International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-17

https://doi.org/10.70917/ijcisim-2026-0145
Article

Multiple Regression Analysis for Tourism Management: A Path of Integration between Tourist Satisfaction Enhancement and Business Operations

Yonghe Yang *

School of Tourism and Economic Management, Lijiang Normal University, Lijiang, Yunnan, 674100, China; lj.0888@163.com

Abstract: In recent years, the tourism industry has gradually developed into a strategic pillar industry of the national economy and has become one of the important industries for development in China. In order to help tourism management, the study utilized socio-economic data and selected 13 indicators of tourism development driving factors. Principal component analysis was applied to extract the main driving factors affecting tourism development. A multiple linear regression model was used to study the interrelationship between tourism development and various driving factors, and the results of principal component analysis were verified by the goodness-of-fit test and the significance test of regression coefficients. The test of the fitted linear regression coefficient of determination $R^2 = 0.994$ indicates that the model fits the data well. The quantitative model of tourism development was established as $F1 = -6.18012 + 2.083ZLX1 + 0.496LX2 + 0.145ZLX3 + 1.154ZLX4$. That is, for every 1% increase in socio-economic conditions, tourism service capacity, tourism resource attractiveness and infrastructure development, tourism development increases by 2.083%, 1.154%, 0.496% and 0.145% respectively. In addition, the study starts from the perspective of business operation and proposes to enhance employee training, improve employee welfare and good customer relations as a way to improve tourist satisfaction.

Keywords: Principal Component Analysis; Multiple Linear Regression; Tourism Management; Tourist Satisfaction

1. Introduction

In 2023, a series of highly anticipated tourist hotspots quickly emerged, including the Zibo tourism boom led by the social media sensation "Zibo barbecue" and the overnight fame of Guizhou's Taipan Village due to "Village BA" [1-2]. Meanwhile, tourism trends have diversified, shifting from fast-paced, high-intensity travel patterns to slower-paced, immersive experiences, and even to rural tourism aimed at escaping stress and seeking spiritual fulfillment. These trends not only integrate destinations with elements like cuisine and sports events to create must-visit spots but also reflect the public's new expectations and strong demand for travel in the post-pandemic era [3-6]. According to survey data from the Ministry of Culture and Tourism, China saw 4.891 billion domestic tourist trips in 2023, an increase of 2.361 billion compared to the same period last year, representing a year-on-year growth of 93.3% [7]. These figures clearly indicate that China's tourism industry is rapidly recovering, the tourism market is undergoing a resurgence, and it possesses immense development potential and remarkable resilience.

Tourism satisfaction evolved from customer satisfaction, which is defined as the psychological state formed by the difference between a customer's psychological expectations of a product or service and their actual feelings after consumption. If the actual feelings exceed expectations, the customer is satisfied; if expectations are high but the actual experience falls short, the customer is dissatisfied [8-11]. Research has found that satisfaction further influences the willingness to make repeat purchases [12]. American scholar Pizam, A. first proposed the concept of tourist satisfaction, thereby laying the theoretical foundation for research in this field. He argued that tourist satisfaction is consumers' evaluation of their travel experience based on a comparison of expectations in a tourism context [13]. Nguyen, Q. N. et al. defined satisfaction as tourists' emotional state after experiencing a trip, and that evaluating tourism satisfaction is a post-purchase process [14]. Pearce, P. L., and Moscardo, G. argue that tourist satisfaction with scenic areas is significantly influenced by multiple factors, including the reasonableness of ticket and service facility pricing, the completeness of rest and entertainment facilities within the scenic area, and the diversity and quality of services provided by the scenic area [15].

In quantitative research on tourist satisfaction, Abdou, A. H. et al. explored the relationship between perceived service quality, tourist satisfaction, and behavioral intentions in tourism during the COVID-19 pandemic. This study used questionnaire surveys and structural equation modeling for quantitative analysis, finding that perceived service quality has a positive impact on tourist satisfaction [16]. Syakier, W. A. and Hanafiah, M. H. focused on the influence of tour guide behavior on tourist satisfaction and behavioral intentions, Through surveys of international tourists, they found that tour guides' professional competence and service attitude have a positive impact on tourist satisfaction and indirectly promote tourists' willingness to revisit and travel word-of-mouth [17]. Nguyen, T. N., and Huynh, V. D. used regression models to quantitatively analyze the relationship between the tourism environment and tourist satisfaction, finding that cultural, socioeconomic, and natural environments have positive but gradually diminishing effects on satisfaction, contributing 38.9%, 33.4%, and 27.7% to satisfaction, respectively [18]. Suwannakul, E., et al. used structural equation modeling to quantitatively analyze the impact of four dimensions—traffic diversion, convenience, airport functional efficiency, and accessibility—on tourist satisfaction and revisit intentions [19]. Haron, H. I. C et al. (2023) conducted a quantitative analysis using SPSS based on the premise that tourist satisfaction mediates the relationship between educational institutions and tourism organizations, and found that tourist satisfaction between educational institutions and tourism organizations significantly influences revisit intention [20]. As evidenced by the above literature, quantitative assessments of tourist satisfaction can be conducted using various tools and methods, including structural equation models and regression analysis models [21].

This paper first utilizes socio-economic data and selects 13 indicators to reflect tourism development. Then principal component analysis is applied to extract the main driving factors affecting tourism development. Then, with tourism development as the dependent variable and socio-economic conditions, attractiveness of tourism resources, infrastructure construction and public service level, and tourism service capacity as the independent variables, a multiple linear regression analysis was conducted to explore the effects of the influencing factors, and the variables were tested for smoothness using Eviews7. Finally, the related strategies to enhance tourists' satisfaction are proposed in three aspects, namely, staff training, staff welfare treatment and customer relationship handling, respectively.

2. Studies on Factors Influencing Tourism Development

2.1. Research Methodology

2.1.1. Principal Component Analysis

(1) Basic Principles and Ideas

a) Basic idea

Principal component analysis is to take the method of mathematical dimensionality reduction to find out a few comprehensive variables to replace the original numerous variables, so that these comprehensive variables can represent the amount of information of the original variables as much as possible, and are not related to each other. This method of statistical analysis, which reduces multiple variables to a few composite variables that are uncorrelated with each other, is called principal component analysis [22]. What principal component analysis does is to try to replace the original variables by a new set of uncorrelated composite variables by regrouping the original variables that have some correlation. If the first linear combination selected, i.e. the first composite variable, is denoted as F_1, and "information" is measured by the variance, the larger the $Var(F_1)$, the more information is contained in F_1. Therefore the F_1 selected among all linear combinations should have the largest variance, so F_1 is called the first principal component. If the first principal component is not enough to represent all the information of the original p variables, then consider selecting F_1 i.e., the second linear combination, in order to effectively reflect the original information, the existing information of F_1 does not need to appear in F_1, which in mathematical language is to require $Cov(F_1, F_2) = 0$, i.e., F_2 is the second principal component, and so on the third, fourth, ... and p th principal components can be constructed.

b) Basic principles

According to the definition of the mathematical model of principal component analysis, in order to perform principal component analysis, it is necessary to derive the principal component coefficients in order to obtain the principal component model based on the original data as well as the requirements of the three conditions of the model. This is the problem that needs to be solved to derive the principal components.

The model is first required to satisfy the following conditions:

① F_i, F_j are uncorrelated with each other $(i \neq j, i, j = 1,2,3.....p)$.

② The variance of F_1 is greater than the variance of F_2 is greater than the variance of F_3.

③ The contributions of $F_1, F_2, \ldots . F_n$ add up to equal 1.

According to the condition ① of the principal component mathematical model requires that the principal components are uncorrelated with each other, and for this reason the covariance array between the principal components should be a diagonal array. That is, for the principal components:

$$F = AX \tag{1}$$

Its covariance array should be:

$$Var(F) = Var(AX) = (AX) \cdot (AX)^{'} = AXX^{'}A^{'} = \Lambda = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_p \end{pmatrix} \tag{2}$$

Let the covariance array of the original data be V , if the original data is normalized then the covariance array is equal to the correlation matrix, i.e., there is:

$$V = R = XX^{'} \tag{3}$$

Then by the principal component mathematical model condition ③ and the nature of orthogonal matrix, if it can satisfy condition ③ it is best to require A to be an orthogonal matrix, that is, to satisfy $AA^{'} = I$, and so the covariance of the original data is obtained by substituting it into the covariance matrix formula of the principal component:

$$Var(F) = AXX^{'}A^{'} = ARA^{'} = \Lambda \tag{4}$$

$$ARA^{'} = \Lambda \quad RA^{'} = A^{'}\Lambda \tag{5}$$

Expanding the above equation yields:

$$\begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{pmatrix} \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{2p} \\ a_{12} & a_{22} & \cdots & a_{p2} \\ \vdots & \vdots & \vdots & \vdots \\ a_{1p} & a_{2p} & \cdots & a_{pp} \end{pmatrix}$$
$$\begin{pmatrix} a_{11} & a_{21} & \cdots & a_{2p} \\ a_{12} & a_{22} & \cdots & a_{p2} \\ \vdots & \vdots & \vdots & \vdots \\ a_{1p} & a_{2p} & \cdots & a_{pp} \end{pmatrix} \cdot \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_p \end{pmatrix} \tag{6}$$

Expanding the left and right sides of the equation, by the property of equality of matrices, the equation derived here from the first column only is:

$$\begin{cases} (r_{11} - \lambda_1)a_{11} + r_{12}a_{12} + \cdots + r_{1p}a_{1p} = 0 \\ r_{21}a_{11} + (r_{22} - \lambda_1)a_{12} + \cdots + r_{2p}a_{1p} = 0 \\ \cdots\cdots\cdots \\ r_{p1}a_{11} + r_{p2}a_{22} + \cdots + (r_{pp} - \lambda_1) = 0 \end{cases} \tag{7}$$

In order to obtain the solution of this chi-square equation, the coefficient matrix determinant is required to be 0, i.e.:

$$\begin{vmatrix} r_{11}-\lambda_1 & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22}-\lambda & \cdots & r_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp}-\lambda_1 \end{vmatrix} = 0 \tag{8}$$

Clearly, λ_1 is the eigenvalue of the correlation coefficient matrix, and $\alpha_1 = (a_{11}, a_{12}, \cdots a_{1p})$ is the corresponding eigenvector. Similar equations can be obtained based on the second column, third column, etc., so that λ_1 is the equation:

$$| R - \lambda I | = 0 \tag{9}$$

of P roots, λ_1 is the characteristic root of the characteristic equation, and a_i is the component of its eigenvector.

The following is another proof that the variances of the principal components are sequentially decreasing. Let the p eigenroots of the correlation coefficient matrix R be $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_p$, and the corresponding eigenvectors be a_j:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} \tag{10}$$

The variance with respect to F_1 is:

$$Var(F_1) = a_1 XX' a_1' = a_1 R a_1' = \lambda_1. \tag{11}$$

Similarly there is: $Var(F_i) = \lambda_i$, i.e., the variances of the principal components are decreasing in order. And the covariance is:

$$\begin{aligned} Cov(a_i X', a_j X) &= a_i' R a_j = a_i' \left(\sum_{\alpha=1}^{2} \lambda_\alpha a_\alpha a_\alpha' \right) a_j \\ &= \sum_{\alpha=1}^{2} \lambda_\alpha (a_i' a_\alpha)(a_\alpha a_j) = 0, i \neq j \end{aligned} \tag{12}$$

In summary, according to the proof the principal component covariance in principal component analysis should be a diagonal matrix whose elements on the diagonal are exactly equal to the eigenvalues of the correlation matrix of the original data, and the elements of the principal component coefficients matrix A are the eigenvectors corresponding to the eigenvalues of the correlation matrix of the original data. The matrix A is an orthogonal matrix.

Thus, the variables $(x_1, x_2, \cdots x_p)$ are transformed to obtain the new composite variables:

$$\begin{cases} F_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1p}x_p \\ F_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2p}x_p \\ \qquad \cdots \\ F_p = a_{p1}x_1 + a_{p2}x_2 + \cdots + a_{pp}x_p \end{cases} \tag{13}$$

The new random variables are uncorrelated with each other and have decreasing variance in order.

(2) Calculation steps of principal component analysis

The sample observation data matrix is:

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pp} \end{pmatrix} \tag{14}$$

Step 1: Normalize the raw data:

$$x_{xj}^{'} = \frac{x_{ij} - \overline{x_j}}{\sqrt{\mathrm{var}(x_j)}} (i = 1, 2, \cdots, n; j = 1, 2, \cdots p) \tag{15}$$

Among them:

$$\overline{x_j} = \frac{1}{n}\sum_{i=1}^{n} x_{ij}, var(x_j) = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - \overline{x_j})^2, (j = 1, 2, \cdots p) \tag{16}$$

Step 2: Calculate the sample correlation coefficient matrix:

$$R = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{pmatrix} \tag{17}$$

For convenience, the correlation coefficient of the normalized data is assumed to be X even after normalization of the original data:

$$r_{ij} = \frac{1}{n-1}\sum_{i=1}^{n} x_{ii} x_{ij} \quad (j = 1, 2, \cdots p) \tag{18}$$

Step 3: Find the eigenvalues $(\lambda_1, \lambda_2, \cdots \lambda_n)$ and the corresponding eigenvectors $a_i = (a_n, a_{i2}, \cdots a_w), (j = 1, 2, \cdots p)$ of the correlation coefficient matrix R using Jacobi's method.

Step 4: Select the significant principal components and write the principal component expressions.

Principal component analysis can get p principal components, however, because the variance of each principal component is decreasing, the amount of information contained is also decreasing, so the actual analysis, generally do not select p principal components, but according to the cumulative contribution rate of each principal component to select the first k principal components, where the contribution rate refers to the proportion of the variance of a particular principal component to the total variance, which is actually the proportion of a particular eigenvalue to the total eigenvalue. The proportion of all eigenvalues combined. That is:

$$\text{Contribution rate} = \frac{\lambda_i}{\sum_{i=1}^{2} \lambda_i} \tag{19}$$

The larger the contribution rate, the stronger the information of the original variable contained in that principal component. The selection of the number of principal components k is mainly based on the cumulative contribution rate of the principal components, i.e., the cumulative contribution rate is generally required to reach 85% or more, so as to ensure that the composite variable can include the vast majority of information of the original variable.

Step 5: Calculate the principal component scores.

According to the standardized raw data, according to each sample, respectively, substituting into the principal component expression, you can get the new data of each sample under each principal component, that is, the principal component score. The specific form can be as follows:

$$\begin{pmatrix} F_{11} & F_{12} & \cdots & F_{1k} \\ F_{21} & F_{22} & \cdots & F_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ F_{n1} & F_{n2} & \cdots & F_{nk} \end{pmatrix} \tag{20}$$

Step 6: Based on the data from the principal component scores, then further statistical analysis can be performed.

2.1.2. Multiple Regression Analysis

The main object of regression analysis is the statistical relationship between the variables of objective things, which is based on a large number of experiments and observations of objective things, and is used to find the statistical regularity hidden in those phenomena that seem to be uncertain. Examining the interdependence of one variable (dependent variable) with the rest of multiple variables (independent variables) is the multiple regression problem [23].

(1) General form of multiple linear regression models

Let the linear regression model of the random variable y with the general variables $x_1, x_2, \cdots x_p$ be:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon \tag{21}$$

Where $\beta_0, \beta_1 \cdots, \beta_p$ is the $p+1$ unknown parameter, β_0 is the regression constant, and $\beta_1 \cdots, \beta_p$ is the regression coefficient. Y is called the explanatory variable (dependent variable), while $x_1, x_2, \cdots x_p$ are p general variables that can be precisely measured and can be controlled, and are called explanatory variables (independent variables).

For a practical problem, if we obtain n sets of observations $(x_n, x_{n2}, \cdots x_{ip}; y)(i = 1, 2, ...n)$, then the linear regression model can be expressed as:

$$y = X\beta + \varepsilon \tag{22}$$

where $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_3 \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}, \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$ is the random error.

(2) Goodness of fit

Goodness of fit is used to test the fit of the regression equation to the sample observations and is reflected in the value of the sample coefficient of determination $R^2 = \frac{SSR}{SST}$.

The sum of squares decomposition formula is:

$$\sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2 \tag{23}$$

The regression sum of squares, $SSR = \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2$, captures the magnitude of fluctuations in the n estimates $\hat{y}_1, \hat{y}_2, ..., \hat{y}_n$ fluctuate in size, which is due to the fact that Y is associated with the independent variables $x_1, x_2, ...x_m$ do have a linear relationship and are related by $x_1, x_2, ...x_m$ variations.

The residual sum of squares, $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$, is the sum of the residuals, in addition to the $x_1, x_2, \ldots x_m$ to Y linearly (including $x_1, x_2, \ldots x_m$ to Y caused by nonlinear relationships and random errors).

The total sum of squared deviations $SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$ embodies the observations $y_1, y_2, \ldots y_n$ total fluctuation size.

From the significance of the regression sum of squares and the residual sum of squares, it can be known that if the greater the proportion of the regression sum of squares in the total sum of squares of deviations, the better the linear regression effect, which indicates that the regression line fits the sample observations better. If the proportion of the sum of squared residuals is large, the regression straight line is not well fitted to the sample observations. The ratio of the sum of squared regressions SSR to the sum of squared total deviations SST is defined as the coefficient of determination $R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$.

The coefficient of determination R^2 is a relative indicator of the goodness of fit of a regression line to a sample of observations, reflecting the proportion of variation in the dependent variable that can be explained by the independent variable. If the coefficient of determination R^2 is close to 1, it means that the vast majority of the uncertainty in the dependent variable can be explained by the regression equation, and the regression equation is a good fit. Conversely, if R^2 is not large, it means that the regression equation is not good and should be modified. However, the problem to be noted is that a large coefficient of determination does not affirm that the relationship between the independent variable and the dependent variable is linear, this is because it is possible that curvilinear regression works better. Especially when the range of values of the independent variable is very narrow, linear regression is usually better, so that linear regression can not be used for extrapolation prediction. At this point we can use the simulated misfit test to determine the true functional relationship between the dependent variable and the independent variable, this test requires repeated observations of the independent variable, and economic data modeling usually can not get repeated observations, so we have to use the residual analysis method to determine the correctness of the regression equation.

2.1.3. Non-Violation of Basic Assumptions

In the basic assumptions of the regression model, we assume that the random error term $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ has the same variance independent or uncorrelated, i.e., for all sample points, there:

$$\begin{cases} E(\varepsilon_i) = 0, i = 1, 2, \ldots, n \\ Cov(\varepsilon_i, \varepsilon_j) = \begin{cases} \sigma^2, i = j \\ 0, i \neq j \quad (i, j, 1, 2, \ldots n) \end{cases} \end{cases} \tag{24}$$

However, when modeling regressions for real-world problems, there are often cases where this assumption is contradicted, one being heteroskedasticity in econometric modeling, viz:

$$var(\varepsilon_i) \neq var(\varepsilon_j), \text{when } i \neq j \tag{25}$$

The other is autocorrelation, i.e:

$$cov(\varepsilon_i, \varepsilon_j) \neq 0, \text{when } i \neq j \tag{26}$$

Therefore, when building the regression model, we should fully consider the impact of these two situations on the regression modeling, and diagnose and deal with the possible heteroskedasticity and autocorrelation. Before using the regression equation for analysis and prediction, we should check whether the model meets the basic assumptions in order to make further modifications to the model.

(1) Heteroskedasticity

Regarding the test of heteroskedasticity, more than 10 diagnostic methods have been proposed, but there is no universally recognized optimal method. In view of the unrepeatable observability of tourism development data, we will use the residual plot analysis method.

Residual plot analysis is an intuitive and convenient method of analysis. It draws scatter plots with the residuals e_i as vertical coordinates and other suitable variables as horizontal coordinates. There are three choices of commonly used horizontal coordinates:

a) Use the fitted value $\hat{y}$ as the horizontal coordinate.

b) Take $x_i(i=1,2,.....p)$ as the horizontal coordinate.

c) Observation time or ordinal number as the horizontal coordinate.

If the regression model is appropriate for the sample data, then the residuals e_i should reflect the properties assumed by e_i , and thus can be used to determine whether the regression model has certain properties. In general, when the regression model satisfies all the assumptions, the N-point walk of the residual plot should be random without any pattern. If there is heteroskedasticity in the regression model, the walks of the points on the residual plot show a corresponding trend, such as the value of the residuals e_i increases (or decreases) with the increase of the y-value with an obvious regularity, and thus it can be assumed that the variance of the model's random error term ε_i is non-chiral and heteroskedasticity exists.

(2) Autocorrelation

When there is serial correlation in the random error terms of a linear regression model, the basic assumptions of the linear regression equation are violated and the following problems will arise:

a) The estimates of the parameters are no longer minimum variance linearly unbiased.

b) The mean square error MSE may grossly underestimate the variance of the error term.

c) It tends to lead to an over evaluation of the t-value and invalidation of the commonly used F-test and t-test.

d) When serial correlation exists, $\hat{\beta}$ remains an unbiased estimator of β , but in any particular sample, $\hat{\beta}$ may seriously distort the true picture of β , i.e., the least squares estimator becomes very sensitive to sampling fluctuations.

e) If the least squares estimation of model parameters is applied without treatment, forecasting and performing structural analysis with this model will bring large variance or even erroneous results.

The DW test is an effective method for testing serial correlation problems. The first-order autoregressive form of the randomly perturbed term is $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$, and in order to test for serial correlation, the constructed hypothesis is $H_0 : \rho = 0$.

To test the above hypothesis, the DW statistic is defined as:

$$DW = \frac{\sum_{i=2}^{n} e_i - e_{i-1}{}^{2}}{\sum_{i=1}^{n} e_i^2} = \frac{\sum_{i=2}^{n} e_i^2 + \sum_{i=2}^{n} e_{i-1}^2 - 2\sum_{i=2}^{n} e_i e_{i-1}}{\sum_{i=1}^{n} e_i^2}, \text{Of which } e_i = y_i - \hat{y}_i \tag{27}$$

Then:

$$DW \sim 2\left[1 - \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2}\right], \hat{\rho} \sim \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2}, \therefore DW = 2[1-\rho] \tag{28}$$

The value of DW ranges from $0 \le DW \le 4$, and the DW distribution is checked based on the sample size n and the number of explanatory variables k (including the constant term) to obtain the critical values d_L and d_v to determine the autocorrelation state of the model as follows:

$0 \le DW \le d_L$ - There is a positive correlation in the error terms $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$.

$d_L < DW \le d_U$ - Cannot determine if there is autocorrelation.

$d_U < DW < 4 - d_U$ - No autocorrelation between the error terms $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$.

$4 - d_U \le DW < 4 - d_L$ - Cannot determine if there is autocorrelation.

$4 - d_L \le DW \le 4$ - The error term $\varepsilon_1, \varepsilon_2, ..., \varepsilon_n$ there is a negative correlation between them.

When the value of DW is around 2, it is safe to assume that the model does not have serial autocorrelation.

2.2. Tourism Development Evaluation Indicators

Based on the principles of comprehensiveness, scientific rationality, targeting and operability, the tourism development evaluation indicators selected in this paper on the basis of reading relevant literature, news, browsing tourism-related websites, etc. are shown in Table 1. A total of 13 indicators are selected to reflect the tourism development.

Table 1. Tourism development evaluation index.

Index	Symbol
The number of A spots	X1
Number of AA spots	X2
AAA number of scenic spots	X3
The number of scenic spots in 4A	X4
The number of scenic spots in 5A	X5
Number of accommodation and catering staff	X6
The number of employees in the third industry	X7
Per capita GDP	X8
The per capita disposable income of urban residents	X9
Urban railway mileage	X10
Urban highway mileage	X11
International airport distance	X12
High train	X13

2.3. Data Sources

By searching China Knowledge Network papers, library databases, the National Tourism Administration and official statistics bureaus of prefecture-level cities and other related materials, we comprehensively analyze domestic and international literature, combing, summarizing, summarizing and calculating them. Understand the current academic research dynamics, track the academic frontier, and then rigorously analyze the dominant factors affecting China's tourism development. Tourism is a comprehensive economic industry, based on tourism resources, providing tourism-related services for tourists, and supported and guaranteed by tourism reception and road transportation. On this theoretical basis this paper selects 13 indicators in Table 1 as the research object. Among them, the number of A-grade, M-grade, AAA-grade, 4A-grade and 5A-grade scenic spots is compiled from the relevant data on the official website of each provincial and municipal tourism bureau. The number of accommodation and catering employees, the number of employees in the tertiary industry, the per capita GDP, and the per capita disposable income of urban residents are compiled from the City Statistical Yearbook and the statistical bulletin of national economic and social development of each city. The mileage of intra-city railroads, intra-city highways and high-speed railroads are calculated by GIS using the 2024 China administrative area railroad map, highway map and high-speed railroad map as the base map. The distance to international airports is obtained by measuring the nearest airports through the distance measurement function in “Earth Online”.

3. Multiple Linear Regression Analysis of Tourism Development

3.1. Analysis of the Main Driving Factors of Tourism Development

The standardized data of the 13 indicator factors were subjected to KMO and Bartlett's test, and the results are shown in Table 2.The test statistic of KMO is 0.896, which is greater than 0.8, indicating that the degree of overlap of information between the variables is relatively high, and it is very suitable for principal component analysis.

Table 2. KMO and Bartlett test indicators Factor.

The sample is sufficient for the sampling of Kaiser-Meyer-Olkin		0.896
Bartlett's Test of Sphericity	Approximate card	381.652
	DF	15
	Sig.	0.000

The factor fragmentation map was plotted as shown in Figure 1. Four driving factors were extracted using maximum variance orthogonal rotation and factor minimum eigenvalues greater than one.

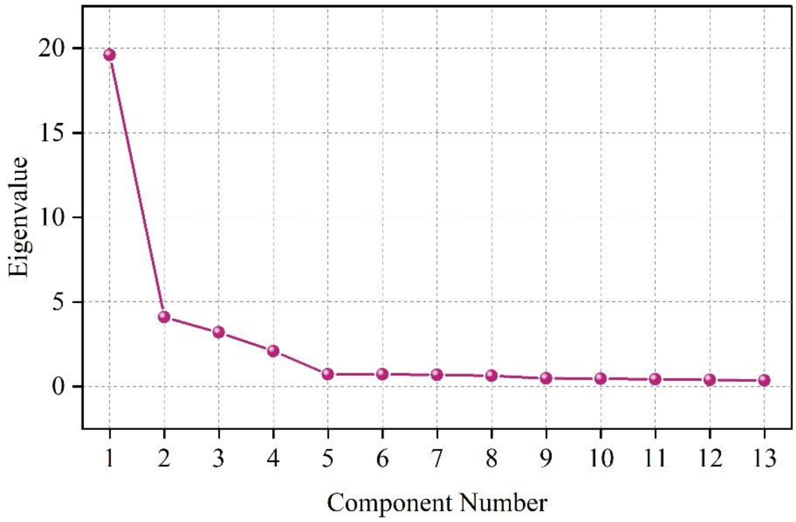


Figure 1. Factor scree plot.

The characteristic roots and cumulative variance contribution rate of the factors affecting tourism development are shown in Table 3, which shows that the cumulative variance contribution rate of the first four extracted characteristic roots is 95.125%, which has adequately summarized most of the data in the indicator system. According to the characteristic roots, variance contribution rates, cumulative contribution rates and the interrelationships among the four extracted factors, it can be categorized into socio-economic conditions, attractiveness of tourism resources, level of infrastructure construction and public services, and capacity of tourism services.

Table 3. The contribution of the influence factor and the cumulative variance.

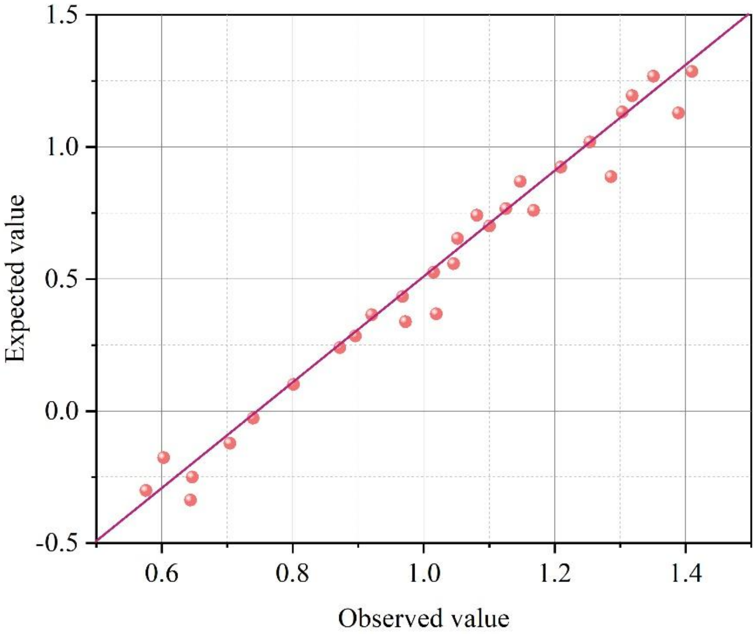
Factor	Initial eigenvalue		
	Characteristic root	Variance contribution	Cumulative contribution
1	18.974	73.023	73.023
2	2.956	11.416	84.439
3	1.681	6.416	90.855
4	1.114	4.27	95.125

The principal component loading matrix is shown in Table 4. Variables with loading factors greater than 0.85 in the first principal component are: per capita GDP, per capita disposable income of urban residents, whose loadings are 0.9 and 0.88, respectively, which mainly reflect the socio-economic development, so this factor is defined as socio-economic conditions. In the second principal component, the variables with loading factors greater than 0.85 are: the number of grade A scenic spots, the number of grade AA scenic spots, the number of grade AAA scenic spots, the number of grade 4A scenic spots, and the number of grade 5A scenic spots, with loadings of 0.913, 0.925, 0.934, 0.945, and 0.886 respectively, mainly reflecting the endowment of tourism resources, and therefore this factor is defined as the attractiveness of tourism resources. In the third principal component, the variables with loading factors greater than 0.85 are: the number of employees in accommodation and catering and the number of employees in the tertiary industry, with loadings of 0.873 and 0.941 respectively, which mainly reflect the reception conditions of the tourism industry, and thus this factor is defined as the level of infrastructure construction and public services. Variables with a loading factor greater than 0.85 in the fourth principal component include: mileage of intra-city railroad, mileage of intra-city highway, distance to international airports and mileage of high-speed railways, with loadings of 0.898, 0.86, 0.853 and 0.86 respectively, which mainly reflect the conditions of transportation and thus this factor is defined as the capacity of tourism services.

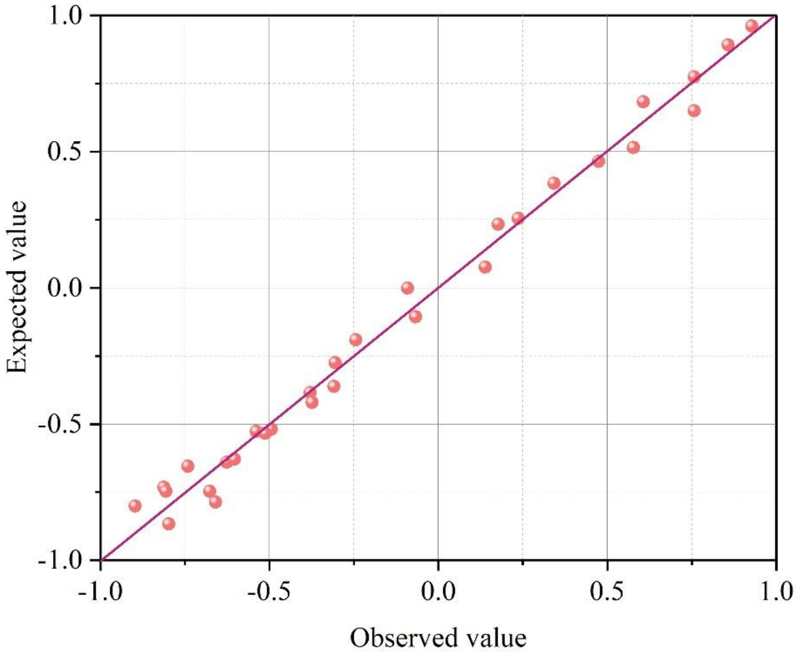
Table 4. Matrices of Factors Loadings.

Var	Principal component			
	1	2	3	4
X1	0.169	**0.913**	0.137	0.262
X2	0.252	**0.925**	0.424	0.331
X3	0.17	**0.934**	0.418	0.266
X4	0.496	**0.945**	0.5	0.305
X5	0.557	**0.886**	0.126	0.473
X6	0.284	0.44	**0.873**	0.167
X7	0.311	0.141	**0.941**	0.259
X8	**0.9**	0.171	0.175	0.28
X9	**0.88**	0.237	0.233	0.33
X10	0.2	0.354	0.367	**0.898**
X11	0.365	0.559	0.189	**0.86**
X12	0.228	0.287	0.31	**0.853**
X13	0.195	0.107	0.182	**0.86**

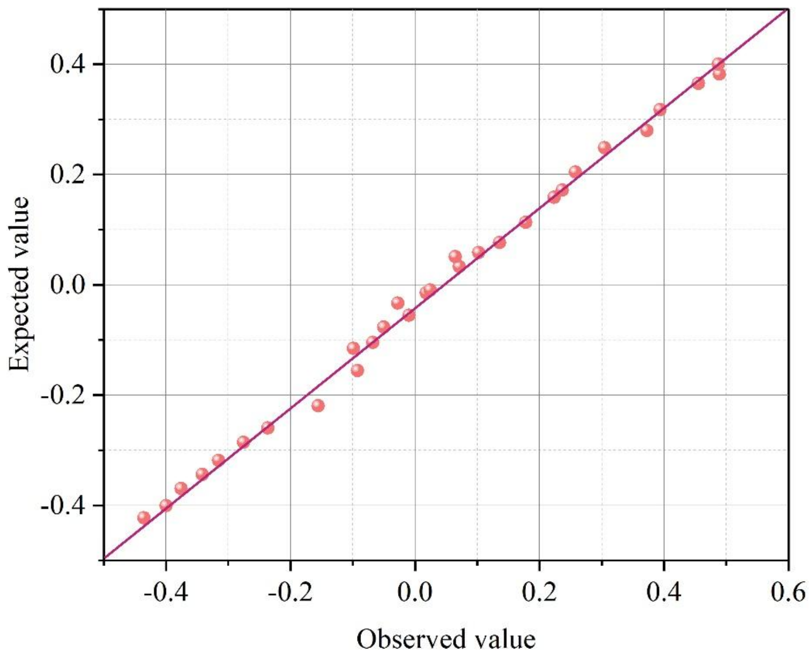
The normal distributions of the 4 principal components are shown in Fig. 2, and (a)~(d) denote the normal distribution plots of socio-economic conditions, attractiveness of tourism resources, level of infrastructure construction and public services, and capacity of tourism services, respectively.The factor eigenroots of the 4 principal components are fully consistent with the normal distribution. Therefore, in the following, the 4 driving factors are taken as explanatory variables, and tourism development is taken as the explanatory variable, and a multiple linear regression model is constructed to investigate the influencing factors of tourism development.



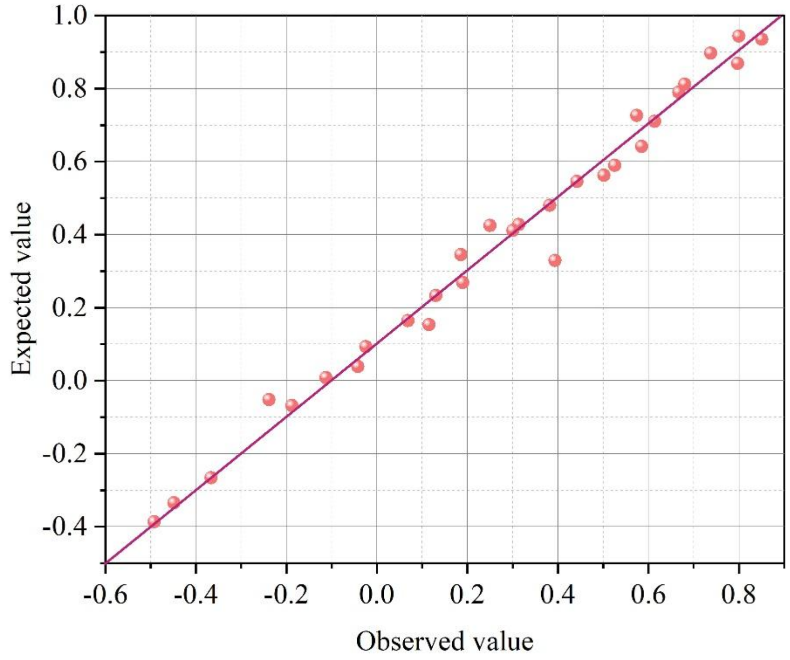
(a) Main component 1 normal distribution



(b) Main component 2 normal distribution



(c) Main component 3 normal distribution



(d) Main component 4 normal distribution

Figure 2. The normal distribution of four main components.

3.2. Multiple Linear Regression Analysis

3.2.1. Correlation Analysis

The explanatory variables were standardized to eliminate the difference in magnitude, and after standardization the variables were noted as ZLX1, ZLX2, ZLX3, ZLX4. In order to understand the correlation between the explanatory variables, correlation analysis was carried out on the explanatory variables using the Pearson correlation test, and the results of the test are shown in Table 5. From the test results, it can be seen that there is a strong correlation between the variables, so the variables are considered for regression analysis using stepwise regression method.

Table 5. Variable correlation analysis results.

Correlation	ZLX1	ZLX2	ZLX3	ZLX4
ZLX1	1	0.962**	0.942**	0.987**
ZLX2	0.962**	1	0.964**	0.988**
ZLX3	0.942**	0.964**	1	0.986**
ZLX4	0.987**	0.988**	0.986**	1

3.2.2. Multiple Regression Modeling

The fitted tourism development indicator F1 was used as an explanatory variable with the standardized explanatory variables for multiple regression analysis, and the variables were subjected to stepwise regression analysis using SPSS 22.0, according to which the multiple regression model was established:

$$Y = \mathrm{a} + aZLX1 + bZLX2 + cZLX3 + dZLX4 + \varepsilon \quad (29)$$

Where a is the intercept term and ε is the perturbation disturbance term, the regression results are shown in Table 6. It can be seen that the R^2 of the regression model is 0.991 and the corrected R^2 is 0.994, and the model fits well. The Durbin Watson test value of the model is 2.003, checking the table of DW critical values at 0.01 significance level, the Durbin Watson test value falls between dU and 4-dU, and there is no autocorrelation of variables.

Table 6. Multivariate regression of the model.

R	R^2	Adj_R^2	Standard bias error	Significant f value changes	Durbin-Watson
0.966^b	0.991	0.994	0.1952	0.000	2.003

The results of the analysis of variances of the model are shown in Table 7. From the results, it can be seen that the model has an F-value of 342.236 and a P-value of 0, which passes the F-test at the 0.05 level of significance, and the model has a high degree of approximation.

Table 7. Analysis of the variation of the model.

	Sum of squares	DF	Mean squared	F	Sig.
Regression	53.621	8	13.263	342.236	0.000^b
Residual error	0.387	12	0.036	-	-
Total	52.635	13	-	-	-

The coefficients of the model are shown in Table 8. The explanatory variables in the regression model passed the t-test at the 5% level, and the coefficients of the variables were able to accurately explain the explanatory variable tourism development F1. It is generally considered that the variance inflation factor VIF is greater than 10 and there is covariance between the variables, and all the values of VIF between the variables are less than 10 and are not affected by multicollinearity.

Table 8. Coefficient of model.

	Nonnormalized coefficient	Standard error	Normalization factor	T	Sig.	Common linear statistics	
	B		Beta			Allowance	VIF
ZLX1	2.083	0.125	1.085	16.695	0.011	0.17	5.667
ZLX2	0.496	0.094	0.261	5.581	0.001	0.325	3.033
ZLX3	0.145	0.049	0.094	2.906	0.015	0.949	1.061
ZLX4	1.154	0.087	0.59	13.208	-0.01	0.375	2.782

3.2.3 Stability Tests

In order to ensure that there is no random trend and to prevent the existence of pseudo-regression in the time series, the unit root test was performed on the variables using Eviews7 and the results of the test are shown in Table 9. From the test results, it can be seen that after the second-order difference test is performed on the explanatory variable F1 and the explanatory variables ZLX1, ZLX2, ZXL3, and ZLX4, the explanatory variables as well as the explanatory variables are smooth sequences, same-order monotonic, and satisfy the conditions of cointegration test at the 5% significance level.

Table 9. ADF test results.

Var.	Test form	ADF	1% threshold	5% threshold	10% threshold	Prob
D(F1,2)	(C,0,2)	-4.37216	-4.20006	-3.17436	-2.72964	0.0058
D(ZLX1,2)	(C,0,2)	-5.54751	-4.20026	-3.17471	-2.73135	0.0016
D(ZLX2,2)	(C,0,2)	-3.35309	-4.12223	-3.14531	-2.71319	0.036
D(ZXL3,2)	(C,0,2)	-3.90403	-4.41946	-3.26037	-2.77178	0.0198
D(ZLX4,2)	(C,0,2)	-6.18012	-4.19886	-3.17598	-2.72828	0.0005

The unit root test is performed on the sequence of residuals e of the multiple linear regression equation and the results are shown in Table 10. According to the test results, it can be seen that the regression equation residual series e is a smooth series, passed the smoothness test, so all get the multiple regression equation as:

$$F1 = -6.18012 + 2.083ZLX1 + 0.496LX2 + 0.145ZLX3 + 1.154ZLX4 \quad (30)$$

Table 10. Model cointegral test results.

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-6.79654	0.0000
Test critical values	-4.0576	-4.0576	-
	-3.12094	-3.12094	
	-2.70064	-2.70064	

Elasticity analysis based on the regression results of the multiple linear regression model shows that socio-economic conditions have the most significant impact on tourism development, with tourism development increasing by 2.083% for every 1% increase in it. The second is tourism service capacity, which increases by 1.154% for every 1% increase in tourism development. While tourism resource attractiveness and infrastructure construction only increase by 0.496% and 0.145% for every 1% increase in tourism development.

4. Strategies for the Integration of Visitor Satisfaction Improvement and Business Operation Paths

4.1. Strengthen Staff Training and Improve Staff Service Awareness

To really do a good job of staff training, travel companies should start from the following aspects: First, innovative training methods to improve training efficiency. Most of the ground service staff of Tourism Company Limited are college and above education, also younger, can be designed by using various new media, with actual cases of the plot, so that employees can participate in the teaching, which can improve the enthusiasm of employees to participate. Secondly, change the training and assessment methods to motivate employees to work. To truly improve the enthusiasm of employees, so that employees have a good sense of service, should change the way of assessment, after the completion of training, can set up a variety of incentives to staff competitions, for the performance of employees can be given appropriate material rewards [24].

4.2. Improvement of Employee Benefits

In view of the low wages and low satisfaction of employees in tourism companies, tourism companies should give appropriate incentives in addition to the most basic wages of employees, especially during the summer and spring transportation period, when the number of travelers increases dramatically and the work pressure of employees increases, so they can give bonuses to employees on duty and so on. At the end of the peak period of summer and spring transportation, the staff's work motivation can be stimulated by giving paid vacations, free air tickets, and opportunities for training and further education to the staff with excellent performance.

4.3. Good Customer Relations

Proper handling of traveler relations will not only stabilize the source of customers, but also enhance the travelers' sense of identification with the travel company. Tourism companies can be targeted to establish a passenger history file, especially for important travelers, record their dietary preferences, seat selection habits, personal taboos, etc., which can help personnel to understand the needs of travelers in advance, and then provide personalized service.

5. Conclusion

The study used principal component analysis to extract four main factors affecting tourism development: socio-economic conditions, attractiveness of tourism resources, level of infrastructure development and public services, and tourism service capacity. After that, a quantitative model of tourism development was established using multiple linear regression analysis as follows: $F1 = -6.18012 + 2.083ZLX1 + 0.496LX2 + 0.145ZLX3 + 1.154ZLX4$. The model indicates that for every 1% increase in socio-economic conditions, attractiveness of tourism resources, level of infrastructure construction and public services, and tourism service capacity, tourism development increases by 2.083%, 0.496%, 0.145%, and 1.154%, respectively. In order to further promote tourism development, the study proposes to strengthen staff training, improve staff welfare and good customer relations.

References

1. Yu, G., Qin, J., Cheng, Z., Li, Z., & Quan, S. (2024). The Transformation of Intelligent Tourism Planning through the" Zibo Barbecue" Craze. Journal of Big Data and Computing (ISSN: 2959-0590), 2(1), 51.
2. Han, K., Song, X., Peng, J., & Liu, Y. (2025). Research on Rural Grassroots Governance from the Perspective of Village BA. Journal of Current Social Issues Studies, 2(3), 193-198.
3. Bec, A., Moyle, B., Timms, K., Schaffer, V., Skavronskaya, L., & Little, C. (2019). Management of immersive heritage tourism experiences: A conceptual model. Tourism Management, 72, 117-120.
4. Cheer, J. M., Belhassen, Y., & Kujawa, J. (2017). The search for spirituality in tourism: Toward a conceptual framework for spiritual tourism. Tourism Management Perspectives, 24, 252-256.
5. Huete-Alcocer, N., & Hernandez-Rojas, R. D. (2022). Does local cuisine influence the image of a World Heritage destination and subsequent loyalty to that destination?. International Journal of Gastronomy and Food Science, 27, 100470.
6. Bazzanella, F., Schnitzer, M., Peters, M., & Bichler, B. F. (2023). The role of sports events in developing tourism destinations: a systematized review and future research agenda. Journal of Sport & Tourism, 27(2), 77-109.
7. Ying, L., Ziyi, Q., Shizhuan, H., Yan, L., & Tongqian, Z. (2024). The development of digital tourism in China. Cogent Social Sciences, 10(1), 2347013.
8. Castro, J. C., Quisimalin, M., de Pablos, C., Gancino, V., & Jerez, J. (2017). Tourism marketing: Measuring tourist satisfaction. Journal of Service Science and Management, 10(3), 280-308.
9. Correia, A., Kozak, M., & Ferradeira, J. (2013). From tourist motivations to tourist satisfaction. International journal of culture, tourism and hospitality research, 7(4), 411-424.
10. Shahrivar, R. B. (2012). Factors that influence tourist satisfaction. Journal of Travel and Tourism Research (Online), 12(1), 61.
11. Saayman, M., Li, G., Uysal, M., & Song, H. (2018). Tourist satisfaction and subjective well-being: An index approach. International Journal of Tourism Research, 20(3), 388-399.
12. Zeng, L., & Yi Man Li, R. (2021). Tourist satisfaction, willingness to revisit and recommend, and mountain kangyang tourism spots sustainability: A structural equation modelling approach. Sustainability, 13(19), 10620.
13. Pizam, A., Uriely, N., & Reichel, A. (2000). The intensity of tourist–host social relationship and its effects on satisfaction and change of attitudes: The case of working tourists in Israel. Tourism management, 21(4), 395-406.
14. Nguyen, Q. N., Nguyen, H. L., & Le, T. D. H. (2020). Relationships among novelty seeking, satisfaction, return intention, and willingness to recommend of foreign tourists in Vietnam. Management Science Letters,

10(10), 2249-2258.
15. Pearce, P. L., & Moscardo, G. (1985). Visitor evaluation: An appraisal of goals and techniques. Evaluation Review, 9(3), 281-306.
16. Abdou, A. H., Mohamed, S. A. K., Khalil, A. A. F., Albakhit, A. I., & Alarjani, A. J. N. (2022). Modeling the relationship between perceived service quality, tourist satisfaction, and tourists' behavioral intentions amid COVID-19 pandemic: Evidence of yoga tourists' perspectives. Frontiers in Psychology, 13, 1003650.
17. Syakier, W. A., & Hanafiah, M. H. (2022). Tour guide performances, tourist satisfaction and behavioural intentions: a study on tours in Kuala Lumpur city centre. Journal of Quality Assurance in Hospitality & Tourism, 23(3), 597-614.
18. Nguyen, T. N., & Huynh, V. D. (2024). Exploring the Impact of Tourism Environment on Tourist Satisfaction in Tourist Sites: An Example of Phong Dien Tourism Village, Vietnam. Case Studies in the Environment, 8(1), 2281531.
19. Suwannakul, E., Khetjenkarn, S., & Wichaidit, S. (2024). How the service quality of regional airport affects tourist satisfaction with airport services and their intention to revisit secondary cities: An exploratory study. Journal of Quality Assurance in Hospitality & Tourism, 25(2), 371-396.
20. Haron, H. I. C., Abdullah, H., Ahmad Tajuddin, S. A. F. S., & Rahim, N. A. A. A. (2023). Bootstrapping on Tourist Satisfactions: Examine the Mediating Effects Between Determinant Factors and Tourist Intention to Revisit Terengganu's Edutourism Destinations. International Journal of Sustainable Development & Planning, 18(6).
21. Bogale, D., & Wondirad, A. (2019). Determinant factors of tourist satisfaction in Arbaminch City and its vicinity, Southern Ethiopia. International Journal of Tourism Policy, 9(4), 320-348.
22. Monica Palma ,Sandra De Iaco ,Claudia Cappello& Veronica Distefano. (2023). Tourism composite spatial indicators through variography and geographically weighted principal components analysis. Annals of Operations Research,342(3),1-19.
23. Xin Jin,Youzhe Wang,Jian Luo & Qunli Sun. (2022). Research on the Development of Tourism Industry in the Post-Epidemic Period——Analysis Based on Multiple Regression Model GM(2,1) Model. Academic Journal of Business & Management,4.0(8.0),
24. Suh,Chang Juck,Kim & Jong Hoon. (2013). The Effects of the Service Recovery Procedure on Process Improvement, Employee Attitude, Capability Improvement and Customer satisfaction. Journal of Korea Service Management Society,14(2),25-48.