Robust Optimization of Travel Management in Europe

Mason Chen[1], Charles Chen[2]
[1]Stanford University OHS, Palo Alto, CA, USA, mason05@ohs.stanford.edu
[2]Applied Materials, Sunnyvale CA, USA, charles.chen.training@gmail.com

Abstract
This paper was addressing travel cost & time while visiting five major cities in Europe. The main objective was to build a robust route which can meet both time & expense requirements. The input variables used were the intra-city time & expense while the response variables were the total time & expense. Several travel design constraints were considered in the model as noise factors. Each input variable was limited to 2 different transportation choices: flight or train. The literature research was conducted about train & flight speed in Europe. Taking train is better when travel distance is under 500km, and flight is better if above 1,000 km. The choice can go either way in between. Most major cities in Europe are in that range. A multiple regression model was built on the mean time & expense. Optimal travel route was set by meeting the desirability functions of 2 travel requirements. To achieve Robust Design, Monte Carlo (MC) Simulation was conducted by including the time & expense variation. MC results have observed 8% risk of not meeting the expense budget. By raising the importance of time desirability, the new optimal route can meet both time & expense requirements.

Key Words: DOE, Predictive Modeling, Europe Travel Monte Carlo Simulation, Statistics, JMP

1. Introduction

Many people like travelling but have three constraints on Time Management, Expense Budget, and Travel Quality. The objectives of this paper are: (1) manage travel in Europe, (2) build a statistical model to predict travel duration and expense, and (3) conduct a Robust Design to optimize the travel plan. A great way to save time and money when travelling across Europe is by taking the trains. Eurostar is more convenient and economic than taking Flights. However, Eurostar only has direct trains among major cities such as London, Paris, Brussel, Lyon, Avignon… Taking trains among other cities in Europe may not be a better choice than taking a Flight.

1.1 Design Europe Travel Package

This paper would design a summer travel package on visiting five major European cities: London, Paris, Amsterdam, Prague and Munich as shown in Figure 1. The starting and ending destination will be at the Paris Charles de Gaulle CDG Airport applicable for most Visitors. CDG airport is the largest international airport in France and the second largest in Europe. This paper would only consider the travelling time and expense of taking either train or/and flight among five destinations. It won't consider the other travel duration and expense during visiting each city.

Figure 1: travel among five major European cities

1.1 "STEMS" Approach

"STEMS" (Science, Technology, Engineering, Mathematics, Statistics) methodology was applied on this project to help define the project scope as shown in Figure 2 STEMS diagram. The Kinetic and Friction Physics "Science" was studied to compare transportation speed and duration between train and flight. "Technology" is understanding the Europe train system. Systematic "Engineering" problem solving techniques such as Decision Flow Chart is utilized to optimize the travel package. "Math" can help calculate the estimated transportation duration and choose the more convenient way among five major destinations. JMP statistical software and DOE Robust Design were used through this paper. All 5 "STEMS" elements are critical to making this project successful [1-3].

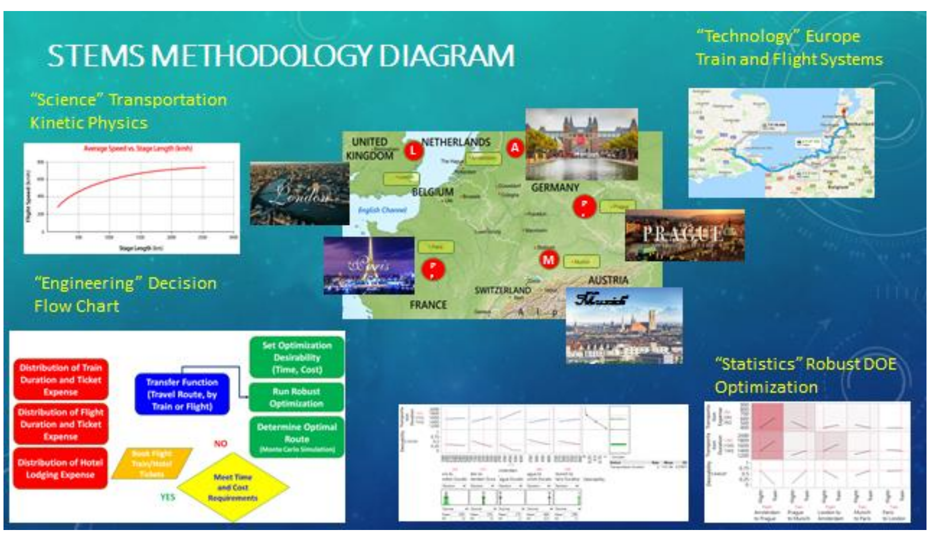

Figure 2: STEMS Diagram

1.3 Identify Design Constraints

To build a regression model, two Response Variables are: (1) Total Intra-City Transportation Time and (2) Total Intra-City Transportation Expense. Three Input Variables are: (1) Travel Route (Sequence), (2) Each Intra-City Transportation Time, and (3) Each Intra-City Transportation Expense.

In addition, to simplify the regression model, the following Design Constraints would be considered throughout this paper:

- Collect all transportation raw data on July 11, 2018 and book tickets by May 26, 2018
- Start and End at Paris CDG Airport or Train Main Station
- Add 3 hours Check-In/out Time for flight, and 1hour Check-In/out Time for Euro Star Train
- Take Economic Flight Seat or 2^{nd} Class Train Seat (all one-way ticket)
- Only consider direct flight or direct train if available. Otherwise, take one stop in transition
- Only consider flight or train after 9am and arrive by 9pm within the same day
- Won't consider driving or ferry transportation among five cities
- Won't consider Flight/Train Delay Factor

1.4 Set Travel Transportation Requirements

To manage this special travel package among five major cities in Europe, two reasonable travel requirements are set as following:

(1) Total Intra-City Transportation Duration < 28.8 Hours:

- Total 12 Travel Days, 9am-9pm, 12 hours*12 Days= 144 Hours
- Less than 20% of 144 Hours is allocated to intra-city transportations < 28.8 Hours
- Other local transportation duration won't be included in this first requirement.

(2) Total Intra-City Transportation Expense < $500 USD:

- Total Budget: < $5,000 USD
- 12 Hotels = $2,400 USD ($200/Night)
- Meals: $900 USD ($75/Day)
- Other Tour/Local Transportations: $1,200 USD ($100/Day)
- Total Intra-City Transportation Expense Budget ~ $500 USD

2. Data Collection and Baseline Analysis

Section 2 will conduct the following baseline analysis: (1) Driving/Flying Distance Data, (2) Average Driving/Flying Speed and Duration, and (3) Project Management Flow Chart.

2.1 Driving/Flying Distance Data

To design the shortest transportation duration among five destination cities, both the driving distance and the flying distance information are provided in the Figure 3 diagram. There are two distance numbers in each intra-city pair. The first number is the Driving Distance and the second one is the Flying Distance. For example, for L (London) – A (Amsterdam) pair, the driving distance is 550km, and the flying distance is 358 km.

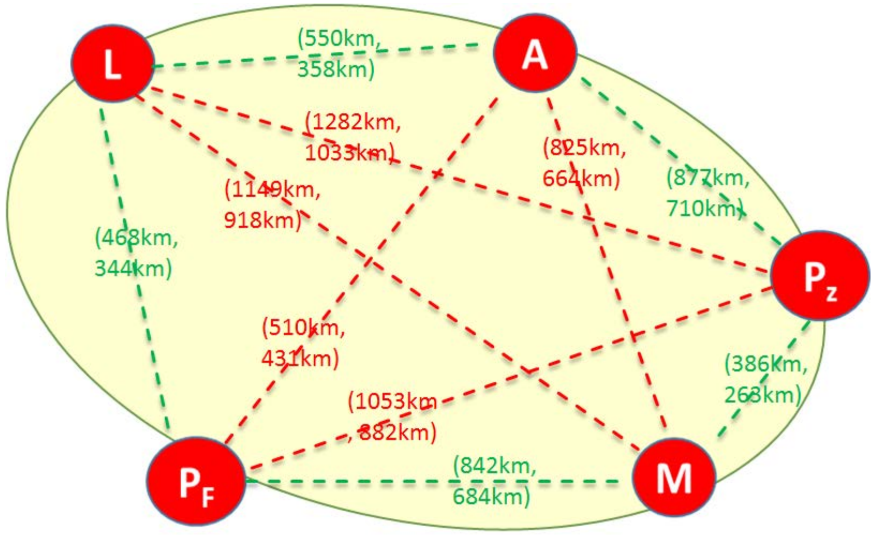

Figure 3: Driving and Flying Distance among five major cities

Based on Figure 3 Distance Diagram, the Shortest Green Route is by Train 3123km, and by Flight 2359km. The distance ratio factor of Driving/Flying is ~ 1.3. The transportation duration ratio will depend on this Distance Ratio and Speed Ratio. To simplify this project, we will only consider two shortest routes (3123km Driving Distance, 2359km Flying Distance)
in Figure 3:

1. Paris-London-Amsterdam-Prague-Munich-Paris (Clockwise)
2. Paris-Munich-Prague-Amsterdam-Paris (Counter-Clockwise)

Though, the one-way ticket price and flying duration may be slightly different between the opposite directions

2.2 Average Driving/Flying Speed and Duration

Depending on the length of the aircraft, it may then take 15-20 minutes for the plane to climb to its cruising altitude and another 15-20mins for preparing landing. There are four forces acting upon an aircraft: (1) Weight (Gravity), (2) Lift – acting perpendicular to the direction of relative motion, (3) Thrust – acting along the direction of motion, generated by engines to move the aircraft forward, and (4) Drag – acting opposite to the relative motion of the aircraft, generated by the air resistance. The lift force "holding" a plane up is generated by airflow over the wings. Lift is only possible if the relative air speed must be large enough. Train acceleration physics is much simpler, and it typically takes 5mins to reach the full speed.

According to the Eurostar Train service, train line speeds are 300 km per hour except within the Channel Tunnel, where a reduced speed of 160 km per hour applies for safety reasons. Based on 80% (full)-20% (Reduced) ratio, the estimated average Eurostar train speed is around 265km/hour. As shown in Figure 4, the average flight speed is drawn vs. stage length. The average flight speed will be significant lower with shorter length due to significant portion of take-off and landing.

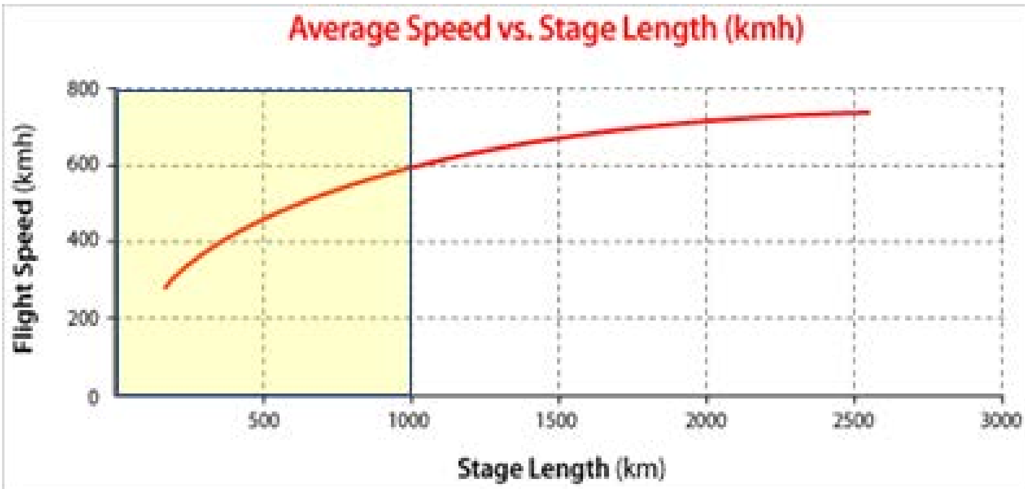

Figure 4: Flight Speed vs. Stage Length

While considering the driving/flying ratios (distance, speed, speed up/down) plus the check in/out time, it may make more sense to take train if distance is shorter than 300kms and take flight if longer than 1,000kms. It would be a close call if the intra-city distance is between 300kms, and 1,000kms (all five intra-city transportation distance are in this range). If taking all routes by train, the total transportation duration goal (below 28.8hours) won't be met. If taking all routes by flight, the total transportation expense goal (below $500 USD) won't be met. We need an optimal route to achieve both goals.

2.3 Project Management Flow Chart

To manage this travel task systematically, a project management flow chart was made as shown in Figure 5. The flow chart starts with red zone by collecting all the raw data of train/flight duration and ticket expense. The lodging expense won't be included in this paper. The Blue Zone would build a transfer function of all input variables which will contribute to two Reponses (total transportation duration and expense). The Green Zone would conduct sensitivity analysis and Robust optimization to determine the Optimal Route. The optimal route would be validated against two requirements and decide whether to book tickets.

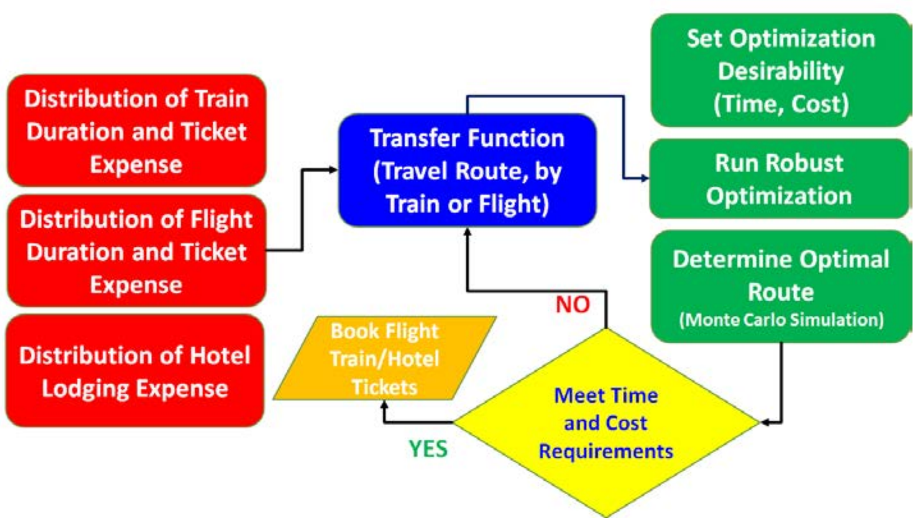

Figure 5: Project Management Flow Chart

3. Route Design Optimization

In Section 3, a special design of experiment (DOE) methodology was utilized to determine the optimal route: (1) Design a Structured DOE, (2) DSD Optimization Result, and (3) Monte Carlo Simulation

3.1 Design a Structured DOE

As shown in Figure 6, a special JMP Definitive Screen Design (DSD) was conducted in order optimize the transportation route. A "definitive screening design" (DSD) would be conducted to optimize the "Neural" algorithm. Here are areas where definitive screening designs are superior to standard screening designs: (1) identify the causes of nonlinear effects by fielding each continuous factor at three levels and (2) avoid confounding between any effects up through the second order [4-8]. There are five one-way transportation segments of categorical input variables (Flight or Train), total 18 DSD runs.

Design

Run	Paris to London	London to Amsterdam	Amsterdam to Prague	Prague to Munich	Munich to Paris
1	Train	Train	Train	Train	Train
2	Flight	Flight	Flight	Flight	Flight
3	Train	Train	Train	Train	Flight
4	Flight	Flight	Flight	Flight	Train
5	Train	Flight	Train	Train	Train
6	Flight	Train	Flight	Flight	Flight
7	Train	Flight	Flight	Train	Train
8	Flight	Train	Train	Flight	Flight
9	Train	Train	Flight	Flight	Train
10	Flight	Flight	Train	Train	Flight
11	Train	Flight	Train	Flight	Flight
12	Flight	Train	Flight	Train	Train
13	Train	Train	Flight	Train	Flight
14	Flight	Flight	Train	Flight	Train
15	Train	Train	Train	Flight	Train
16	Flight	Flight	Flight	Train	Flight
17	Flight	Flight	Flight	Flight	Flight
18	Train	Train	Train	Train	Train

Figure 6: DSD Design Metrix

To ensure the DSD structure, three examination criteria was done before conducting the DSD simulation runs on Neural Network algorithm seen in Figure 7a, and 7b.

Power Analysis

Significance Level 0.05

Anticipated RMSE 1

Term	Anticipated Coefficient	Power
Intercept	1	0.973
Paris to London	1	0.888
London to Amsterdam	1	0.956
Amsterdam to Prague	1	0.956
Prague to Munich	1	0.956
Munich to Paris	1	0.956

Apply Changes to Anticipated Coefficients

Figure 7a: Power Analysis

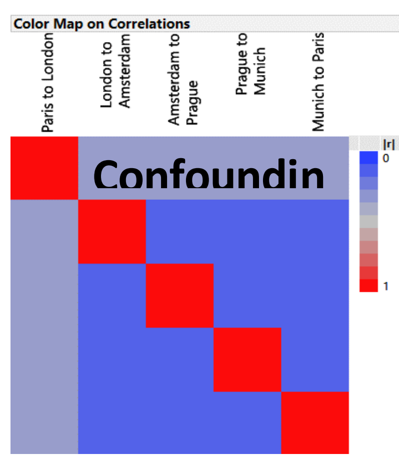

Figure 7b: Confounding Color Map

The first "Power" analysis is to check whether the DSD Run Size is sufficient. If run size is too small, the 95% confidence interval of any effect term will be very wide. Then, the Power level would indicate the probability of the predicted sign is still valid. In Figure 7a, all power levels are above 88% (little Run Size concern). The second "Confounding" color-map analysis is to investigate whether any Resolution II or Resolution II confounding concerns between any main effect. The confounding severity is indicated by color map (from 0% correlation in Blue to 100% correlation in Red). The diagonal is always in red color. In Figure 7b, there is very mild Resolution II confounding (correlation = 0.33) due to Categorical variables. Therefore, no severe Confounding concern was noticed.

3.2 DSD Optimization Result

Section 3.2 will provide the DSD results of Section 3.1 DSD execution. The objective of this DSD is to demonstrate the optimal transportation route to minimize two intra-city transportation goals: total duration and total expense. The DOE results were shown in Figure 8 JMP Profiler Analysis. Among five transportation segments, Amsterdam to Prague segment has shown the biggest impact to both Expense and Duration responses. Taking flight is a much better choice than taking train. The reason is there is no Euro Train from Amsterdam to Prague. There is also no direct train available between two cities. This train route is not very popular, and the train ticket is way more expensive than taking a flight.

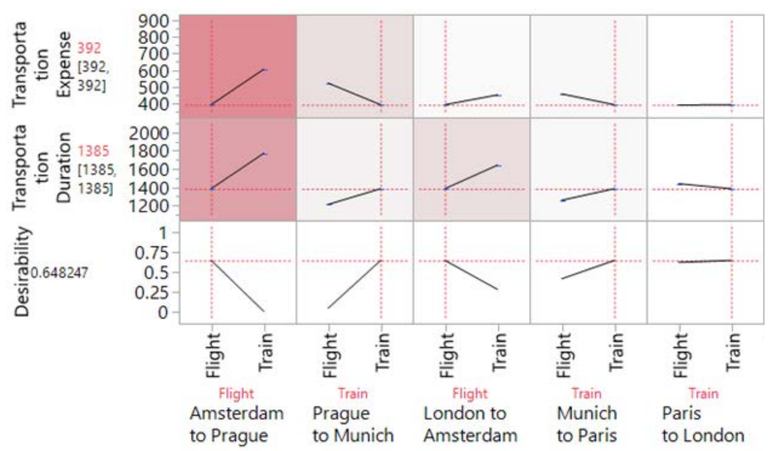

Figure 8: Profiler Sensitivity and Optimization

Next, there are two competing patterns between two responses on: (1) Prague to Munich, and (2) London to Amsterdam. Prague to Munich segment has more impact on the expense and taking direct train would significantly reduce the expense around $100 USD as compared to taking a flight. Instead, for the London-Amsterdam segment, taking a flight can shorten the transportation duration time by more than 200 minutes. The driving/flying distance ratio of London-Amsterdam route is 550km/358km ~ 1.54 higher than the typical 1.3 ratio in Europe. From London-Amsterdam, train needs to pass Paris first. The flight can be across Ocean. The fourth sensitive segment Munich to Paris also favors train choice to meet the expense requirement more. The last segment Paris to London favors train to meet the duration requirement more. There is no wonder most visitors would take Euro Star train across English Straight. The optimal design can achieve the expected expense at $392 USD (below $500 USD) and duration $1,385mins (below 2,880mins). Though, the overall optimal design can only meet both requirements at 65% desirability.

3.3 Monte Carlo Simulation

Section 3.2 optimal route design would not consider the noise factors such as duration distribution and ticket price distribution. Even within the same day at the same station, the transportation duration and the ticket price may be different from peak hours to off-peak hours. To accurately estimate the total transportation duration and expense, there is a need to consider all these uncertain noise factors. JMP Profiler Monte Carlo simulator is very powerful to simulation these random noise factors.

Monte Carlo simulation, or probability simulation, is a technique used to understand the impact of risk and uncertainty in financial, project management, cost, and other forecasting models. Their essential idea is using randomness to solve problems that might be deterministic in principle. They are often used in physical and mathematical problems and are most useful when it is difficult or impossible to use other approaches. Monte Carlo methods are mainly used in three problem classes [9,10] optimization, numerical integration, and generating draws from a probability distribution. Monte Carlo Simulation enables you to discover the distribution of model outputs as a function of the random variation in the factors and model noise. The simulator in the profilers provides a way to set up the random inputs and run the simulations, producing an output table of simulated values.

As shown in Figure 9a and 9b, JMP Profiler Simulators use Normal Distribution to simulate the two responses (9a Duration, 9b Expense) by including the variability of flight duration and ticket price. The distribution standard deviation was determined by calculating from the real duration/ticket price data. The two requirements are also input as upper specific limit (USL) for JMP Profiler to estimate the non-conforming defect rate. The Monte Carlo Simulation results have shown 0% probability of not meeting the 28.8hours duration requirement and 8% probability of not meeting $500 USD transportation budget. Even though the predicted expense is $392 (seems enough buffer from $500), when considering the noise ticket price distribution, there is still 8% chance that the total transportation expense may exceed $500 USD. To avoid this 8% risk, it may need to book the flights/trains earlier and not scheduled during the peak hours.

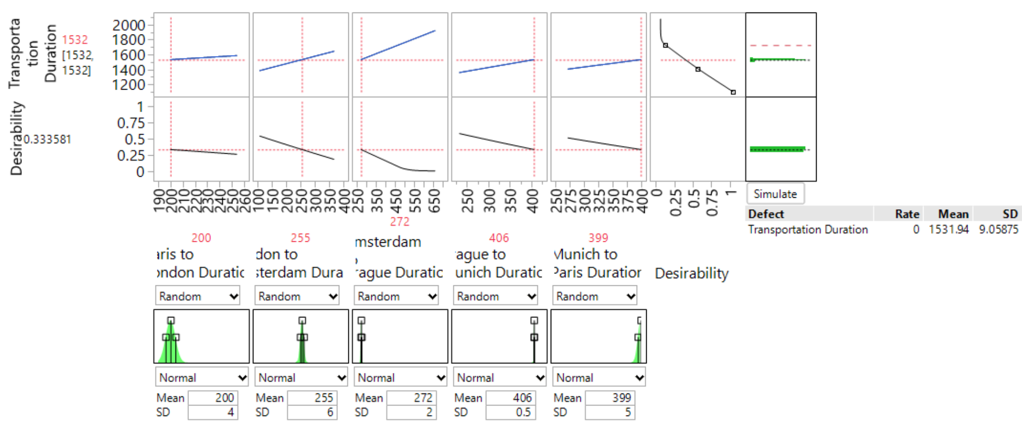

Figure 9a: Monte Carlo Simulation of Transportation Duration

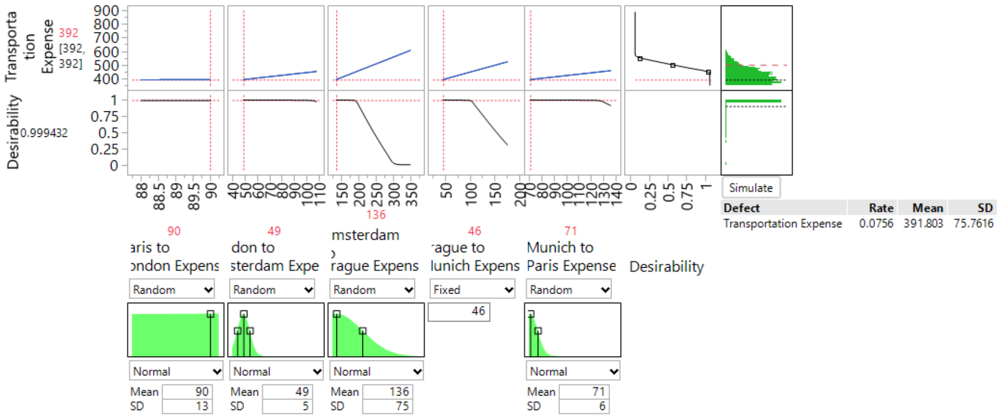

Figure 9b: Monte Carlo Simulation of Transportation Expense

3. Conclusions

This paper has demonstrated an effective "STEMS" methodology of managing Travel in Europe by studying the Transportation Systems in Europe. Designed a structured DOE and build predictive models of minimizing both the travel duration and expense. Monte Carlo simulation method was conducted by considering the variability of random flight/train duration and expense. Observed 8% risk probability of not meeting the transportation expense budget. The same methodology can be applied to travel in China, Japan, Taiwan where the High-Speed Train system is well established.

Acknowledgments

I would like to thank my statistics and math advisor Dr. Charles Chen for helping and supporting me throughout this STEMS project.

References

1. Mason C., (2018 July) "Multivariate Statistics of Antioxidant Chocolate", IWSM Bristol Proceedings, Vol 2 37-40
2. Mason C., (2018 July), "Choose Healthy Chocolate", IEOM Europe Proceedings, 434-441
3. Wu, Anna Dong. "Starbucks and Cardiovascular Disease Prevention." IEOM, IEOM Society, 26 July 2018, www.ieomsociety.org/paris2018/papers/424.pdf.
4. Errore, A., Jones, B., Li, W., and Nachtsheim, C. (2016), "Using Definitive Screening Designs to Identify Active First- and Second-Order Factor Effects," forthcoming, Journal of Quality Technology.
5. Jones, B. and Nachtsheim, C.J. (2011), "Efficient Designs with Minimal Aliasing," Technometrics, 53:1, 62-71.
6. Jones, B. and Nachtsheim, C.J. (2013), "Definitive Screening Designs with Added Two-Level Categorical Factors," Journal of Quality Technology, 45, 121-129.
7. Jones, B. and Nachtsheim, C.J. (2016), "Blocking Schemes for Definitive Screening Designs," Technometrics, 58:1, 74-83.
8. Miller, A. and Sitter, R. R. (2005), "Using Folded-Over Nonorthogonal Designs," Technometrics, 47(4), 502-513.
9. Anderson, Herbert L. (1986). "Metropolis, Monte Carlo and the MANIAC" (PDF). Los Alamos Science. 14: 96–108.
10. Box, G.E.P. and Draper, N.R. (1987), Empirical Model–Building and Response Surfaces, New York: John Wiley and Sons.