-Pubs.

Bowen, B. D., Headley, D. E., & Luedtke, J. R. (1991). *Airline quality rating 1991* (Rep.). West Lafayette, IN: Purdue e-Pubs.

Bowen, B. D., Headley, D. E., & Lutte, R. K. (1993). The airline quality rating: Developing an industry standard. *Journal of Aviation/Aerospace Education & Research, 4*(1), 33-39. Retrieved from http://commons.erau.edu/jaaer/vol4/iss1/8

Erstad, A., Jednachowski, M., Bowen, B. D., Meehan, R., & Bowen, E. (2013). *Modeling organizational performance: Differentiating factors in economic periods prior, during, and post recession*. Retrieved from http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1017&context=atgrads

Forbes, S. J., & Lederman, M. (2007). The role of regional airlines in the US airline industry. *Advances in Airline Economics*, *2*, 193-208.

Headley, D. E., & Bowen, B. D. (1997). International airline quality measurement. *Journal of Air Transportation World Wide, 2*(1), 55-63.

Kinney, J. J. (2002). *Statistics for science and engineering*. Boston: Addison Wesley.

Kruschke, J. K. (2010). An open letter to editors of journals, chairs of departments, directors of funding programs, directors of graduate training, reviewers of grants and manuscripts, researchers, teachers, and students. Retrieved from http://www.indiana.edu/~kruschke/AnOpenLetter.htm

Lund, A., & Lund, M. (2013). Repeated measures ANOVA. Retrieved from
      https://statistics.laerd.com/statistical-guides/repeated-measures-anova-statistical-
      guide.php

Rhoades, D.L. & Waguespack, B. (2000a). Service quality in the U.S. airline industry:
      Variations in performance within airlines and between airlines and the industry.
      Journal of Air Transportation World Wide, 5(1), 60-77.

Rhoades, D.L. & Waguespack, B. (2000b). Judging a book by its cover: The relationship
      between service and safety quality in US national and regional airlines. *Journal of
      Air Transport Management, 6,* 87-94.

Rhoades, D. L., & Waguespack, B. (2008). Twenty years of service quality performance
      in the US airline industry. *Managing Service Quality*, *18*(1), 20-33. doi:
      10.1108/09604520810842821

Shaw, R. G., & Mitchell-Olds, T. (1993). ANOVA for unbalanced data: An overview.
      Ecology, 74(6), 1638-1645.

Tiernan, S., Rhoades, D.L., & Waguespack Jr. B. (2008). Airline service quality:
      Exploratory analysis of consumer perceptions and operational performance in the
      USA and EU. *Managing Service Quality, 18,* 212-224. doi:
      10.1108/09604520810871847

U. S. Department of Transportation, Office of Aviation Enforcement and Proceedings,
      Aviation Consumer Protection Division. (2011). Air travel consumer report.
      Retrieved from
      http://airconsumer.ost.dot.gov/reports/2011/March/2011MarchATCR.pdf

U. S. Department of Transportation, Bureau of Transportation Statistics. (2013). Airline
      on-time performance and causes of flight delays. Retrieved from
      http://apps.bts.gov/help/aviation/#q9