

# Optimizing Pilot Connection Time Using PROC REG and PROC LOGISTIC

Andy Hummel, Delta Air Lines, Atlanta, GA

## ABSTRACT

As any airline traveler knows, connection time is a key element of the travel experience. A tight connection time can cause angst and concern, while a lengthy connection time can introduce boredom and a longer than desired travel experience. The same elements apply when constructing schedules for airline pilots. Like passengers, pilot schedules are built with connections. Delta Air Lines operates a hub and spoke system that feeds both passengers and pilots from the spoke stations and connects them through the hub stations. Pilot connection times that are tight can result in operational disruptions whereas extended pilot connection times are inefficient and unnecessarily costly. This paper will demonstrate how Delta Air Lines utilized SAS® PROC REG to analyze historical data in order to build operationally robust and financially responsible pilot connections.

## INTRODUCTION

A typical Delta pilot flying a domestic schedule will fly up to five flights in a work day which can consist of four connections through a hub station. Delta's largest hub is Atlanta, and the largest domestic aircraft fleet is the McDonnell Douglas MD-88. Each day there are thousands of pilot connections in the hubs. From a strict optimization perspective, the ideal pilot schedule would consist of minimal connection times thereby allowing the most flight time in the least amount of work day. However, experience has shown tight connection times result in operational disruptions and unhappy passengers since there is little room to absorb operational disruptions. The key is to find the equilibrium of minimal connection time and robust operational reliability.

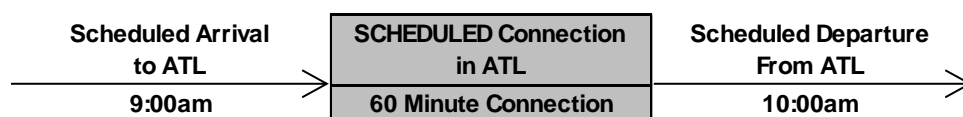
## HOW TO MEASURE SUCCESS

From a passenger perspective, the key to a successful flight is arriving safely and on time. Obviously, if a flight departs on time the likelihood of arriving on time is high. A key metric Delta Air Lines measures is the success rate of on time departures, also known as "d-zero" or "D0". A flight leaving on time or early will achieve D0 while a flight departing late will result in D0 failure. Delta places a large emphasis on achieving a high rate of D0.

Delta operates a hub and spoke structure, and for a flight to achieve D0 the preceding flight must arrive on time in order for pilots to connect to the next flight. If flight A is the first flight of the day to a hub station and arrives late, the arrival delay from flight A will result in a departure delay to flight B and compound throughout the day to delays to flights C, D and E. Building an operational buffer between flights allows the system to absorb disruptions and allows late arriving flights to turn into on time departures.

## DEFINING "SCHEDULED" AND "EFFECTIVE" CONNECTION TIME

Connection time plays a large role in the success of D0 and there are two definitions of connection time used for this study. "Scheduled" connection time is simply the time between a *scheduled* arrival time of a flight to hub and the *scheduled* departure time of the next outbound flight. In the example below, the pilot was scheduled to arrive in Atlanta (ATL) at 9:00am and depart at 10:00am resulting in a 60 minute "scheduled" connection.



"Effective" connection time is the difference between *actual* arrival time and the *scheduled* departure time. In the example below, the actual arrival time of the flight was 9:30am and the scheduled departure time was 10:00am resulting in a 30 minute "effective" connection time. In other words, effective connection time is the connection time the pilot has to achieve a successful D0 departure on his next flight.

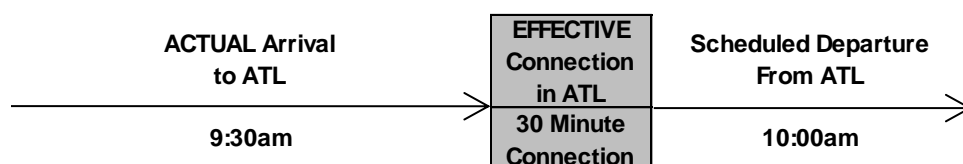


Figure 1. Comparison of schedule and effective connection time

A scheduled connection time of 60 minutes allows the pilot an ample amount of time to connect between flights. However, the 30 minute late arrival of the first flight now gives the pilot only 30 minutes to connect to his next flight, greatly decreasing the likelihood of D0 on the second flight.

In general at a system level, the mean *effective* connection time for Delta is GREATER than the *scheduled* connection time. This indicates that inbound connection flights arrive earlier than scheduled and the pilots have more time to connect than scheduled.

Figure 2 below compares the mean scheduled connection time on the X axis and mean effective connection time on the Y axis. For all observations of scheduled connection time, effective connection time is greater.

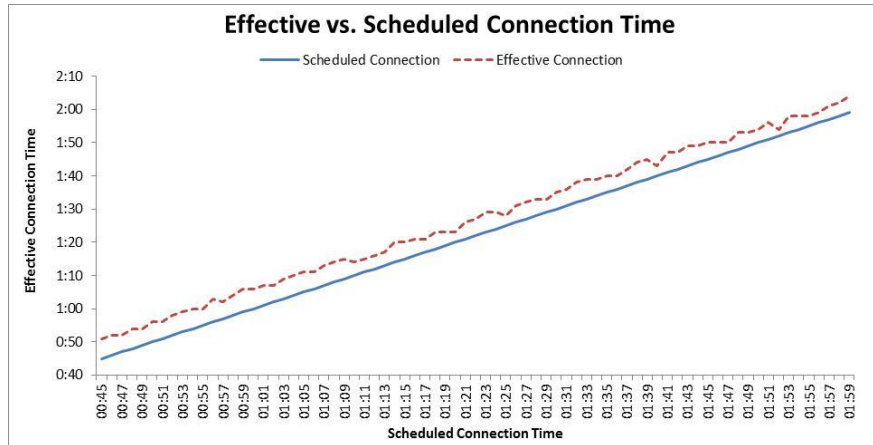


Figure 2. Comparison between Scheduled and Effective pilot connection time.

## RELATIONSHIP BETWEEN D0 AND CONNECTION TIME

Intuitively, the assumption was made that the more effective connection time a pilot has the greater the success of D0. A larger connection time allows for the absorption of operational disruptions and gives the pilot a higher chance of connecting to his outbound flight on time. Historical analysis using SAS showed that this was true, to a degree.

There is a very strong positive correlation between effective connection time and D0 between 0 and approximately 55 minutes of effective connection time. As effective connection time increased from 0 minutes, the success of D0 also increased. After 55 minutes the success of D0 plateaus and little additional benefit is observed. There are many contributors to D0 including weather, maintenance, air congestion and airport ground traffic, and pilot connection time is not the sole influence. In the role of building pilot schedules, pilot connection time is the only variable that can be influenced and will be the focus.

Figure 3 below displays effective connection time along the X axis and success rate of D0 along the Y axis and demonstrates the increase in the rate of D0 between and 60 minutes of effective connection time.

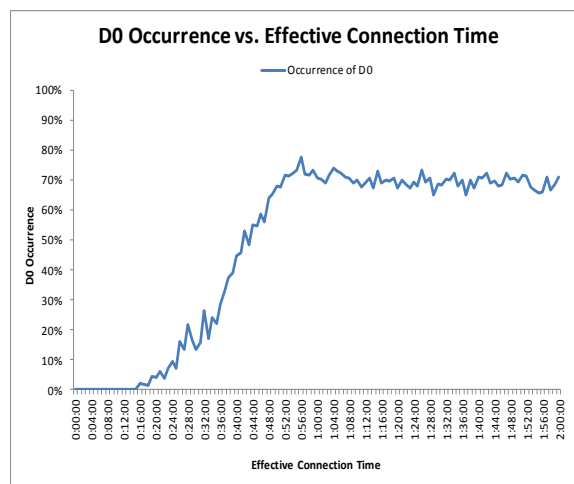


Figure 3. Relationship between D0 and effective pilot connection time.

## OTHER IMPACTS ON D0

There is a clear relationship between effective connection time and D0. While pilot connection time is an influence that can be controlled, there are other variables that have an impact on D0 and need to be identified. Using basic statistical procedures, such as PROC CORR and PROC GPLOT, two other variables have been identified that affected the success of D0: time of day and time of year.

In order to account for time of day and time of year, dummy variables were created. Time of day was divided into four distinct periods: 00:00 - 08:59 (Timeofday1), 09:00 - 11:59 (Timeofday2), 12:00 – 16:59 (Timeofday3), and 17:00 - 23:59 (Timeofday4).

Time of year was divided into three periods, Summer (June, July, August), Winter (December, January, February) and Off Peak Months (September, October, November, March, April, May). Using PROC LOGISTIC, a model was built to predict D0 using the independent variables of following model using the code below:

```
• PROC LOGISTIC DATA=Final DESCENDING;
  MODEL D0_Count = Effective_connect Timeofday1 Timeofday2 Timeofday3
  Summer OffPeakMonths /RSQ LACKFIT;
  RUN;
```

The model is:  $\text{Log (Probability of D-zero)} = -1.0487 + 0.000287 (\text{effective connection time in seconds}) + 1.0782 (\text{Timeofday1}) + 0.6688 (\text{Timeofday2}) + 0.1215 (\text{Timeofday3}) + 0.2240 (\text{OffPeakMonths}) - 0.2024 (\text{Summer})$

The Logistic model was chosen due to the ability to predict a dichotomous dependent variable. The interest was in knowing whether or not D0 would be satisfied based on the independent variables. The predicted value of D0 reflected the probability that D0 was met based on the predictor variables. The simple and multiple regression models were not suited for our investigation since our dependent variable was binary. The DESCENDING option was used so that 1 represented 100% chance of meeting D-zero. The RSQ and LACKFIT options were used to ascertain how well the model fits the data.

After the model was built, the estimated likelihood of D0 was found using the regression equation provided by the model. The following example illustrates the application of the model:

- Suppose that we were interested in knowing the likelihood of D-zero on July 4, 2014 at 10:00am, with 55 minutes of effective connection time
- We would use the regression formula above to determine the probability
- $\text{Probability of D-zero} = \text{Exponential}^{-1.0487 + 0.000287 (3300 \text{ effective connection time in seconds}) + 0.6688 (1 \text{ Timeofday2}) - 0.2024 (1 \text{ Summer})} / 1 + \text{Exponential}^{-1.0487 + 0.000287 (3300 \text{ effective connection time in seconds}) + 0.6688 (1 \text{ Timeofday2}) - 0.2024 (1 \text{ Summer})}$
- $\text{Probability of D-zero} = \text{Exponential} (0.3648) / 1 + \text{Exponential} (0.3648)$
- $\text{Probability of D-zero} = 1.44 / 2.44 = 0.590$  or 59%
- The Model predicts a 59% probability that D-zero will be met if a pilot has 55 minutes of connection time on July 4, 2014 at 10:00am.

In order to estimate effective connection time using the model, D0 was set equal to a value between 0.0 and 1.00 that corresponded to a desired level of D0 success. Values for the categorical independent variables were entered based on the time of day and year that was desired to be predicted. Effective connection time was then solved for to determine how much effective connection time was needed for that level of D0. Using this methodology, management was provided with a table that estimated how much effective connection time was needed during different times of the day for the summer months. Figure 4 displays a table of various levels of D0 and effective connection time.

Summer Recommendations for D0				
Levels of D0	00:00 - 08:59	09:00-11:59	12:00-16:59	17:00-23:59
Connection Time For D0=60%	42 minutes	63 minutes	84 minutes	93 minutes
Connection Time For D0=65%	50 minutes	71 minutes	92 minutes	101 minutes
Connection Time For D0=70%	58 minutes	79 minutes	100 minutes	109 minutes

Figure 4. Table of desired D0 and predicted required effective connection time.

## ACHIEVING OPERATIONAL SUCCESS WITH D0

The analysis has shown that that effective connection time is correlated to achieving an on-time departure (D0). Using PROC LOGISTIC a model was built to predict D0 based on effective pilot connection time, time of day, off peak period and time of year.

However, as defined earlier, effective connection time is the difference between *actual* inbound arrival time and *scheduled* outbound time. Pilot schedules are built weeks in advance and the actual inbound arrival time is obviously unknown, and scheduled connection times must be used. Therefore, the goal is to build a model to predict the scheduled connection time needed to achieve a defined level of effective connection time at various time of day, and times of year.

PROC REG was used to predict scheduled connection time based on the independent variables of: effective connection time, time of day, off peak period and time of year. Once scheduled time is known, it can then be applied when building pilot schedules to arrive at a desired level of D0.

Figure 5 is a visual of the process flow. First a defined level of desired D0 is set internally. PROC REG is used to predict the amount of *effective* connection time needed to achieve that level of D0. A separate PROC REG is then used to predict the required amount of *scheduled* connection time needed for the predicted level of *effective* connection time. The level of scheduled connection time is used when constructing pilot schedules.

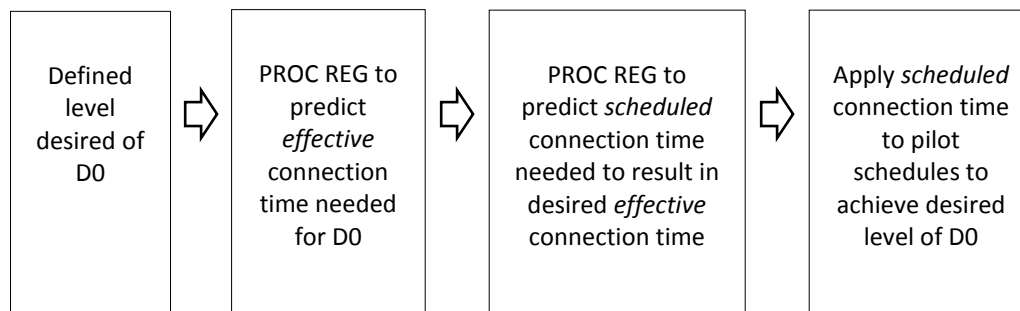


Figure 5. Visual of process flow.

## CONCLUSION

From a passenger's perspective, departing on time, or D0, is essential to a positive travel experience. From an operational and pilot scheduling perspective D0 is required to run an on time operation. The primary influence Delta Air Lines has from a scheduling perspective to achieve D0 is the amount of time a pilot has to connect from one flight to the next.

The success of D0 is not solely dependent on pilot connection time and by using SAS other variables that impact D0 (time of day and season of year) were identified. Once the model was built a desired level of D0 success was set for a defined time of day and season of the year to predict an effective connection time needed to achieve that level of D0. Then a second model was built to determine the amount of scheduled connection time needed.

The use of SAS has allowed Delta Air Lines to be much more surgical in setting pilot connections times. Prior to this analysis, pilot connection time was set at a very broad level with little knowledge of how connection time impacts the operation. We now have the ability to more efficiently build pilot schedules while increasing operational performance.

## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Andy Hummel  
Delta Air Lines  
General Manager – Pilot Crew Resources  
Andrew.Hummel@Delta.com  
404-715-1270

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.