The Plight of the Late Flight

An Analysis of Delay Improvement over 25 Years

Prepared for ST599 - Big Data Analysis by:
Andrew Bernath
Heather Kitada
Ethan Edwards

Due date: 5/10/2014

Introduction and Motivation

Numb with frustration, you glare at the clock for the 30th time. A longing glance at the attendant provides no consolation; they can't usher you onto a plane that hasn't arrived yet. *Oh well,* you think, *back to Angry Birds*. Thank goodness for smartphones. Few can boast that flight delays have not affected them in some way and therefore, it is worthwhile to ask whether the air travel delay for any particular airline is increasing or decreasing over time. The purpose of this report is to address this question through analysis of 25 years of flight data from the Bureau of Transportation Statistics (BTS). Additionally, this report will explore whether sampling flights from airlines stratified by airport size provides a reasonable estimate of mean delay.

Methods Summary

For a traveler, a natural choice for a measure of total flight delay is the arrival delay, defined as the difference between the scheduled and actual arrival times. This delay is of foremost interest to a traveler as it corresponds with their ultimate goal of arriving at their destination on time. The arrival delay also has the advantage that a flight which arrives at or earlier than its scheduled time would be considered on time or early, regardless of a late departure.

The BTS data represents the entire population of flights in the United States between 1987 and 2013, so that population means and standard deviations are directly calculable. Additionally, a sampling scheme was developed to address the assumption that flights originating from airports of similar traffic volume would share delay patterns. In addition to partitioning by carrier and year, data were further stratified into three categories. Airports with "small," medium," and "large" flight volumes were defined by ordering the average number of flights from every airport over the last 25 years and dividing them into thirds for the aforementioned categories. Strata were sampled in a manner that was proportional to the number of flights. Sample means and variances were weighted and finite population correction was taken into account to appropriately create the desired population estimates.

Using population and sample means, the improvement in delay from year x to year y is given as the change in mean arrival delay, $\mu_y - \mu_x$. A positive value corresponds to an increase in mean delay while negative values correspond to a decrease in mean delay The median change in mean arrival delay for each airline was the primary metric used to assess a carrier's improvement in arrival delay.

Results and Findings

This section presents findings for the year-to-year changes in mean arrival delays of carriers for which BTS has records of at least 10 consecutive years. Tables 1 and 2 (below, *right*) list the median year-to-year changes for population and sample data, respectively. Figures 1 and 2 (*below*) display the changes in mean delay over the 25 years of BTS data and provide a visual comparison of the population parameters with the sample estimates. Points lying above the reference line represent increases in mean delay from the previous year, while points lying below the reference line refer to decreases in delay from the previous year. Note that, from Tables 1 and 2, AirTran (FL) and American Eagle (MQ) have the best improvement in mean delay for the population and sample means, respectively.

Table 1: Median Change in Delay - Population

Carrier	Median
FL	-0.563
NW	-0.316
TW	-0.208
HP	-0.207
US	-0.147
HA	-0.143
CO	-0.128
MQ	-0.073
UA	-0.054
WN	-0.052
AS	0.059
DL	0.421
AA	0.561
OO	0.640
B6	0.986
EV	2.636

Table 2: Median Change in Delay - Sample

Carrier	Median
MQ	-1.688
TW	-1.212
CO	-1.155
FL	-1.073
DL	-0.679
AA	-0.595
AS	-0.463
NW	-0.218
B6	-0.178
WN	-0.145
UA	-0.066
US	0.171
HP	0.702
HA	0.864
OO	1.478
EV	2.878

To represent the performance of the sampling scheme, a frequency histogram (*Figure 3*, *below)* displays the bias of our sample estimates of the population means per airline, per year. Though slightly right skewed, the center of the histogram appears to be centered close to 0 with a grand mean estimate of 0.335 minutes delay. This suggests that the sampling scheme was effective for estimating population parameters.

Fig. 1: Change in Mean Delay - AirTran

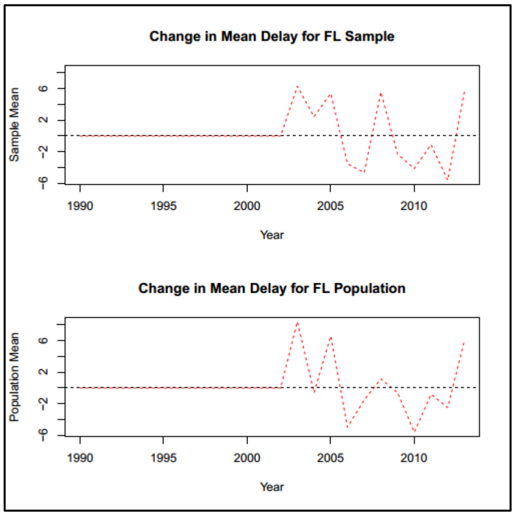
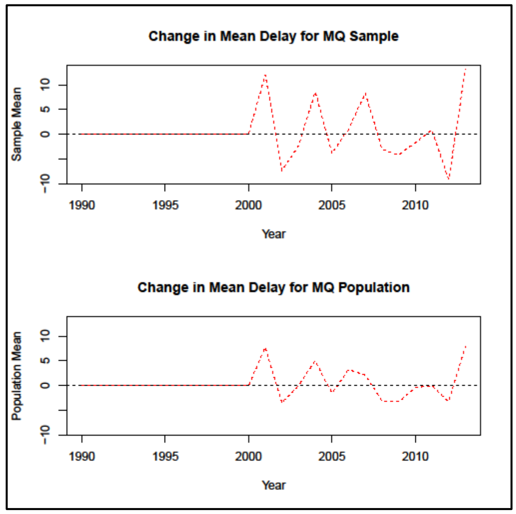


Fig. 2: Change in Mean Delay – American Eagle



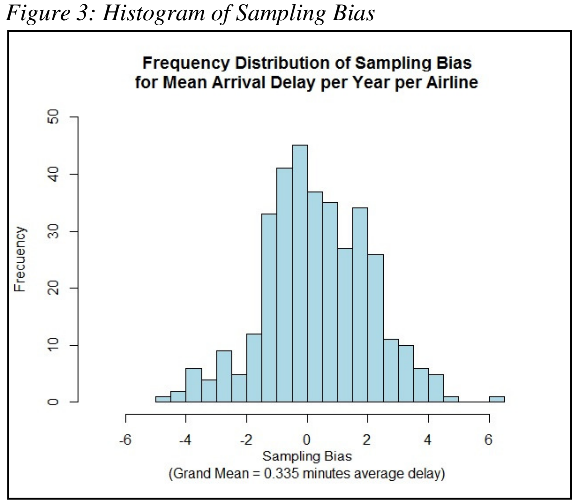
Obstacles

Determination of a metric with which to characterize air travel delay was not a trivial task. Several alternatives were considered but ultimately, inconsistencies in the data prevented most metrics metric from being useful. Delays other than arrival and departure are missing for years prior to 2003, possibly due to changes in monitoring and reporting. Several airlines had very few years of data and one airline (Pacific Southwest Airlines) had no data due to purchase by USAir in 1987.[1] Additionally, from 2007 to 2008, there is a substantial jump (upwards of a 5-6 fold increase) in several delay categories. It is likely that this was a result of several airlines making substantial changes in flight schedules due to rising fuel prices and the onset of the economic downturn.[2]

Several challenges in the sampling scheme stemmed from the lengthy computing time necessary to collect a representative sample. Initially, all airports were intended to be proportionally sampled but this required an unreasonable computing time. Even after revision to three size categories, sampling took several hours to complete. Surprisingly, the sampling computations took considerably longer to execute than those of the population.

Applications and Future Research

Figure 3: Histogram of Sampling Bias



The data used in this study is part of a larger BTS flight data set, in which there is much potential for improvement on these results. Relaxing the assumption that arrival delay best characterizes total delay would allow us to understand how various sources of delay (such as weather, late aircraft, National Air System, etc.) affect travel and how they are changing through time. Determination of a measure of delay from sources that the carrier has direct control over would be useful both for carriers seeking to improve services and passengers choosing a reliable carrier. Stratification can also be performed using geographic region, time of year (months, seasons), non-stop vs. connected flights, etc. and compared for accuracy. Finally, incorporating estimates of "padding" (the practice of intentionally setting extended flight times to decrease apparent delay) by airlines may provide a more accurate estimate of delay.

[1] Trinkle, Kevin. PSA History. Pacific Southwest Airlines History website. http://www.psa-history.org/about_psa/history. Published 1995. Accessed May 10, 2014.

[2] Office of the Inspector General. *Aviation Industry Performance: A Review of the Aviation Industry, 2008-2011*. Washington, DC: United States Department of Transportation; 2012. Number: CC-2012-029.