Analysis and Discussion on the Problem of Airport Taxi

Zihan Wu 1, Zihao Shi 1, Junchen Cui 2 and Hui Xu 2*

1 Department of Communication Engineering, Faculty of Engineering, Yanbian University, Yanji, China.
2 Department of Business School, Faculty of Electronic Commerce, Dalian University of Technology, Dalian, China.
*Corresponding author email id: 792586313@qq.com
Date of publication (dd/mm/yyyy): 19/11/2019

Abstract – Taking the problem of taxi in the airport as the background, this study mainly focuses on the optimal allocation of taxi resources in the airport. Through analyzing results of some related factors and using dendrogram to compare the benefit and loss of the scheme, a decision-making model is established to realize the "matching relationship" between taxi resources and flights in different time and space. On the basis of the collected data, the mean value model is established to solve the related variables. Basing on this model and principal component analysis (PCA), this paper uses MATLAB fitting and geometry software to analyze and test the model rationalization and the dependence of related factors.

Keywords – Decision-making Model, Mean Value Model, Principal Component Analysis, MATLAB Fitting.

I. INTRODUCTION

Taxi is one of the main means of transportation nowadays; however, taxi drivers may find it difficult to make the optimal decision because of the number of people in line, waiting time and the distance, and there are more uncertain factors need to be considered in reality. How to build a model and test its rationality based on these factors has become a difficult problem at present.

In recent years, researchers in many academic fields often measure the rationality of the simulated influencing factors of a scheme by data fitting. The core problem is to test the rationality of the decision-making model which is built by refining the most important factors from the comprehensive consideration of the relevant factors. In 2017, Huang [1] et al. and others constructed an index evaluation system from society, economy and environment, and used principal component analysis to evaluate the rationality of Xi’an’s urban ecological environment from 2009 to 2015. In 2017, researchers [2] proposed an "evolutionary" principal component analysis method which is based on the traditional principal component analysis method proposing classification rules based on Euclidean Distance. They take the actual load of users in a certain area as an example, and use cosine similarity theorem to fit all kinds of users’ curve shapes to verify the effectiveness of the proposed algorithm and examined the "evolutionary" principal component analysis method. In 2018, Meng [3] et al adopted improved PCA based on Spass indicating that Beijing improves the urban human-water harmony from social, economic and other related factors. The result was consistent with the actual situation of Beijing that determined that the model is reasonable and feasible.

In the background of airport taxi problem, we need to establish a decision-making model to explore whether the relevant factors of its influence mechanism can be used as conditions to simulate the decision-making model, and analyze the dependence of the relevant factors of the model. Now we need to solve a major problem:

How to choose the main influencing factors to establish the decision-making model which substitution variables of its factors in the decision - making model that can help drivers to make the optimal decision and maximize the
benefits.

We can divide the main question into two small tasks:

1. Comprehensively analyze the influencing factors of the model, the number of passengers and the driver's income, and then establish a decision-making model.

2. Collect the data of the taxi in an airport in a city, and establish the rationality of the model and the dependence of relevant factors.

II. MODEL ESTABLISHMENT AND SOLUTION

2.1. The Model of Problem 1 and Its Solution

Due to the different number of aircraft landing at the same time, the number of passengers is different. Moreover, the various time will also affect the proportion of passengers who choose to take a taxi. For example, in the midnight when public vehicles stop, more passengers will choose a taxi. In addition to the time that the driver has to wait for the number of passengers to reach a certain amount, the number of taxis released in each batch by the airport management personnel, the time consumed in each batch and the number of vehicles in front of the storage pool will affect the waiting time of the driver. Therefore, according to the above factors, this study establishes a model which the revenue of taxi drivers is represented by the profit per minute, which is regarded as a constant value. Then calculates and compares the time when the driver returns to the city from the airport and the waiting time of the driver in the storage pool ignoring the uncertain factors such as traffic jam. Finally, the decision is made by weighing the benefit and loss of the two schemes through dendrogram, and the rationality of the model will be proved in the second question.

For convenience, we choose to calculate the losses of waiting and not waiting respectively. Smaller the losses represent larger benefit.

<table>
<thead>
<tr>
<th>n₁</th>
<th>Number of taxis released in each batch</th>
<th>n₂</th>
<th>Average number of passengers per car</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>Average taxi passengers</td>
<td>D</td>
<td>Average D minutes a flight landing</td>
</tr>
<tr>
<td>m</td>
<td>Number of taxis waiting in the storage pool</td>
<td>t₁</td>
<td>Time required for each batch of passengers to get on the taxi</td>
</tr>
<tr>
<td>t₂</td>
<td>Drivers’ total waiting time</td>
<td>x</td>
<td>The average amount of money a driver earns per minute</td>
</tr>
<tr>
<td>T</td>
<td>Time taken from airport to downtown</td>
<td>Z</td>
<td>Amount of loses</td>
</tr>
</tbody>
</table>

Among them, n₁, n₂, K, t₁, x, T and so on are the exact numbers which can be determined by the relevant statistical data of the airport, while D and m are the main uncertainties that drivers consider when making decisions.

The total waiting time of the drivers t₂ is composed of two parts. The first part is the time for waiting for the number of passengers to reach a certain amount, that is to say, the number of passengers n₂ should meet the number of people when each batch of taxis n₁ is released. The second part is the time for the total number of passengers to get on the taxi (Time reconnected is not taken into account).

Goal: compare Z_wait with Z_not_wait.
\[ t_2 = \left(\frac{n_1 \times n_2 \times D + t_1}{k}\right) \left(\frac{m}{n_1} + 1\right) \]  
(1) 

\[ Z_{\text{wait}} = t_2 \times x \]  
(2) 

\[ Z_{\text{not wait}} = T \times x \]  
(3) 

So we can build the tree diagram model as follows

![Tree diagram model](image)

Fig. 1. Tree diagram model.

2.2. The Model of Problem 2 and its Solution

First, we need to calculate the average number of passengers who choose to take a taxi in each flight K.

According to the inquiry, a civil aviation aircraft has about 100-200 seats. In this paper, 150 is selected as the maximum passengers per aircraft.

<table>
<thead>
<tr>
<th>Airline</th>
<th>Route</th>
<th>Sortie</th>
<th>Passenger throughput</th>
<th>Inbound</th>
<th>Outbound</th>
<th>Stopover</th>
<th>Passenger load factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA</td>
<td>Beijing-Zhengzhou-Shaoyang</td>
<td>688</td>
<td>67012</td>
<td>31263</td>
<td>35749</td>
<td>57.06</td>
<td></td>
</tr>
<tr>
<td>AQ</td>
<td>Changsha-Shaoyang-Haikou</td>
<td>1362</td>
<td>158698</td>
<td>43546</td>
<td>115152</td>
<td>70614</td>
<td>61.98</td>
</tr>
<tr>
<td>KY</td>
<td>Kunming-Shaoyang-Hangzhou</td>
<td>790</td>
<td>76813</td>
<td>27273</td>
<td>49540</td>
<td>23146</td>
<td>6404</td>
</tr>
<tr>
<td>G5</td>
<td>Chongqing-Shaoyang-Changsha</td>
<td>1326</td>
<td>82397</td>
<td>36145</td>
<td>46252</td>
<td>10426</td>
<td>69.71</td>
</tr>
<tr>
<td>JR</td>
<td>Changsha-Shaoyang</td>
<td>252</td>
<td>8505</td>
<td>4163</td>
<td>4342</td>
<td>66.45</td>
<td></td>
</tr>
<tr>
<td>9C</td>
<td>Shenzhen-Shaoyang-Lanzhou</td>
<td>626</td>
<td>65214</td>
<td>13157</td>
<td>52057</td>
<td>39497</td>
<td>57.38</td>
</tr>
<tr>
<td>MU</td>
<td>Xian-Shaoyang</td>
<td>20</td>
<td>1007</td>
<td>410</td>
<td>597</td>
<td>31.43</td>
<td></td>
</tr>
<tr>
<td>9C</td>
<td>Shenzhen-Shaoyang-Shenyang</td>
<td>100</td>
<td>10048</td>
<td>2151</td>
<td>7897</td>
<td>6351</td>
<td>55.48</td>
</tr>
</tbody>
</table>

By substituting the data in Table 1 into the mean value model

\[ F = \frac{\sum_{i=0}^{n} f_i}{n} \]  
(4) 

It can be found that under normal conditions, the average passenger load factor rate of each aircraft is \( F = 57.94\% \) (where \( f_i \) represents the occupancy rate of a certain aircraft), and the number of frequent passengers on each aircraft is \( F \times 150 = 86.91 \). 

There is no airport express line connecting the city center in the roadside transportation system of Pudong airport. According to the data in 2018, the public transport dominated by Metro Line 2, maglev train and airport
shuttle bus has insufficient attraction due to service time, congestion, ticket price and other reasons, with a share rate of about 31.6%, resulting in a high proportion of individualized transport in Pudong Airport, with private cars, taxis and Internet buses accounting for 54.6%. Although taxis have decreased in recent years due to the rapid growth of online car-hailing, they still share 16.5% of the total number of passengers in the mode of passenger distribution. The average number of taxis per day is nearly 10000.

Here we select 16.5% as the proportion of all passengers who choose taxi. After calculation, it is found that each aircraft will bring \(86.91 \times 16.5\% = 14.34\) passengers who choose to take a taxi. For the convenience of calculation, 14 people are selected here. That is, in the model of question one, \(K = 14\).

The following two figures indicate the trend of passenger flow with time in the taxi waiting area of Beijing Capital International Airport.

![Fig. 2. Chart of passengers flow (20160322-20160324) (Abscissa: time ordinate: number of people).](image)

It can be seen from the figure that although the number of people in the peak every day is different, the general trend of the passenger flow in the taxi waiting area is almost the same. There will be a cruciality between 8:00 and 9:00. The peak hours of the passenger flow in the airport taxi waiting area are 13:00, 16:00, 19:00 and 22:00.

However, the most direct factor influencing the passenger flow is the number of flights arriving at the airport at various times. The taxi waiting area will be easier to gather more passengers after flight landing. Therefore, the number of passengers in airport taxi waiting area can be analyzed in combination with the arrival of airport passengers.

![Fig. 3. Number of arriving flights at Beijing Capital International Airport in each period (Abscissa: time ordinate: number of people).](image)
Figures above reflect that the trend of the number of arrivals and taxi passenger flow is basically consistent, so we can believe that the main factor determining the passenger flow is the number of arrivals.

According to figure 4, we can find that \( Z_{\text{not wait}} = 81 \), and it takes 34 minutes \((T)\) to get to the urban, and the driver can earn about 2.3 yuan \((x)\) per minute.

At the same time, through the data collection, we know that the average time of each batch of taxis is about 0.17 minutes, and the average number of passengers per taxi is 1.28.

According to the data collected and calculated above, we can simplify the model of question 1.

Goal: compare with \( Z_{\text{wait}} \) and \( Z_{\text{not wait}} \)

\[
t_2 = \left(\frac{m \times n_k}{k}\right) \times D + t_1 \left(\frac{m}{n_k} + 1\right)
\]

\[
= \left(\frac{8 \times 1.28}{14}\right) \times D + 0.17 \left(\frac{m}{8} + 1\right)
\]

\[
Z_{\text{wait}} = t_2 \times x = 2.3 \times t_2
\]

\[
Z_{\text{not wait}} = T \times x = 81
\]

Table 2. Flight volume of corresponding period (this table is integrated into one page table)

<table>
<thead>
<tr>
<th>Time period, flight quantity, flight interval</th>
<th>Time period, flight quantity, flight interval</th>
<th>Time period, flight quantity, flight interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>00:00-01:00</td>
<td>20</td>
<td>3min</td>
</tr>
<tr>
<td>01:00-02:00</td>
<td>30</td>
<td>2min</td>
</tr>
<tr>
<td>02:00-03:00</td>
<td>26</td>
<td>2.3min</td>
</tr>
</tbody>
</table>
According to the data in Table 2, it can be as low as 0.7 minute flight in the peak period and about 3 minute flight in the trough period. Therefore, we can analyze it in two cases.

### Table 3. Comparative Analysis of earnings in peak and low period

<table>
<thead>
<tr>
<th>Time period, flight quantity, flight interval</th>
<th>Time period, flight quantity, flight interval</th>
<th>Time period, flight quantity, flight interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>03:00-04:00</td>
<td>7</td>
<td>8.7min</td>
</tr>
<tr>
<td>04:00-05:00</td>
<td>1</td>
<td>60min</td>
</tr>
<tr>
<td>05:00-06:00</td>
<td>8</td>
<td>8.7min</td>
</tr>
<tr>
<td>06:00-07:00</td>
<td>67</td>
<td>0.9min</td>
</tr>
<tr>
<td>07:00-08:00</td>
<td>84</td>
<td>0.7min</td>
</tr>
<tr>
<td>08:00-09:00</td>
<td>77</td>
<td>0.77min</td>
</tr>
<tr>
<td>09:00-10:00</td>
<td>46</td>
<td>1.3min</td>
</tr>
<tr>
<td>10:00-11:00</td>
<td>57</td>
<td>0.95min</td>
</tr>
<tr>
<td>11:00-12:00</td>
<td>78</td>
<td>0.77min</td>
</tr>
<tr>
<td>12:00-13:00</td>
<td>47</td>
<td>1.3min</td>
</tr>
<tr>
<td>13:00-14:00</td>
<td>56</td>
<td>1.1min</td>
</tr>
<tr>
<td>14:00-15:00</td>
<td>51</td>
<td>1.2min</td>
</tr>
<tr>
<td>15:00-16:00</td>
<td>56</td>
<td>1.1min</td>
</tr>
<tr>
<td>16:00-17:00</td>
<td>61</td>
<td>1min</td>
</tr>
<tr>
<td>17:00-18:00</td>
<td>65</td>
<td>0.9min</td>
</tr>
<tr>
<td>18:00-19:00</td>
<td>63</td>
<td>1min</td>
</tr>
<tr>
<td>19:00-20:00</td>
<td>57</td>
<td>1min</td>
</tr>
<tr>
<td>20:00-21:00</td>
<td>32</td>
<td>2min</td>
</tr>
<tr>
<td>21:00-22:00</td>
<td>20</td>
<td>3min</td>
</tr>
<tr>
<td>22:00-23:00</td>
<td>29</td>
<td>2min</td>
</tr>
<tr>
<td>23:00-24:00</td>
<td>2</td>
<td>30m</td>
</tr>
</tbody>
</table>

**Peak D = 0.7**
- **Z_{wait} = Z_{not\ text{wait}} (m = 403)**  
  The income of waiting are the same as those of not waiting
- **Z_{wait} > Z_{not\ text{wait}} (m >403)**  
  The income of waiting are less than those of not waiting
- **Z_{wait} < Z_{not\ text{wait}} (m<403)**  
  The income of waiting are more than those of not waiting

**Trough D = 3**
- **Z_{wait} = Z_{not\ text{wait}} (m=111)**  
  The income of waiting are the same as those of not waiting
- **Z_{wait} > Z_{not\ text{wait}} (m >111)**  
  The income of waiting are less than those of not waiting
- **Z_{wait} < Z_{not\ text{wait}} (m<111)**  
  The income of waiting are more than those of not waiting
2.2.1 Rationality of Analysis Model and Dependence on Relevant Factors

In the Figure 1, the main factors affecting the driver’s interests are the average time \( D \) arrival flight and the number of existing taxis \( m \) in the storage pool. In order to verify the rationality of the model, the PCA method is selected to verify whether the above two factors are the main factors.

First of all, in addition to the average \( D \) minute arrival flight and the number of existing taxis in the storage pool \( m \), there are also factors such as the average number of passengers per taxi \( n_2 \), the time required for each batch of passengers to get on the train \( t \) and so on. These four factors were analyzed by PCA.

Step 1: Select two influencing factors, and standardize each index component and transform the formula:

\[
X_{ij} = \frac{Y_{ij} - \bar{Y}_i}{S_i} \quad (j=1,2,...,n)
\]

(9)

\[
\bar{Y}_i = \frac{1}{n} \sum_{k=1}^{n} Y_{ki}
\]

(10)

**Sample Standard Deviation:**

\[
S_i = \sqrt{\frac{1}{n-1} \sum_{k=1}^{n} (Y_{ki} - \bar{Y}_i)^2}
\]

(11)

Step 2: take the standardized two factors as X1 and X2 to make the scatter diagram.

Step 3: Rotate axes X1 and X2 by 45 °, and draw new axes Y1 and Y2. Observe the variance of the two factors on coordinate axes Y1 and Y2.

Table 4. Flight interval \( D \) and number of taxis in the storage pool \( m \).

<table>
<thead>
<tr>
<th>Number of taxis</th>
<th>Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.20292</td>
<td>-0.21156</td>
</tr>
<tr>
<td>-1.18181</td>
<td>-0.29127</td>
</tr>
<tr>
<td>-0.43235</td>
<td>-0.26736</td>
</tr>
<tr>
<td>-0.51679</td>
<td>0.242845</td>
</tr>
<tr>
<td>-1.19236</td>
<td>4.33245</td>
</tr>
<tr>
<td>-1.30847</td>
<td>-1.72065</td>
</tr>
<tr>
<td>-1.21347</td>
<td>-0.39491</td>
</tr>
<tr>
<td>-0.749899</td>
<td>-0.38933</td>
</tr>
<tr>
<td>2.206594</td>
<td>-0.34708</td>
</tr>
<tr>
<td>1.74214</td>
<td>-0.37498</td>
</tr>
<tr>
<td>0.243222</td>
<td>-0.38933</td>
</tr>
<tr>
<td>-0.5379</td>
<td>-0.34708</td>
</tr>
<tr>
<td>-0.36901</td>
<td>-0.36302</td>
</tr>
</tbody>
</table>
Table 4 shows the data of flight interval D and the number of taxis m in the storage pool after the standardization transformation.

<table>
<thead>
<tr>
<th>Number of taxis</th>
<th>Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.116553</td>
<td>-0.35505</td>
</tr>
<tr>
<td>0.570451</td>
<td>-0.36302</td>
</tr>
<tr>
<td>-0.062889</td>
<td>-0.37099</td>
</tr>
<tr>
<td>1.6788.6</td>
<td>-0.37099</td>
</tr>
<tr>
<td>1.953256</td>
<td>-0.29127</td>
</tr>
<tr>
<td>0.359336</td>
<td>1.940868</td>
</tr>
</tbody>
</table>

From Fig. 4, we can analyze that the impact of flight interval D and the number of waiting taxis in the storage pool on decision-making is basically the same.

Fig. 4. (Left) Horizontal axis: flight interval D vertical axis: number of taxis in car storage pool m (right) horizontal axis: flight interval D vertical axis: time required for each batch of passengers to board t1.

Fig. 5. (Left) Horizontal axis: flight interval D vertical axis: Average number of passengers per car n2 (right) horizontal axis: number of taxis in the storage pool vertical axis: time required for each batch of passengers to get on the taxis t1.
Figure 5 reflects that flight interval $D$ has a greater impact on decision-making than the time $t$, required for each batch of passengers to get on the taxis, and flight interval $D$ has a greater impact on decision-making than the average number of passengers $n_2$ per taxi.

![Figure 5](image)

**Fig. 5.** Number of taxis waiting in the storage pool $m$ vertical axis: Average number of passengers per car $n_2$.

Figure 6 reflects that the number of taxis $m$ in the storage pool has a greater impact on decision-making than the average number of passengers $n_2$ per taxi. From the above results of PCA, we know that the number of taxis $m$ and the flight interval $D$ in the car storage pool are indeed the main factors affecting the decision, and we can also draw the conclusion that the number of taxis $m$ in the car storage pool has a slightly greater impact on the decision-making than the flight interval $D$. Therefore, it is reasonable to regard the number of taxis $m$ and the flight interval $D$ in the car storage pool as the main influencing factors of the model in question 1.

Only the factors such as the capacity of each flight, the number of taxis released in each batch, the distance from the city, the profit of taxi driver per minute, etc. are given that the model could be solved. Therefore, this model highly depends on the factors.

### III. Conclusion

Based on the airport taxi problem, this paper systematically studies the number of taxi passengers \(^4\) and the impact of the relevant factors we discussed, and how to establish a dendrogram model so that drivers can get an optimal decision judging by the waiting time and relative interests. After we set up the dendrogram model, we calculate the average passenger load rate $F$ of each aircraft and the average number of passengers $K$ who choose to take a taxi in each flight under normal conditions by collected data. According to the mean value model, and then we can get the taxi passenger flow, the average time consumption of each batch of taxis, the average number of passengers per taxi, the average number of flights inbounding and the storage of taxis. By substituting these variables into the formula, model I can be simplified, and the situation can be divided into peak and low period. In these two cases, the comparison of the result data can help drivers make decisions with large relative interests.

In the above process, in order to verify the rationality of the model, we accurately verified that two factors ($D$, $m$) are the main factors of the model and analyzed the dependence of relative factors through data fitting and PCA.

The model established in this paper can be extended to large cities \(^5\). For instance, through mobile APPs, we can know that the average $D$ minute landing flight. At the same time, in large city, the number of waiting taxis $m$
are displayed at the car storage pool [6], so we can refer to the decision model established by us for these two variables to make the optimal decision.

REFERENCES


AUTHOR’S PROFILE

Zihan Wu. 1998. Study in the department of communication engineering of Yanbian University (Yanj, China). email id: 792586313@qq.com

Hui Xu, Associate Professor and master tutor of mathematics department, School of science, Yanbian University. Research Interests: Mathematics Education Technology, Mathematical Modeling and Intelligent Computing.

Zihao Shi, Department of Communication Engineering, Faculty of Engineering, Yanbian University, Yanji, China.

Junchen Cui, Department of Business School, Faculty of Electronic Commerce, Dalian University of Technology, Dalian, China.