Airline Flight Delays and Flight Schedule Padding: An Investigative Report

Dominique Burgauer and Jacob Peters

Systems Engineering 302
Professor Tony E. Smith
Systems Engineering Dept., School of Engineering & Applied Science
University of Pennsylvania, Philadelphia, PA 19104
{dburgau,jacobp}@seas.upenn.edu

Submitted as a final project for Systems Engineering 302, December 2000

Abstract

In this report we will investigate airline flight schedules and delays for one specific route from Los Angeles to San Francisco during the month of September 2000. We find that airlines do tend to pad their schedules to give the appearance that flights arrive on time more frequently than they actually do. Also, we were unsuccessful in finding a multiple regression that accurately predicted either minutes late or actual flight time based on the data at hand. This is most likely due to the fact that external factors, such as weather, affect flight delays above anything else. In Section 1, we present the motivation for our research. Section 2 describes the data in detail. Section 3 presents some initial data observations. Sections 4 and 5 contain regression analysis and plots examining the relationships between certain variables. Sections 6 and 7 summarize our findings and outline areas for further research.

1. Motivation of Research

In the past decades airline travel has become a mainstream means of transportation. Fueled by a tremendous surge in business and leisure travel since the Airline Deregulation Act of 1978, the commercial airline industry has grown rapidly. The deregulated system on the whole has handled the expansion well, adding new routes, new competitors, increased flight frequency, increased capacity on larger planes, and a complex, yet, functional pricing system. Consumers also now benefit from the introduction of frequent flyer programs, rewarding loyal customers with free travel, and new luxuries introduced in upper-class cabins.

Nonetheless, this rapid expansion over the past 22 years has not been smooth. The air travel system has expanded in spurts, where periods of growth and increased service were followed by industry recessions and cost cutting. Most recently, the economic boom of the late 1990's has fueled a 10.8% growth in available passenger seat-miles from 1999Q1 to 2000Q1 as the number of flights has increased[1]. However, this growth has ironically strained the air travel industry on the ground, where capital-intensive infrastructures—airport gates, runways, and air traffic control systems—have begun to cause massive delays (not to mention safety hazards!). Certain airports have antiquated runways—they are either too close together or not oriented correctly for certain wind conditions—exacerbating delays with any small amount of weather. Other airports do not have enough space to move planes around from runways to gates without congestion. And to top it all off, in an area of rapidly increasing computing power, air traffic control systems have not been upgraded and thus cannot handle both in-air control and ground control of planes.

Delays have thus become a standard element of air travel. The sample used in this report has a mean delay of almost sixteen minutes. However, the median delay, counting early arrivals as negative numbers, is zero minutes. So, just how common are delays? It is common to hear the pilot announce to the passengers that even though the plane was delayed in taking off and possibly leaving the gate, the lost time will be made up in the air for an early arrival. Surely, there is a limit to the amount of time that can be made up in-flight as each plane flies at a specific cruising speed. It seems unreasonable that a short flight leaving Los Angeles three minutes late from the gate could arrive in San Francisco fifteen minutes early. Thus, there must be another explanation as to why flights can encounter delays and still end up arriving early.

Since the industry is highly competitive, airlines are eager to avoid the publication of negative performance results. Until recently, this was not a difficult task, since most people are confused by the complexity of the industry and the statistics airlines provide about their performance. In fact, real data about flight times has been difficult to access, even though it has been tracked by the U.S. Department of Transportation. However, the recent uproar about the airline industry performance having a horrible record of flight delays in the year 2000 has brought this information to the public.

[1] U.S. Department of Transportation, Office of Airline Information, http://www.bts.gov/oai/indicators/top.html#PassengerService

Now, for the first time, large amounts of detailed data regarding the industry are publicized on the Internet, allowing us to personally investigate airline statistics. No longer must the consumer listen to airlines about their on-time performance when personal experiences dictate different conclusions. With this data in hand, the consumer can look at the raw flight data and make new conclusions about airlines' real performance.

2. Data

The data used in this study was collected by the United States Department of Transportation (U.S. DOT). The department tracks flight data on the major domestic airlines and recently built a website to allow public access to the databases. The data is accessible at

`http://www.bts.gov/ntda/oai/`

Federal law mandates that any air carrier with more than 1% of total domestic scheduled passenger revenue must report flight statistics. Thus, only the following air carriers are represented in this database: Alaska, America West, American, Continental, Delta, Northwest, Southwest, TWA, United, and USAIR, limiting our analysis of flight delays and scheduling practices. However, these airlines in total cover over 90% of the total domestic operating revenues.

2.1 Sample

We collected specific flight information for the purposes of our investigation. The route selected was Los Angeles International Airport (LAX) to San Francisco International Airport (SFO) and included all flights with scheduled departure times from September 1, 2000, 0:00AM to September 30, 2000, 11:59PM. The sample contains 1182 flights. Of these flights, 1048 flew, and 134 were canceled.

This specific sample was chosen for the following reasons:

- As a short distance, shuttle route, there are many flights per day, allowing us to observe multiple flights under similar daily conditions such as weather, special events, and other uncontrollable daily factors that could contribute to delays. It allows us to analyze delays and scheduling against to time of day.

- Four carriers, Alaska, American, Delta, and United, serve this route, offering a look at how airlines operate under competition. It turns out that scheduled flight times vary significantly between airlines.
- The short length increases the effect of differences in scheduled flight lengths between airlines with respect to general average flight lengths.
- The flying time on this north-south route is less strongly influenced by wind than on a cross-country flight (demonstrated by the hour difference when flying east-west and west-east).
- For convenience of analysis, the route originates and terminates in the same time zone.

2.2 Data Processing

The data obtained from the U.S. DOT listed the following statistics for each flight within the sample:

1. *Carrier* {Alaska, American, Delta, United}
2. *Flight Number*
3. *Origin* {Airport Code: LAX}
4. *Destination* {Airport Code: SFO}
5. *Date* {MM/DD/YY}
6. *Day of Week* {Mon, Tue, Wed, Thu, Fri, Sat, Sun}
7. *Scheduled Departure Time* {HH:MM AM/PM}
8. *Actual Departure Time* {HH:MM AM/PM}
9. *Scheduled Arrival Time* {HH:MM AM/PM}
10. *Actual Arrival Time* {HH:MM AM/PM}
11. *Minutes Late* {+Late/-Early}

The first transformation on this data occurred after the observation had been made, that some flights had actual departure times and actual arrival times of 12:00 AM, regardless of their scheduled times. As *Minutes Late* for these flights was listed as zero, it was obvious that this was a representation for cancelled flights. Also, some flights had only their *Actual Arrival Times* as 12:00 AM with *Minutes Late* as zero, meaning these flights were also most likely cancelled. As there is no method of numerically analyzing the cancelled flight data for our purposes, these flights were removed from the main data set. See section 6 for more discussion on this removal.

Additional calculations were made on the data, creating new flight statistics for analysis. Each of the four statistics involving time (7, 8, 9, and 10) was replicated in additional corresponding columns with the time represented as minutes after 12:00 AM. This was done purely for use in future calculations and to enable scheduled departure time to be used as a predictor variable in regressions.

Subtracting the *Scheduled* or *Actual Arrival Times* from the corresponding D*eparture Times* yielded both the *Scheduled* and *Actual Flight Length* in minutes. Computing these statistics was straightforward, except in cases where flights arrived on the next day. Then, purely subtracting would lead to a negative flight time as a flight departing at 10:00 PM and arriving at 1:00 AM would have a flight length of –1260 minutes. Therefore, to calculate the lengths of these flights, a day's worth of minutes (24 * 60) was added to the final result, accounting for the change of day.

Minutes Departing Late was calculated by subtracting the *Actual Departure Time* from the *Scheduled Departure*, resulting in negative minutes for flights that left early. The difference between *Actual* and *Scheduled Flight Length* was computed in a similar fashion. *Scheduled Flight Length* was subtracted from *Actual Flight Length*, resulting in the difference where negative results mean that the flight took less time to fly than anticipated.

Finally, two categorical variables were created: *isOutlier* and *isRushHour*. Flights were classified as an outlier if the number of *Minutes Late* for the flight was more than $1.5f$ away from the closest quartile (25% or 75%) where f=| 75% quartile – 25% quartile |. This is the classical definition of how to identify outliers. In our case the quartiles were 24 and –9, so the upper cutoff for outliers was defined at 73.5. The lowest cutoff was below the lowest data point, so there were no outliers.

Creating the metrics to decide which hours during the day would be classified as belonging to a "Rush Hour" period was more difficult. Looking at a distribution of *Scheduled Departure Times*, the borders of morning and afternoon rush hour were loosely defined. We used this distribution as a guide and along with our own judgment decided that rush hour flights were flights that had *Scheduled Departure Times* from 6:00 AM to 8:59 AM and 4:00 PM to 7:59 PM.

For cosmetic reasons, we also fixed some of the original data changing the *Day* from {Mon, Tue, Wed…} to {1.Mon, 2.Tue, 3.Wed…} and to the dates in the *Date* variable we changed all one-digit representations of the day to two-digit representations (with a leading zero).

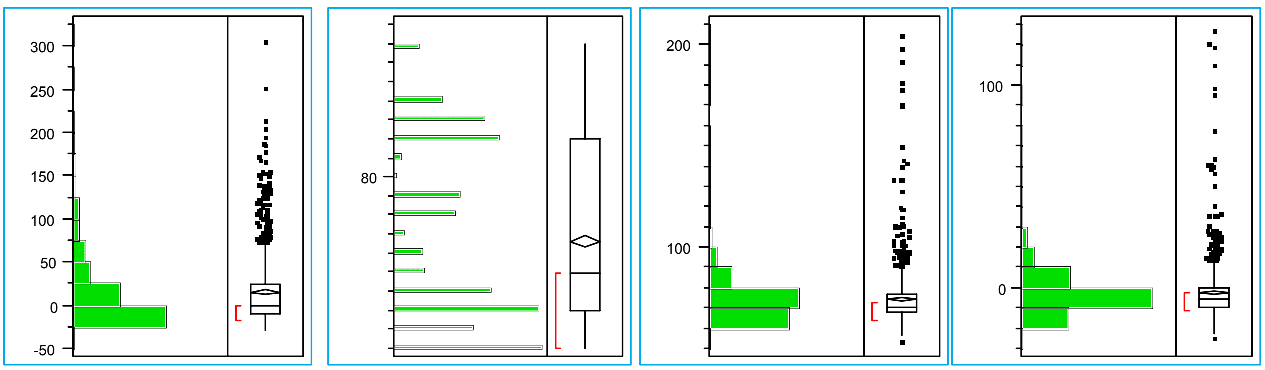
Since different type of aircraft can have a significant impact on actual flying time, we gathered data about the models that the four airlines use for flights from LAX to SFO from

Travelocity.com[2]. All four carriers use aircraft made by Boeing, and cruising speeds for the aircraft are listed on Boeing's webpage[3]. Note that while each of the following planes is a Jet-aircraft, their speeds differ considerably.

Carrier	**Aircraft used**	**Max. Speed**
Alaska	MD-80	504 mph
American	MD-80	504 mph
Delta	Boeing 727-200 / 757	605 mph / 530 mph
United	Boeing 737-300 / 737-500	495 mph / 495 mph

3. Initial Data Observations

To get a feel for the data, we first looked at the distributions of the following variables:



Minutes Late		Scheduled Length (Min)		Actual Length (Min)		Diff btw Actual & Sched.	
Mean	15.855	Mean	76.643	Mean	74.347	Mean	-2.296
Std. Dev	40.883	Std. Dev	4.798	Std. Dev	14.020	Std. Dev	14.059
Median	0.00	Median	75.000	Median	71.00	Median	-5.00

Here we can obviously see that even though the mean *Minutes Late* is over 15 minutes, around half of all flights are early. It is also clear that both the median and mean *Actual Flight Lengths* are less than the both the *Scheduled Flight Length* mean and median. In addition, note that the mean and median differences between each *Scheduled* and *Actual Flight Length* are negative, meaning that flights on the whole when judged from gate departure time were significantly early. This data becomes even more interesting when broken down by airline:

Airline	Mean Scheduled Flightime [Min]	Scheduled Standard Deviation [Min]	Mean Actual Flightime [Min]	Actual Standard Deviation [Min]
Alaska	73.4	1.0	72.2	5.0
American	75.5	4.0	71.8	11.7
Delta	84.7	1.7	78.0	8.5
United	76.2	4.4	72.1	7.5

[2] Travelocity.com: http://www.travelocity.com/
[3] Boeing: http://www.boeing.com/

Here we see that all airlines most probably do pad their schedules, although to varying degrees. Alaska's *Actual Flight Lengths* are only slightly above their mean *Scheduled Flight Length*, whereas Delta adds almost seven minutes to its *Scheduled Flight Length*.

The shortest actual flight during the sample period took 54 minutes (on American Airlines). This means LAX to SFO can be flown in that amount of time. However, since *Actual Flight Length* is measured gate to gate, this figure also includes taxiing to the runway, waiting in line to takeoff, and taxiing back to the gate after landing. This could explain why airlines have different scheduled flight times. For instance, maybe one airline's gate is farther away from the runway than another's.

Clearly, these preliminary results cannot lead to conclusions just yet. However, the above observations shall be more closely investigated in the following sections of this report. The differences between airlines will be discussed further in section 4.3.

4. Single Regression and Plot Analysis

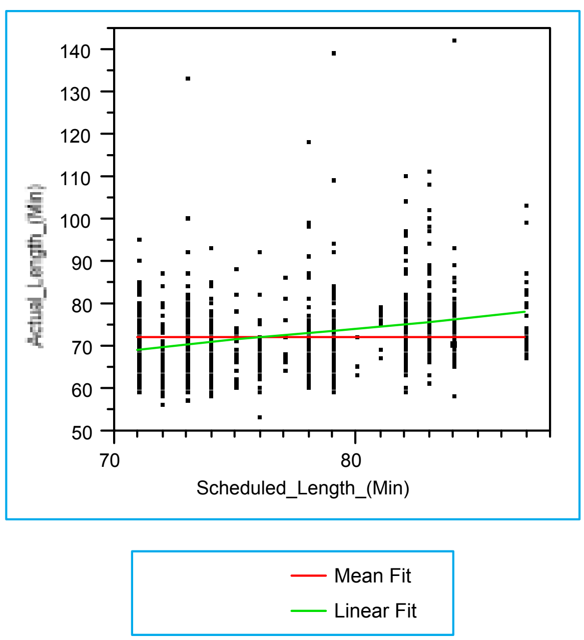
In a beginning effort to study the flight data in more detail, we will do singular regressions where we can and plot analysis for categorical variables in other cases. It should be noted at this point that all analysis from this point will be performed on the non-outlier data of 949 flights. These regressions will be based primarily on the three Gauss-Markov assumptions of linearity, independence and homoscedasticity of our data. The first two assumptions will be examined after the first regression, while the third assumption of independence deserves further discussion at this point.

The structure of the airline industry makes it impossible to assume that flights are independent. The airline system is a large network where many common resources are shared among flights. This dependency causes delays of two types: flight-independent system delays which affect all flights and flight specific-delays that affect specific flights. The first type of delay includes weather delays, airport congestion, and other such system wide delays. The second type is mostly limited to the fact that each plane is scheduled for flights back to back throughout the day. If the plane is delayed on its first flight, this can delay the remaining flights that plane must fly during the day. This effect is especially noticed on shuttle flights where a handful of planes fly back and forth all day between two cities on very tight schedules. The two delays are related as the first type can cause the second.

However, correcting for this dependence between flights is impossible within the scope of this paper and with the available data. It would be necessary to have data on which planes were used in each flight to track how delays began and were subsequently perpetrated through the schedule. Other changes, such as airlines purposely delaying a flight to accommodate extra passengers from another cancelled flight, cannot be seen in this data either. So, for the purposes of this report, we assume that flights are independent. Even with this assumption, it will still be possible to show effects of schedule padding and attempt to predict the actual flying times and minutes late for each flight.

4.1 Actual Flight Length and Scheduled Flight Length Regression

If one were trying to predict *Actual Flight Length*, one might assume that *Scheduled Flight Length* would be a good predictor variable. However, contrary to common sense, *Scheduled Flight Length* gives little insight into *Actual Flight Length*. Here is a simple Least-Squares regression predicting *Actual Flight Length* by *Scheduled Flight Length*:



Mean Fit

Mean	72.49737
Std Dev [RMSE]	8.86569

Linear Fit

Actual_Length_(Min) = 31.3448 + **0.5365 Scheduled_Length_(Min)**

Summary of Fit

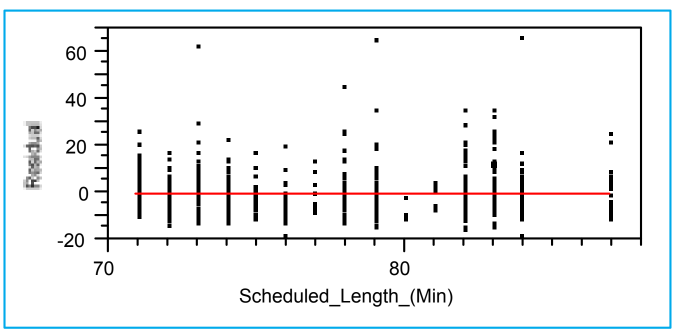
RSquare	**0.084927**
RSquare Adj	0.083961
Root Mean Square Error	8.485345
Observations (or Sum Wgts)	949

Parameter Estimates

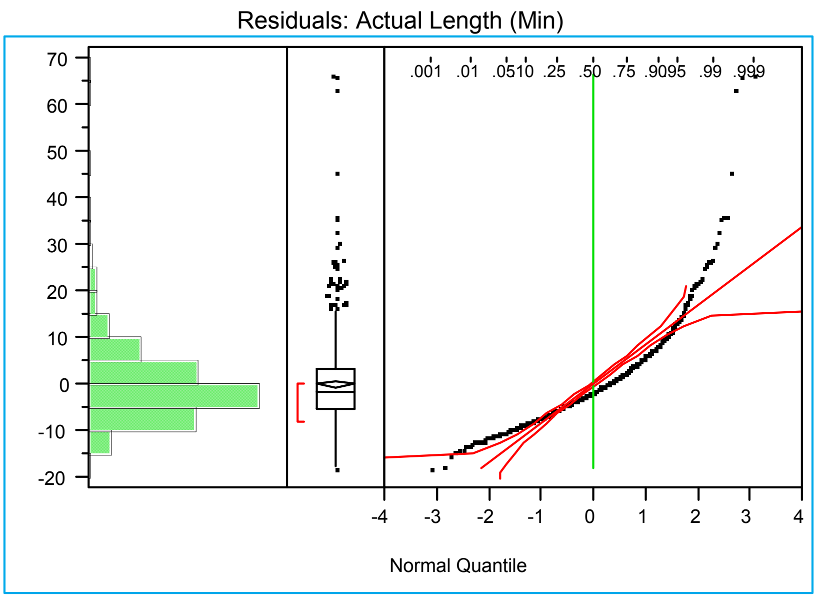
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	31.344801	4.398243	7.13	<.0001
Scheduled_Length_(Min)	**0.5365047**	0.057227	9.37	<.0001

Ideally, every one-minute increase in *Scheduled Flight Length* should result approximately in an increase of one minute in *Actual Flight Length*. The line fit, however, yields a slope of 0.54, meaning that for every minute increase, actual time increased by half a minute. But, with RSquare at 0.085, it is clear that *Scheduled Flight Length* does not account for most of the variance in *Actual Flight Length*.

There are also some interesting observations that can be made directly from the regression graph itself. First, notice that the fastest flight times (below 60 minutes) occur for most of the scheduled lengths. Second, notice how for each scheduled length value that has more than a few flights, the actual times are well spread throughout the 60-minute to 90-minute range.



This residual plot shows that the data does follow the linearity and homoscedasticity assumptions required for the regression. The variance in residuals at different scheduled lengths seems to be fairly consistent, with no identifiable trends. The location of the residuals does not give any indication that a data transformation is necessary to improve linearity. This was investigated further by trying different transformations (ln(x) and x^2). For each transformation, the corresponding regression was performed and the residuals were analyzed. In all cases, the residuals suffered from the fact that in the real data there is a lower bound on how early a flight can be, but not an upper bound on how late a flight can be. Therefore, actual lengths varied much less on the lower side of the mean than on the upper side. Here is a distribution analysis and normal quantile plot for the residuals for the regression above:

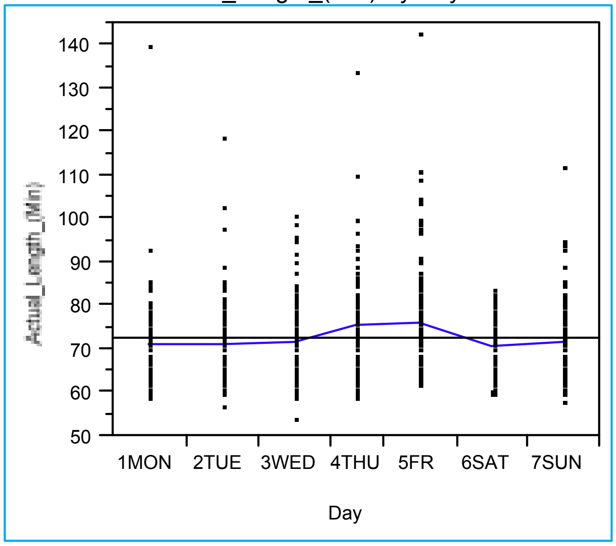


We can see that while the bulk of the data looks normal, there are outliers on the upper side of the mean. It should also be noted that while the mean of the residuals is zero (as defined by a Least-Squares regression), the median is negative, meaning that this regression over-predicted actual times for more than half the flights. This is most likely caused by the weight of the few but significant largest actual lengths.

4.2 Actual Flight Time and Minutes Late by Day of Week Plot

It seems therefore, that some other factor besides *Scheduled Flight Length* causes variance in *Actual Flight Length*. We next examined how *Day* affects *Actual Flight Length*, and the results proved to be very interesting. First, we looked at the *Scheduled Flight Length* by *Day* to see if airlines scheduled their flights differently according to the day of the week. We found that each day had the same mean scheduled flight time at around 76.5 minutes (with minor variances under a minute due to the difference in the number of flights per day from removing the cancelled flights). But, the *Actual Flight Length* formed a distinct pattern, where the means on Thursday and Friday were longer.

Actual_Length_(Min) By Day

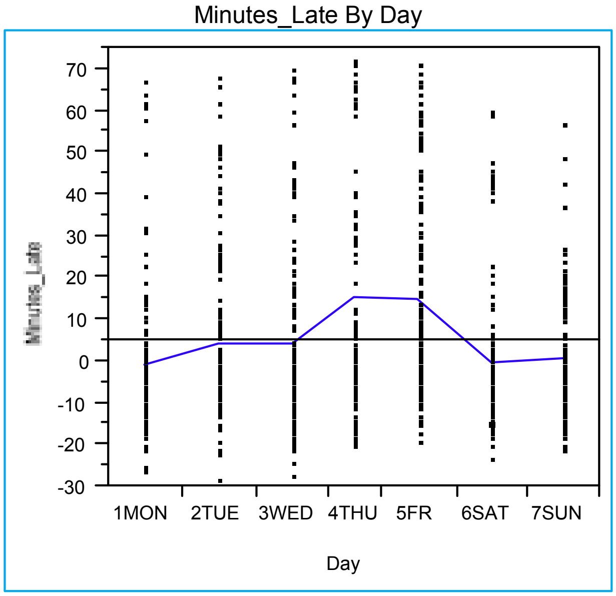


Means and Std Deviations

Level	Number	Mean	Std Dev	Std Err Mean
1MON	146	71.0616	7.8926	0.6532
2TUE	141	71.1986	8.4568	0.7122
3WED	141	71.7943	8.2916	0.6983
4THU	101	75.7822	10.8771	1.0823
5FRI	146	76.0411	10.8099	0.8946
6SAT	137	70.4964	5.3853	0.4601
7SUN	137	71.8905	8.1319	0.6948

It is also apparent that Thursdays during the month of September 2000 had more troubles as only 101 flights flew compared to the rest of the days that had about 40 more. This is an instance where not being able to account for cancelled flights weakens our ability to predict delays.

Because the *Day* had such a strong effect on *Actual Flight Length*, we decided to look at how it affected *Minutes Late*. Here is a plot of those two variables:

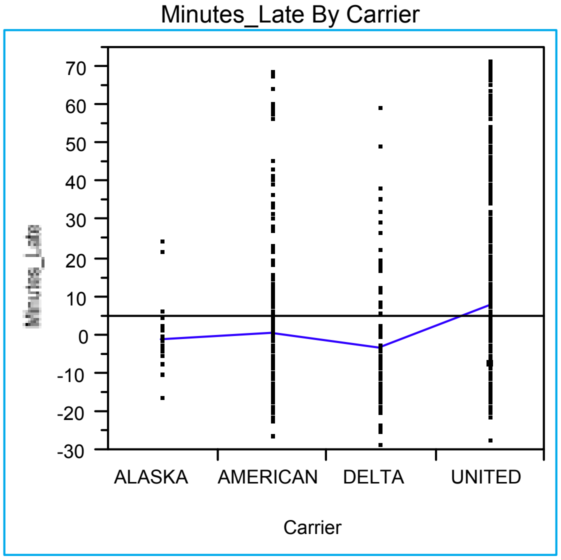


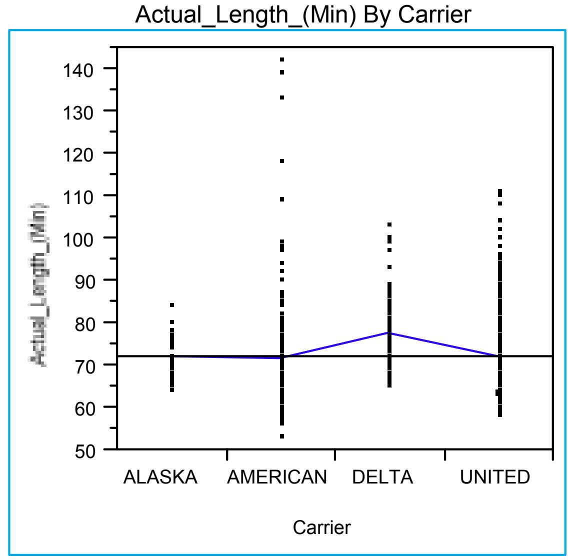
Means and Std Deviations

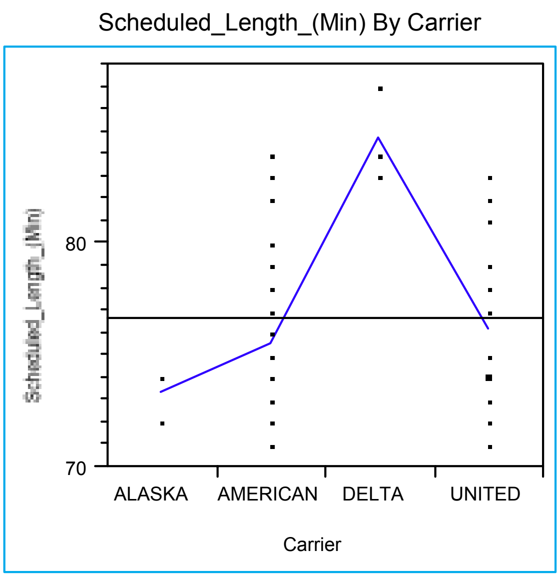
Level	Number	Mean	Std Dev	Std Err Mean
1MON	146	-0.6507	17.3478	1.4357
2TUE	141	4.4043	20.1218	1.6946
3WED	141	4.0355	22.6446	1.9070
4THU	101	15.3564	28.1970	2.8057
5FRI	146	15.0548	22.9509	1.8994
6SAT	137	-0.0657	18.0519	1.5423
7SUN	137	0.4964	15.2470	1.3026

We see an even stronger relationship here between *Day* and *Minutes Late*. It is clearly obvious that flights on Thursday and Friday tend to arrive later than flights on all other days. Having sampled four weeks of flights, the relationships between *Day* and both *Minutes Late* and *Actual Flight Length* should be fairly sound, as it is unlikely that certain external sources of variation hit the same days of the week for four straight weeks. The specific effects of the *Day* on *Actual Flight Length* and *Minutes Late* will be explored further in multiple regressions in Section 5.

4.3 Actual Flight Length and Minutes Late Against Carrier Plot







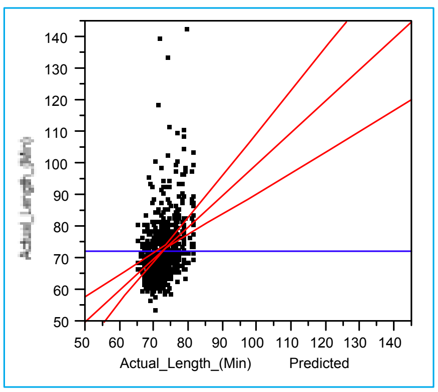
The first plot is deceiving if one does not take into account the padding factor. Alaska, American, and Delta all seem to be right on schedule, while United appears to be approximately 8 minutes late on average. However, if one considers the fact that Delta schedules in excess of 8 minutes more than the next lower competitor, the plot appears in a whole different light. Though Delta seems to be on time or even slightly early, they schedule their flights to be longer than necessary in the first place, since we know that much lower actual means of flight times are possible and achieved by all the other carriers. The second plot shows that Delta's flights do take longer than their competitors' flights, and we can see from the third plot that they do schedule their flights to take longer. Nevertheless, Delta seems to have gone overboard in this respect. Perhaps this policy is an attempt to improve their on-time statistics.

5. Multiple Regressions

So far, we have individually shown how different explanatory variables affect *Minutes Late* and *Actual Flight Length*. We have shown that some of these variables are significant for our purposes. However, we have yet to examine how these single variables behave in a multivariable regression. Here we attempt to predict *Minutes Late* and *Actual Flight Length* using a combination of explanatory variables.

5.1 Actual Flight Length Multiple Regression

The explanatory variables that make most sense in using to predict *Actual Flight Length* are: *Scheduled Flight Length*, *Day*, *Carrier*, and *Scheduled Departure Time*.



Response: Actual_Length_(Min)
Summary of Fit

RSquare	0.151559
RSquare Adj	0.141598
Root Mean Square Error	8.21406
Mean of Response	72.49737
Observations (or Sum Wgts)	949

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Lower 95%	Upper 95%
Intercept	23.690117	6.980929	3.39	0.0007	9.9898215	37.390413
Carrier[ALASKA-UNITED]	0.6471238	1.292486	0.50	0.6167	-1.889422	3.1836693
Carrier[AMERICA-UNITED]	-0.224432	0.702784	-0.32	0.7495	-1.603669	1.1548048
Carrier[DELTA-UNITED]	-0.043249	1.020832	-0.04	0.9662	-2.046664	1.9601663
Day[Fri-Wed]	3.4131441	0.634292	5.38	<.0001	2.1683253	4.657963
Day[Mon-Wed]	-1.489063	0.634241	-2.35	0.0191	-2.733782	-0.244344
Day[Sat-Wed]	-2.422852	0.6531	-3.71	0.0002	-3.704582	-1.141122
Day[Sun-Wed]	-0.722141	0.653301	-1.11	0.2693	-2.004264	0.5599824
Day[Thu-Wed]	3.4570215	0.74299	4.65	<.0001	1.9988807	4.9151622
Day[Tue-Wed]	-1.389689	0.643454	-2.16	0.0310	-2.652488	-0.12689
Scheduled_Departure_Time	0.0023986	0.001107	2.17	0.0306	0.0002254	0.0045718
Scheduled_Length (Min)	**0.6151789**	0.081186	7.58	<.0001	0.4558483	0.7745094

The first observation about this multiple regression is that it has an RSquare value of 0.15 which means it does not account for most of the variance. However, after looking at many combinations of explanatory variables, this was the regression that yielded highest RSquare value. This value indicates that in this regression, the explanatory variables only account for 15% of the variability in the *Actual Flight Length* data. This suggests there is something else creating additional variability or that *Actual Flight Length* is truly random. Considering the large number of other possible variables that could *Actual Flight Length*—including weather, personnel issues, Air Traffic Control congestion and more—it is likely that there are better explanatory variables for predicting variability.

However, we can learn a lot from this regression by looking at the some of the parameters and their significance. We notice that the parameter for *Scheduled Flight Length* has increased slightly from the single regression in section 4.1, by a tenth of a minute. Furthermore, we can individually look at the effects of the different values of the two categorical variables: *Day* and *Carrier*. Here we see that both Thursday and Friday are the only two days with positive parameters, meaning flights on these days take longer than flights on the other das of the week (using Wednesday as a comparison base). These parameter estimates have a 95% confidence interval that is entirely in the positive range, meaning that we can be confident that on these days flights do take longer. We can also be sure that Saturday's flights tend to be earlier than Wednesday's and so do Tuesday's and Monday's, although to a lesser degree. It is also apparent that with respect to *Actual Flight Length*, the selection of an airline, while it might make a slight difference on average as shown previously, does not have a significant effect on the variance of *Actual Flight Length*. The 95% confidence intervals demonstrate this clearly as for all airlines, they span zero, which means we cannot be sure that the effect of carrier choice is not zero.

To look further into the significance of these variables and to look for ways to increase RSquare, we ran a stepwise regression with the same explanatory variables.

Response: Actual_Length_(Min)
Current Estimates

RSquare	RSquare Adj
0.1499	0.1463

Lock	Entered	Parameter	Estimate	nDF	"F Ratio"	"Prob>F"
X	X	Intercept	23.7875481	1	0.000	1.0000
_	_	Carrier{AMERICAN&UNITED&ALASKA-DELTA}	?	1	0.048	0.8274
_	_	Carrier{AMERICAN-UNITED&ALASKA}	?	2	0.047	0.9541
_	_	Carrier{UNITED-ALASKA}	?	3	0.147	0.9313
_	X	Day{Sat&Mon&Tue&Wed&Sun-Thu&Fri}	-2.3569946	2	32.678	0.0000
_	X	Day{Sat&Mon&Tue-Wed&Sun}	-0.4801612	1	2.302	0.1295
_	_	Day{Sat-Mon&Tue}	?	1	1.320	0.2509
_	_	Day{Mon-Tue}	?	2	0.665	0.5148
_	_	Day{Wed-Sun}	?	1	0.013	0.9076
_	_	Day{Thu-Fri}	?	1	0.003	0.9569
_	X	Scheduled_Departure_Minute	0.00260538	1	7.474	0.0064
_	X	Scheduled_Length_(Min)	0.62197657	1	101.859	0.0000

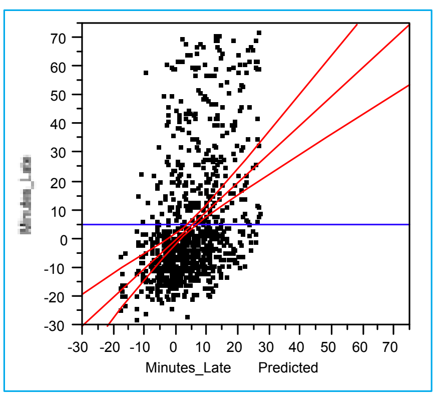
Step	Parameter	Action	"Sig Prob"	RSquare
1	Scheduled_Length_(Min)	Entered	0.0000	0.0849
2	Day{Sat&Mon&Tue&Wed&Sun-Thu&Fri}	Entered	0.0000	0.1408
3	Scheduled_Departure_Minute	Entered	0.0051	0.1479
4	Day{Sat&Mon&Tue-Wed&Sun}	Entered	0.1295	0.1499

Step History

This stepwise regression re-confirms our previously achieved RSquare value. It also shows which variables from the Least-Squares regression contribute to explaining the variability in *Actual Flight Length*. As expected, *Scheduled Flight Length* comes as the most important explanatory variable. Then, we confirm our previous analysis, that flights on Thursday and Friday are longer than on other days (step 2) and then we see that Saturday, Monday and Tuesday are also slightly earlier than Wednesday and Sunday. *Scheduled Departure Minute* also adds a positive effect to flight length. But still, we have not been able to accurately find a method of predicting *Actual Flight Length* with the data at hand.

5.2 Minutes Late Multiple Regression

After having failed to find a significant multiple regression to explain *Actual Flight Length*, we will attempt the same process on *Minutes Late*, although we exclude *Scheduled Flight Length*.



Response: Minutes_Late
Summary of Fit

RSquare	0.173257
RSquare Adj	0.164443
Root Mean Square Error	19.7071
Mean of Response	5.166491
Observations (or Sum Wgts)	949

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Lower 95%	Upper 95%
Intercept	-13.89305	2.33103	-5.96	<.0001	-18.46777	-9.318338
Carrier[ALASKA-UNITED]	-6.060252	3.070906	-1.97	0.0487	-12.087	-0.033508
Carrier[AMERICA-UNITED]	2.1909866	1.560404	1.40	0.1606	-0.871352	5.2533256
Carrier[DELTA-UNITED]	-4.836445	1.956016	-2.47	0.0136	-8.675185	-0.997706
Day[Fri-Wed]	9.6764367	1.521783	6.36	<.0001	6.6898925	12.662981
Day[Mon-Wed]	-6.276885	1.521647	-4.13	<.0001	-9.263162	-3.290607
Day[Sat-Wed]	-4.841224	1.566522	-3.09	0.0021	-7.915568	-1.766879
Day[Sun-Wed]	-5.993801	1.566371	-3.83	0.0001	-9.067851	-2.919751
Day[Thu-Wed]	10.228968	1.779017	5.75	<.0001	6.7375947	13.72034
Day[Tue-Wed]	-1.250205	1.543722	-0.81	0.4182	-4.279804	1.7793936
Scheduled_Departure	0.0162775	0.002125	7.66	<.0001	0.0121078	0.0204473

Again, this regression suffers from an extremely low RSquare value of 0.17. However, while this regression fails to explain the variability in delays, it does shed some light on the categorical variables. We see that here that using Wednesday as a base, Thursday's and Friday's flights tend to be much more delayed than on other days. And it is apparent that Monday, Saturday and Sunday tend to have shorter delays. However, differing from the previous regression, the choice of Delta for carrier does have a significant effect on reducing delays. We can be 95% confident that Delta has shorter delays that United, as the confidence interval for the Delta parameter does not include zero.

To better understand what is happening in this regression, we can look at a stepwise regression.

Response: Minutes_Late

Current Estimates

RSquare	RSquare Adj
0.1728	0.1676

Lock	Entered	Parameter	Estimate	nDF	"F Ratio"	"Prob>F"
X	X	Intercept	-7.9667429	1	0.000	1.0000
_	X	Carrier{DELTA&ALASKA&AMERICAN-UNITED}	-6.0313093	3	17.592	0.0000
_	X	Carrier{DELTA-ALASKA&AMERICAN}	-1.4415867	2	4.643	0.0098
_	X	Carrier{ALASKA-AMERICAN}	-4.0879667	1	3.726	0.0539
_	X	Day{Mon&Sat&Sun&Wed&Tue-Fri&Thu}	-6.7289698	2	49.542	0.0000
_	X	Day{Mon&Sat&Sun-Wed&Tue}	-2.1598926	1	8.136	0.0044
_	_	Day{Mon-Sat&Sun}	?	1	0.181	0.6702
_	_	Day{Sat-Sun}	?	2	0.207	0.8130
_	_	Day{Wed-Tue}	?	1	0.016	0.9004
_	_	Day{Fri-Thu}	?	1	0.047	0.8281
_	X	Scheduled_Departure_Minute	0.0161834	1	58.748	0.0000

Step History

Step	Parameter	Action	"Sig Prob"	RSquare
1	Day{Mon&Sat&Sun&Wed&Tue-Fri&Thu}	Entered	0.0000	0.0760
2	Scheduled_Departure_Minute	Entered	0.0000	0.1193
3	Carrier{DELTA&ALASKA&AMERICAN-UNITED}	Entered	0.0000	0.1574
4	Day{Mon&Sat&Sun-Wed&Tue}	Entered	0.0041	0.1647

5	Carrier{DELTA-ALASKA&AMERICAN}	Entered	0.0187	0.1696
6	Carrier{ALASKA-AMERICAN}	Entered	0.0539	0.1728

Here we see that the single most significant factor in predicting delays is whether or not a flight is on Thursday or Friday. We also see again that *Scheduled Departure Time* (or more simply time of day) contributes significantly to the regression as well. We plotted *Minutes Late* (Y) versus *Scheduled Departure Time* (X) and found that while the minutes late vary significantly at each time of day, the delays do tend to get bigger as the day progresses. This effect is most likely related to the question of data independence as delays early in the day compound as the day progresses and resources are re-used from flight to flight. It is also apparent that United has longer delays that the other carriers and that Delta and Alaska have shorter delays.

6. Findings

Our results were consistent with our initial experience-based observations that airlines tend to "pad" flight schedules to give the appearance that flights arrive on time in most instances. We have shown that certain airlines practice this more than others as in the case of Delta on this particular route. We were unsuccessful, however, in finding a group of explanatory variables that could predict flight delays or actual flight length to a reasonable extent, although in the process we gained insight into certain factors that do significantly influence flight delays and actual flying time.

It has been shown that on two days of the week, Thursday and Friday, flights tend to be delayed longer and actual flight lengths increase on average. This result seems reasonable considering that these two days correspond with both the end of the business week and the beginning of the weekend travel period, increasing the total load on the air travel system and causing delays. The stepwise regression on minutes late shows that delays on Thursdays and Fridays increase by almost seven minutes as compared to all other days. On a flight that lasts about an hour, this delay can be significant in percentage terms. The data also confirms that flights tend to be slightly more delayed later in the day.

Also, we found that flying Delta improves one's chance of arriving early, although this sense of arriving early is false as Delta's flights average longer in minutes than the other carriers. Clearly then, the best choice of airline is Alaska Airlines as their mean actual flight time is very low and their standard deviation is small, although this decision is made with a narrow perspective. In terms of convenience and flexibility, United offers the most flights by far. American, United and Delta offer greater connectivity with other domestic destinations reachable from San Francisco. It is also apparent that United cancels the largest percentage of its flights, reducing convenience. However, this might be a direct effect of having a tighter flight schedule as if there is bad weather one afternoon, Alaska might not have any flights cancelled whereas United might have to cancel a handful.

	Flights Cancelled	Flights Flown	Total	% Cancelled
Alaska	2	28	30	**6.67%**
American	10	240	250	**4.00%**
Delta	2	86	88	**2.27%**
United	120	694	814	**14.74%**

It is thus very difficult to pick a metric or a combination of metrics to evaluate which airlines perform best. However, it is possible to see that airline flight scheduling is a statistical science and airlines knowingly create longer flight schedules to reduce apparent delays.

7. Further Research

There is tremendous room for additional research in this area. Unless the variability in *Minutes Late* and *Actual Flight Length* is truly random, there are other explanatory variables which could predict these values. There are many reasons for flight delays including weather, ground crew problems, mechanical problems, passenger problems, etc. Data for some of these factors is probably collected by the airlines and closely guarded for good reason. For example, airlines do not want the public to have detailed knowledge of mechanical problems and the delays they cause. Other data, such as historical weather, is not publicly available on the Internet (as checked through various search engines), but it is collected. A further analysis into the real causes and factors influencing delays would require significant effort and expense in obtaining the necessary data.