Analysis on the Correlation of Traffic Flow in Hainan Province Based on Baidu Search

To cite this article: Caixia Chen and Chun Shi 2018 IOP Conf. Ser.: Earth Environ. Sci. 128 012055

View the article online for updates and enhancements.
Analysis on the Correlation of Traffic Flow in Hainan Province Based on Baidu Search

Caixia Chen, Chun Shi
Department of E-commerce, School of Information Science and Technology, Hainan Normal University, Haikou 571158, P.R. China.
email: 914922505@qq.com

Abstract: Internet search data records user's search attention and consumer demand, providing necessary database for the Hainan traffic flow model. Based on Baidu Index, with Hainan traffic flow as example, this paper conduct both qualitative and quantitative analysis on the relationship between search keyword from Baidu Index and actual Hainan tourist traffic flow, and build multiple regression model by SPSS.

1. Introduction
Baidu Index is a data sharing platform based on huge amount of Baidu users' behaviour data. Through Baidu Index, we can acquire the actual search volume of relevant keyword with certain Baidu Index value. We build the consumer tourist behavior model based on the travel destination selection process, then search the primary keyword from the model by Baidu Index recommendation function to acquire corresponding keyword search volume. Finally, we build multiple regression model to analyze the relationship between the search index and the historical tourist flow of Hainan from 2014 to 2016 provided by Hainan Bureau of Statistics, and draw the conclusion.

2. Keyword selection
According to the relevant analysis of the degree of concern to the tourist destination before the tourists travel, the key points related to the destination (Hainan) traffic flow are selected according to the name of the tourist destination (Hainan), the actual route of travel, the destination of tourism destination and other tourist destinations word. According to this, this article selected "Hainan", "Hainan ticket(HNTI)", "Sanya ticket(SYTI)", "Hainan high-speed rail(HNRH)", "Hainan car rental(HNCR)", "Hainan traffic" as the benchmark key words, and then these benchmark keywords Baidu index search, Step to find out the actual search volume of these keywords, the second step to find out the relevant keywords, the third step, the use of webmaster tools such as the test, screening out the amount of search accounted for the first 80% of the key words: "Hainan high-speed rail", "Hainan ticket (Haikou + Sanya)", "Hainan car rental" 3 key words. According to this, the influence of the search volume of these three keywords on the traffic flow in Hainan is revealed, and the influence of the network search volume on the actual traffic flow in Hainan is revealed.

3. Data collection and processing
Baidu index tools were used to collect three keywords from January 1, 2011 to December 17, 2016, the actual daily search volume and monthly search volume. Through the Hainan Statistical Bureau, "Hainan Statistical Monthly" query to 2014, 2015, 2016 three years of Hainan actual number of tourist volume,
the above all the data into the SPSS software system, obtained Figure 1 "Comparison of the actual number of tourists in Hainan (ATHN) and the actual search volume of Baidu keyword (ASHN)"

![Figure 1 Comparison of the actual number of tourists in Hainan and the actual search volume of Baidu keyword](image)

**Figure 1** Comparison of the actual number of tourists in Hainan and the actual search volume of Baidu keyword

Obviously, the actual search volume of Baidu index related keywords and Hainan tourism tourists the actual throughput of the trend is basically the same, with relevance.

4. Analysis of Traffic Flow Correlation

The first step, the data for descriptive analysis, mainly to determine whether the data collected is stable (Table 1).

| Table 1. Descriptive Statistics |
|---------------------------------|
|                                | N | Minimum | Maximum | Mean | Std |
| Hainan tourist volume           | 36| 241.6    | 641.2   | 365.50 | 89.396 |
| HNRH                           | 36| 45.3      | 76.6    | 57.89 | 8.376 |
| HNTI                           | 36| 125.3     | 291.7   | 195.53 | 35.484 |
| HNCR                           | 36| 27.8      | 53.1    | 33.50 | 5.629 |
| N of Valid Cases               | 36|           |         |       |     |

The second step, the data summary, mainly to determine the degree of fitting the model (Table 2).

| Table 2. Model summary         |
|---------------------------------|
| Model                          | R Square | Adjusted R Square | Std Error of Estimate | R Square Change | F Change | df1 | df2 | Sig. F Change |
| Regression                     | .949a     | .900              | .891                 | .900            | 96.097    | 3   | 32  | .000         |

*Predictive variables: (constant), car rental Hainan, high-speed rail Hainan, ticket.

The third step, the variables of multiple regression analysis, analysis of variance (Table 3).

| Table 3. ANOVAb                |
|---------------------------------|
| Model                          | Sum of Squares | df | Mean Square | F | Sig. |
| Regression                     | 251763.511     | 3  | 83921.170   | 96.097 | .000a |
| Residual                       | 27945.489      | 32 | 873.297     |      |      |
5. Establish traffic flow forecasting model

Through the analysis of a large number of the above data, and through the application of SPSS software, the establishment (Table 4, Table 5):

Table 4. Coefficients*

| Model          | Unstandardized Coefficients | Collinearity diagnostics |
|----------------|-----------------------------|--------------------------|
|                | B   | Std Error | t   | Sig. | Tolerance | VIF |
| (Constant)     | -186.920 | 36.782 | -5.082 | .000 |           |     |
| HNRH           | 2.086 | .918 | 2.273 | .030 | .422 | 2.368 |
| HNTI           | .984 | .279 | 3.534 | .001 | .255 | 3.915 |
| HNCR           | 7.140 | 1.475 | 4.842 | .000 | .362 | 2.762 |

Table 5. Collinearity diagnostics*

| Model | Dimension | Eigen value | Condition Index | Variance Propontions |
|-------|-----------|-------------|-----------------|-----------------------|
|       | (Constant) | HNRH | HNTI | HNCR |
| 1     | 3.970     | 1.000 | .00 | .00 | .00 |
| 2     | .018      | 15.038 | .67 | .00 | .11 | .06 |
| 3     | .008      | 22.019 | .08 | .50 | .03 | .53 |
| 4     | .004      | 29.867 | .25 | .50 | .87 | .41 |

* Dependent variable: Hainan passenger throughput

Through the above analysis, we can draw the multiple regression coefficients, the establishment of the following multiple regression model:

\[ Y = -186.92 + 2.086 \times \text{GTHN} + 0.984 \times \text{JPHN} + 7.140 \times \text{ZCHN} \]  \hspace{1cm} (1)

Among them, \( Y \) on behalf of Hainan tourism tourists forecast throughput, \( \text{GTHN} \) on behalf of Baidu Index Key words "high-speed rail Hainan", \( \text{JPHN} \) on behalf of Baidu Index Keyword "ticket Hainan", \( \text{ZCHN} \) on behalf of Baidu index keyword "car rental Hainan".

In the "Hainan high-speed rail", "Hainan ticket" and "Hainan car rental" through the significance test, at 5% significance level, I believe that "Hainan high-speed rail", "Hainan ticket" and "Hainan car" can be through the statistical T test, so that "Hainan high-speed rail", "Hainan ticket" and "Hainan car rental" on the Hainan passenger throughput there is a significant impact.

From the collinear diagnosis table, we can see that the tolerance of "Hainan high iron", "Hainan ticket" and "Hainan car rental" is more than 0.1, and the variance expansion factor is less than 5, so there can be no multi-collinearity. From the above results, we can see that when the other conditions of the same conditions, the "high-speed rail Hainan" indicators for each unit of growth, Hainan passenger throughput increased by 2.086 units; "ticket Hainan" indicators for each unit, Hainan passenger...
throughput growth 0.984 units; "Hainan rental car" indicators for each unit of growth, Hainan passenger throughput increased by 7.14 units.

6. Analysis conclusion
Using tourist flow of Hainan as the object, based on the actual tourist flow of Hainan and the Baidu index search data of keyword including “Hainan high-speed rail”, “Hainan airline ticket”, “Hainan car rental”, this paper analyze the relationship between the tourist flow of Hainan and Baidu keyword with econometrics theory, and reach the conclusion through the designed model.

Based on the results of the above regression analysis, we can draw the following conclusions:

(1) "Hainan high-speed railway" Baidu Index search volume and Hainan tourist flow presented a significant positive correlation, with correlation coefficient of 2.086. We can assume that "Hainan high-speed" Baidu Index search volume drive Hainan tourist flow to move positively. The higher "Hainan high-speed" Baidu Index search volume is, the higher is the Hainan tourist flow. On the contrary, the lower "Hainan high-speed" Baidu Index search volume is, the lower is the Hainan tourist flow.

(2) "Hainan airline ticket" Baidu Index search volume and Hainan tourist flow presented a significant positive correlation, with correlation coefficient of 0.984. We can assume that "Hainan airline ticket" Baidu Index search volume drive Hainan tourist flow to move positively. The higher "Hainan airline ticket" Baidu Index search volume is, the higher is the Hainan tourist flow. On the contrary, the lower "Hainan airline ticket" Baidu Index search volume is, the lower is the Hainan tourist flow.

(3) "Hainan car rental" Baidu Index search volume and Hainan tourist flow presented a significant positive correlation, with correlation coefficient of 2.086. We can assume that "Hainan car rental" Baidu Index search volume drive Hainan tourist flow to move positively. The higher "Hainan car rental" Baidu Index search volume is, the higher is the Hainan tourist flow. On the contrary, the lower "Hainan car rental" Baidu Index search volume is, the lower is the Hainan tourist flow.

Acknowledgments
Thanks to the support by NSF of Hainan Province (No.617110 and No.617121), National Natural Science Foundation of China (61362016) and National special project of international cooperation in science and technology (2014DFA13140).

References
[1] Wei Tengzhou. Research and Simulation of Urban Traffic Flow Forecasting Model [J]. Machinery Research and Application, 2016, (02): 112-114.
[2] Zhou Zhenglin, Zhang Xin. Research on traffic flow forecasting model based on artificial neural network [J]. Information Technology, 2006, (10): 36-38.
[3] Yang Jun, pan, Jia Limin. Traffic Flow Forecasting Model Based on Least Squares Support Vector Machine [J]. Journal of Beijing Jiaotong University, 2011, (02): 114-117+136.
[4] Guo min. Research on the Correlation between Baidu Index and Visitor Volume Based on SPSS [J]. Contemporary Economy, 2015, (29): 136-137.
[5] Cui Tobest. Large-data age under the background of the brand car sales forecast empirical study [D]. Henan University, 2014.
[6] Li Shixia, Tian zhi mei. An Analysis of Factors Affecting Network Attention in Tourism Destination Based on Baidu Index - Taking Qingdao as an Example [J]. Journal of Capital Normal University (Natural Science Edition), 2014, (01): 56-59.
[7] Huangxiankai, Zhang Lifeng, Ding Si. A Study on the Relationship between Baidu Index and Tourist Area Visitors - A Case Study of Beijing Imperial Palace [J], 2013, (11): 93-100.
[8] Andrew P. Davidson, Yin Mi. The internet and the occidental tourist: Analyses of Taiwan’s tourism website from the perspective of western tourists [J]. Information Technology & Tourism, 2005, Vol.7, pp.91-102.
[9] Maria Lexhagen. The importance of value-added services to support the customer search and purchase process on travel websites [J]. Information Technology & Tourism, 2005, Vol. 7 pp. 119-135.

[10] Ginsberg, Mohebbi, Patel, Brammer, Smolinski, Brilliant. Detecting influenza epidemics using search engine query data [J]. Nature, 2009, 457: 1012-1014.