Analysis and Practice of the Synergistic Application of Four Methods and Two Software in Tourism Market Research

Zhengxin Chang
Tourism Management Major, Business, School, Tianjin University of Finance and Economics, Tianjin, China

Abstract: Moderating effect, mediating effect, one-way analysis of variance (ANOVA), and two-way/multi-way ANOVA are important analytical methods in the field of tourism market and tourism research. Among them, the moderating effect can reveal the differences in variable relationships under different conditions; the mediating effect can analyze the internal mechanism of action between variables; one-way ANOVA can effectively compare the significant differences in tourism preferences and behaviors among different groups (e.g., age, region); and two-way/multi-way ANOVA further explores the comprehensive impact of the interaction of multiple factors on tourism phenomena. This paper systematically integrates textbooks, online courses, and academic literature, and uses SPSS and R language software to deeply explain the theory, calculation process, and practical application of the above four methods, laying a foundation for subsequent tourism research. Due to the large theoretical system of the four methods and the limitations of the author's research time, energy, and funds, this paper has not yet deeply explored the linear regression foundation required for moderating effect and mediating effect, the application of R language mediating effect visualization tools (e.g., mediationPlot package), and the processing strategy when data is not significant in two-way/multi-way ANOVA, which will be further studied in the future.

Keywords: Tourism Market Research; Market Analysis; Mediating Effect; Moderating Effect; Analysis of Variance (ANOVA)

1. Introduction

In tourism market research, four quantitative techniques-moderation, mediation, one-way analysis of variance (ANOVA), and multi-factor ANOVA-form a practical toolkit for explaining heterogeneity in tourist behavior, disentangling underlying mechanisms, and testing group differences across typical industry scenarios. SPSS and R are the primary software environments supporting this toolkit, together covering data preprocessing, model specification, estimation, diagnostics, and reporting across a full analysis pipeline. The present study consolidates these "four methods and two software" into a coherent, reusable workflow tailored to tourism problems such as factors shaping homestay sales and cross-industry comparisons of service complaints, while also acknowledging current pain points and limits that practitioners confront in day-to-day analysis[1].

Methodologically, the paper clarifies the logic linking total, direct, indirect, and interaction effects, and it emphasizes decisions that materially affect interpretation in practice. For moderation, mean-centering is highlighted as a default step to make parameters interpretable around realistic "typical" values rather than arbitrary zero points; this guards against misinterpretation when the raw scale lacks a meaningful zero. For variance analysis, we position one-way and multi-factor ANOVA as complementary tools: the former addresses mean differences across three or more groups, whereas the latter extends to multiple factors and their interactions-an expansion that is frequently relevant in tourism contexts where managerial interventions and contextual attributes co-occur. For mediation, we connect conceptual diagrams to operational tests, outlining classical approaches (e.g., coefficient-difference and coefficient-product) and their large-sample inference, as groundwork for later, software-specific implementation[2].

The study's integrative stance is motivated by the dispersion of teaching materials and applied papers that often present a single method in isolation. By threading concepts, models, implementation, and reporting into one workflow, we aim to raise the consistency and reproducibility of empirical work in tourism. Concretely, the paper articulates how the unified workflow translates into day-to-day choices-constructing and centering interaction terms, diagnosing multicollinearity, reporting effect sizes, and adopting robust language when main or interaction effects are non-significant-so that empirical claims remain interpretable and managerially useful across software environments[4].

Finally, we are transparent about scope and limitations that merit further development in future work. Three areas are flagged in particular: strengthening the exposition of linear-regression foundations for moderation and mediation; broadening visualization options for mediation in R; and systematizing strategies for interpreting non-significant findings in multi-factor ANOVA. These issues are common in tourism applications and, by identifying them up front, the paper provides a realistic bridge between methodological ideals and the constraints of applied research practice. Live-streaming commerce provides a natural testbed for the methods reviewed in this paper. In a 2×2 design, human vs. virtual streamer type interacts with product type (hedonic vs. utilitarian) to shape purchase intention; specifically, human streamers outperform virtual streamers for hedonic products, while the two do not differ for utilitarian products (two-way ANOVA with significant interaction; simple effects reported) (Yan et al., 2024)[7]. Moreover, mental imagery quality mediates the effect for hedonic products but not for utilitarian products, illustrating a conditional (moderated) mediation pattern well-suited to the mediation tools we summarize below (Yan et al., 2024). These findings align with our emphasis on (a) interaction terms/ANOVA for moderation and (b) bootstrap or PROCESS-style tests for mediation in applied tourism research[9].

2. Basic Principles of Moderating Effect

2.1 Definition

Moderation stands as a pivotal methodological notion in social science inquiries, serving as a crucial approach for scholars to probe into the interconnections among multiple variables. If the association (the extent and orientation of the regression slope) between the dependent variable Y and the independent variable X alters as a third variable Z varies, then Z is deemed to exert a moderating function between Y and X, and Z is referred to as the moderator variable (see Figure 1: Schematic Diagram of Moderating Effect).

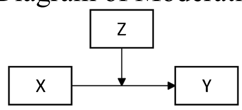

Figure 1. Schematic Diagram of Moderating Effect

2.2 Product-Term Method

To verify the moderating effect, the Product-Term Method is commonly used ① This method creates a new variable as the product of two other variable as the product of two other variable : $X_3=X_2\times X_1$ The product variable X_3 together with its component terms X_1 and X_2 is incorporated into the regression model to simulate the interaction among the two original variables. The regression model is as follow:

$$E[Y_i]=\beta_0+\beta_1X_{1i}+\beta_2X_{2i}+\beta_3X_{1i}X_{2i} \quad (1)$$

X_1X_2 $X_2\times X_1$ X_2 (see Figure 2: Principle of the Product-Term Method; assumingas the moderating variable, as the predictor variable, as the product term, can also be construed as the moderator variable.)

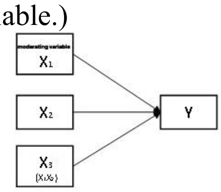

Figure 2. The Principle of the Product-term Method

For comparison, the additive model (without product term, exploring the main effect of linear regression) is:

$$E[Y_i]=\beta_0+\beta_1X_{1i}+\beta_2X_{2i} \quad (2)$$

In Equation (1.2): β_1: Change in the mean value of Y when X_1 increases by 1 unit (with X_2 unchanged).

β_2: Change in the mean value of Y when X_2 increases by 1 unit (with X_1 unchanged).

$$E[Y_i]=\beta_0+\beta_1X_{1i}+\beta_2X_{2i}+\beta_3X_{1i}X_{2i} \quad (3)$$

In Equation (1.1) (non-additive model): β_1 : Impact of X_1 on Y when (X_2=0). β_2: Impact of (X_2) on Y when (X_1=0).(β_3): Change in the slope coefficient Y with respect to X_2 when X_1 (moderating variable) increases by 1 unit (mathematically identical if X_2 is the moderating variable).To simplify understanding,

Original term:

$$E[Y_i]=\beta_0+\beta_1X_{1i}+\beta_2X_{2i}+\beta_3X_{1i}X_{2i} \quad (4)$$

assume X_1=1 (original value) and combine like terms:

$$E[Y_i]=\beta_0+\beta_1 1+\beta_2 X_{2i}+\beta_3 1X_{2i} \\ =(\beta_0+\beta_1)+(\beta_2+\beta_3)X_{2i} \quad (5)$$

When X_1=2 (increases by 1 unit):

$$E[Y_i]=\beta_0+\beta_1 2+\beta_2 X_{2i}+\beta_3 2X_{2i} \\ =(\beta_0+2\beta_1)+(\beta_2+2\beta_3)X_{2i} \quad (6)$$

The intercept change reflects the "shift of the conditional benchmark": β_1 is the conditional effect when X_2=0 , and the intercept (under X_2=0) changes with X_1 (determined by β_1).

2.3 Centering Issue

Data centering refers to subtracting the mean of each score from the score itself to obtain centered scores. Generally speaking, when studying moderating effects, the predictor variables that need to be input comprise the interaction term (XZ) and each of the variables (X, Z). Nevertheless, as the interaction term is derived directly from the product of variables, the covariance among the predictor variables will grow remarkably large. Robinson and Schumacker that data centering is a vital step to guarantee the accurate interpretation of interaction effects. When all variables are quantitative variables, data centering is an essential process. The authors proposed that uncentered data increases multicollinearity, which will cause the contribution of some variables to the regression model to decrease or even disappear.

In other words, because predictor variables overlap and interweave with one another, it's hard for us to judge how much influence predictor variables (X, Z, and XZ) have on the dependent variable (Y). Robinson and Schumacker used the variance inflation factor (VIF) index to intuitively show us the difference in multicollinearity between centered and uncentered data.

Simply put, the knowledge points about centering can be summarized into three items:

Centering purpose 1: Reduce multicollinearity.

When not centered, original variables X and Z are highly correlated with the interaction term XZ (for example, when X is very large, XZ will also be very large), which leads to an increase in the variance inflation factor (VIF) and makes model parameter estimation unstable.

Centering purpose 2: Enhance the rationality of parameter interpretation.

In an uncentered model, the absolute zero point of a variable is used as the reference. However, if the zero point has no practical meaning, the interpretation of coefficients will fail. In a centered model, the mean point of the variable is used as the reference. The mean is the center of the data and usually corresponds to realistic "typical" or "average" scenarios, making the interpretation of coefficients return to reality, such as the age mentioned in the following text.

The author used R language to randomly generate data, perform centering calculation on the data, and display it visually. As shown in Figure 3 below, it can be seen that the mean point has changed, from 50.9 to 0.

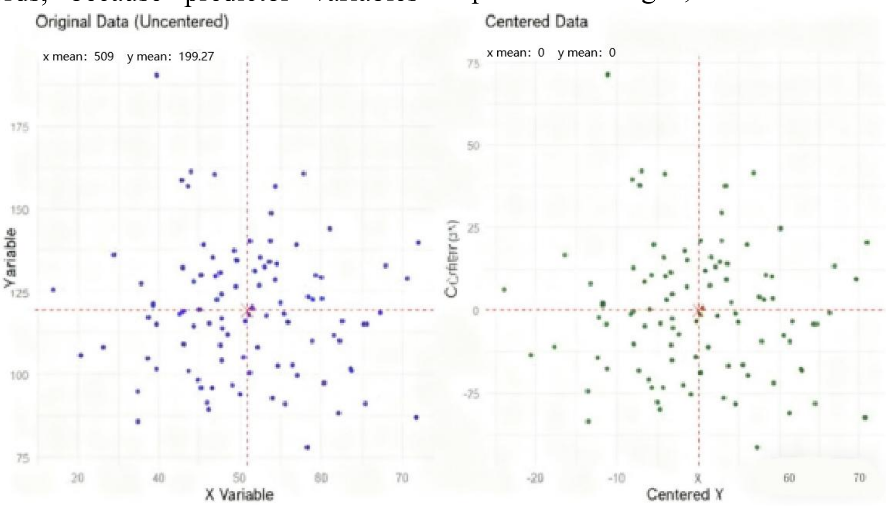

Figure 3. Distribution change before and after data centralization

2.4 Application in Tourism Research

In recent years, the online homestay short-term rental industry has shown a vigorous development trend. When booking a homestay, potential guests often need to perceive and judge the host, based on the perception of the host's emotions, attitudes and behaviors, form an initial impression and preference for the host, and then generate decision-making behavior [3] Chen Dongzhi others explored the influence mechanism of variables such as homestay ratings on whether there is warmth-competence value in the host's self-reported content on homestay sales. This paper chooses to explain the influence of homestay ratings on the host's self-report with warmth but without competence on homestay sales:

Case1: Streamer Type × Product Type (Two-Way ANOVA)

A 2 (streamer: human vs. virtual) × 2 (product: hedonic vs. utilitarian) experiment shows a significant interaction on purchase intention. For hedonic products, human streamers lead to higher purchase intention than virtual streamers (e.g., $M_{human} = 5.78$ vs. $M_{virtual} = 5.07$; $F(1, 101) = 13.69$, $p < .001$). For utilitarian products, the human–virtual difference is nonsignificant (e.g., $F(1, 112) = 0.08$, $p > .05$). This is a textbook example of a cross-over/attenuated interaction: include the product-type dummy, streamer-type dummy, and their interaction in OLS; or run a two-way ANOVA and report the interaction F and simple effects. (Yan et al., 2024).

Case 2: Narrative Style × Product Type (Two-Way ANOVA)

Using a 3 (narrative style: Professional Introduction vs. Entertainment Interaction vs. Scene Storytelling) × 2 (product type: Search vs. Experience) design, results show a significant interaction on purchase intention. For search products, Professional Introduction (PI) and Scene Storytelling (SS) outperform Entertainment Interaction (EI); for experience products, PI and EI outperform SS. Means and ANOVA tables indicate the cross-over/attenuated pattern, and flow experience operates as a mechanism linking styles to intention (ANOVA and PROCESS outputs reported). This gives a clean, teaching-ready moderation example for two-way ANOVA and simple effects. (Zhang et al., 2025).

Table 1. Moderation Effect Table

Study	Independent Variable	Intercept	Host's Self-reported Content	Host's Self-reported Content × Homestay Rating
Main Effect	Warmth Lack of Competence	0.528***	0.021***	×
Moderation Effect		0.527***	0.021***	0.013***

Note. ** indicates p<0.05;*** indicates p <0.001

Data Source: Literature compilation

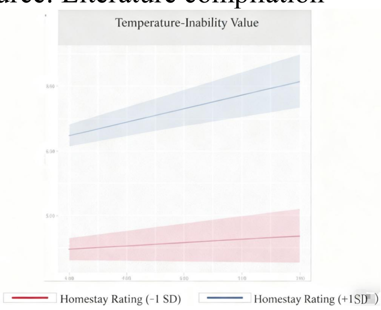

Figure 4. Moderation Effect Diagram of Homestay Ratings

Note. Literature compilation

It can be seen from Table 1 above that the "warmth-lack of competence" value in hosts' self-reports has a significant positive correlation with sales. For homestays with high ratings, the positive impact of the "warmth-lack of competence" value in self-reports on sales is significantly enhanced, and the visual display is shown in Figure 4.

3. Basic Principles of Mediation Effectiveness

3.1 Definition

When examining the impact of an independent variable X on a dependent variable Y, if X impacts Y by acting on a variable M, then M is referred to as a mediator (or mediating variable). The influence of X on Y via the mediating variable M constitutes the mediation effect. Hence, in the causal chain from the independent variable X to the dependent variable Y, the mediating variable occupies an intermediate role. It can also be said that the mediating variable transmits the effect of the independent variable on the dependent variable. Moreover, the mediation relationship also implies a temporal sequence among variables: the occurrence of X precedes that of M, and the occurrence of M precedes that of Y.

The theoretical derivation is as follows. First, understand the total effect of X on Y, and the relationship diagram is as follows in Figure 5:

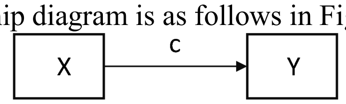

Figure 5. Additive Model

Suppose the coefficient of X's effect on Y at this time is c, and this is the total effect. The formula is as follows:

$$Y=i_1+cX+\varepsilon_1 \tag{7}$$

When the mediation of M occurs, the relationship shown in the following figure 6 arises:

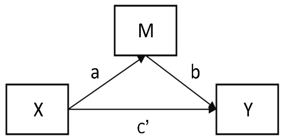

Figure 6. Non-additive Model

It is easy to see that three equations appear at this time: the effect of X on Y, with a coefficient of c'; the coefficient of X's effect on M is a; and the effects of X and M on Y, with coefficients c' and b respectively.

$$Y=i_1+cX+\varepsilon_1 \quad (8)$$

$$M=i_2+aX+\varepsilon_2 \quad (9)$$

$$Y=i_3+c'X+bM+\varepsilon_3 \quad (10)$$

The theoretical formula of mediation effect is the result derived above.

3.2 Complete Mediation and Partial Mediation

c is the overall effect of X on Y, ab is the mediating effect via the mediator variable M, and c′ is the direct effect. When there is just one mediator variable, overall effect = direct effect + mediating effect (c = c′ + ab).

As for partial mediation effect, it means that both the direct effect and the mediation effect exist-that is, X partially influences Y directly (c′ should be statistically significant), and partially influences Y via the mediator variable M (ab should be statistically significant). The complete mediation effect means that X cannot affect Y directly and must be transmitted through M; at this time, the coefficient$c'$is 0 , c=ab In layman's terms: if Zhang San can either go through Li Si or find Wang Wu by himself, it is partial mediation; if Zhang San can only find Wang Wu through Li Si, it is complete mediation.

3.3 Stepwise Causation Method

The stepwise causation method is the most basic method for testing mediation effects. When sorting out and learning about this method, it has multiple names, such as stepwise regression method, stepwise testing method, and traditional mediation analysis method. Since this method was proposed by scholars Baron and Kenny, it is also called the Baron and Kenny method, and all these names have the same meaning. The testing steps of this method are as follows:

$$Y=i_1+cX+\varepsilon_1 \quad (11)$$

$$M=i_2+aX+\varepsilon_2 \quad (12)$$

$$Y=i_3+c'X+bM+\varepsilon_3 \quad (13)$$

For the above equations, the conditions for the establishment of the mediation effect are: (1) c in equation (1) is significant; (2) a in equation (2) is significant; (3) b in equation (3) is significant; (4) $|c'|<|c|$.

3.4 Coefficient Difference Method

As can be seen from the above analysis, $c= c'+ab$, and the total effect equals the direct effect plus the indirect effect. The coefficient difference method is used to test whether there is a difference between the total effect c and the direct effect c' If a mediating variable exists,$c'$is only a part of c ,so c cannot be equal to c', if no mediating variable exists, then$c=c'$. Therefore, we judge whether a mediation effect exists by testing the difference between c andc',The null hypothesis is: c−c'=0,(i.e., the mediating variable does not exist and the mediation effect does not exist). After obtaining a sample, we get the OLS (Ordinary Least Squares) estimates and standard errors of c and c'. Then, we construct a t-statistic (which follows an asymptotic normal distribution under a large sample) to check whether $c-c'$is significantly not equal to 0. The judgment criterion is $|t|>1.96$ $|t|>1.96$(for a two-tailed test with $\alpha=0.05$),If this condition is met, the null hypothesis ($H0: c-c'=0$), is rejected, and the mediation effect is considered significant.

3.5 Coefficient Product Method

If an indirect effect exists, then $a\neq0$, $b\neq0$ If $a=0$ or $b=0$, that is $ab=0$, the mediation effect does not exist. Null hypothesis: $ab=0$, i.e., the mediation effect does not exist. After obtaining the sample, we obtain the estimated values of a and b, as well as the standard errors. Then, we construct a t-statistic (which follows an asymptotic standard normal distribution under a large sample) to check whether ab is significantly not equal to 0. The calculation method and process are the same as those of the coefficient difference method.

Upon careful deduction, it is not difficult to find that the above two methods both derive from the formula c=c′+ab, The coefficient difference method indirectly determines whether a mediation effect exists by testing $c-c'=0$ the coefficient product method directly determines whether a mediation effect exists by testing $ab=0$ and the two methods have the same meaning.

3.6 Application in Tourism Research

Mediation Example (Mental Imagery Quality): In the same stream-shopping context, mental imagery quality mediates the effect of human (vs. virtual) streamer on purchase intention for hedonic products, but not for utilitarian products. Reported conditional indirect effects for hedonic products are significant (e.g., indirect effect ≈ 0.43–0.56, 95% CI excludes 0), whereas utilitarian products show nonsignificant indirect paths. This pattern illustrates how a theoretically meaningful internal state can transmit the effect of a design factor onto behavioral intention-exactly the mediation logic we implement with bootstrap/PROCESS or lavaan. (Yan et al., 2024).

Mediation Example (Flow Experience): In live-stream shopping, narrative styles (professional introduction vs. entertainment interaction vs. scene storytelling) shape flow experience, which in turn increases purchase intention. PROCESS analyzes report significant indirect effects via flow when comparing PI to EI/SS, complementing the two-way ANOVA evidence above. This is a compact illustration of a psychological mediator that translates message design into intention. (Zhang et al., 2025).

Evidence note: see the research model in Figure 1 and *ANOVA/PROCESS results on Tables 6 and 7; Figure 1 interaction plot.

Chen Dongzhi et al. verified that homestay hosts' affinity-seeking strategies indirectly affect guests' destination cooperation intentions through the mediating role of emotional value. Although Qiu Hanqin[6] et al. used the Bootstrap method to verify the mediation effect, the interpretation method is still the same as the above content. In Table 2 below, "Point estimate" refers to the effect size, "T statistics" and "P value" indicate significance. If the absolute value of the t-statistic is larger and the p-value is smaller (for example, all p-values in the table are 0), there is more reason to reject the null hypothesis and believe that the effect truly exists. Then, the mediation effect in the original text is significant, and the mediation effect explanation rate is 13.19% (0.229 × 0.311 / 0.54).

Table 2. Bootstrap Mediation Effect

Effects	Point estimate	T statistics	Bootstrapping				P value
			Percentile 95% CI		Bias-corrected percentile 95% CI		
			Lower	Upper	Lower	Upper	
Standardized total effect	0.54	13.24	0.482	0.638	0.48	Upper 0.640	0
Standardized indirect effect	0.229	4.148	0.103	0.32	0.103	0.319	0
Standardized direct effect	0.311	4.172	0.205	0.497	0.206	0.498	0

Note. Literature compilation

4. One-Way Analysis of Variance

4. 1 Analysis of Variance

Analysis of variance ascertains whether categorical independent factors exert a significant influence on numerical dependent factors by examining whether the means across multiple populations are identical. The testing method involves analyzing data errors. The case selected in this paper is as follows in Table 3:

Table 3. Numerical Display

Observation Value	industry			
	Retail Industry	Tourism	Airlines	Home Appliance Manufacturing
1	57	68	31	44
2	66	39	49	51
3	49	29	21	65
4	40	45	34	77
5	34	56	40	58
6	53	51		
7	44			

Note. From the Internet

Disputes often arise between consumers and producers, sellers, or service providers. At this time, consumers will complain to consumer associations.

In order to evaluate the service quality of several industries, consumer associations selected different enterprises as samples in the retail, tourism, airlines, and home appliance manufacturing industries respectively. Among them, 7 enterprises were selected in the retail industry, 6 in the tourism industry, 5 in the airline industry, and 5 in the home appliance manufacturing industry.

As shown in Figure 7 the visualization below, it is not difficult to find that there are obvious differences in the number of complaints among different industries. The home appliance manufacturing industry has a relatively high number of complaints, while airlines have a relatively low number of complaints. Within the same industry, the number of complaints against different enterprises also varies significantly.

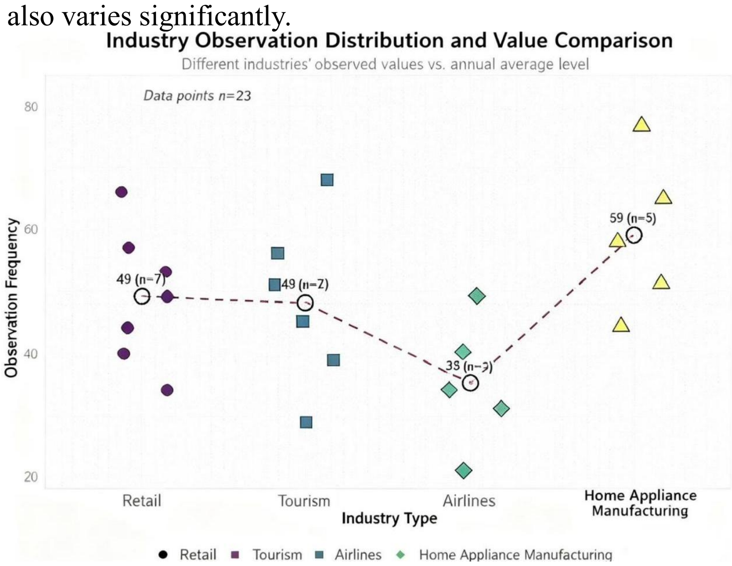

Figure 7. Descriptive Statistics

A scatter plot cannot provide sufficient evidence to prove that there are significant differences in the number of complaints among different industries. At this time, analysis of variance is introduced to explain this difference.

4.2 Basic Ideas and Principles of Analysis of Variance

The core idea of analysis of variance: We are interested in means, but we need to rely on variance to judge whether there are differences between means.

Within-group error: The difference between each observation value of the sample under the same level (population) of a factor. For example, the difference in the number of complaints among different enterprises in the same industry. This kind of error is a random error.

Group-between error: The data error between different levels (such as different industries) is called group-between error. For example, the difference in the number of complaints received by enterprises in different industries. This kind of error may be a random error caused by sampling itself, or a systematic error caused by the systematic factors of the industry itself. Therefore, the group-between error is the sum of random error and systematic error.

4.3 Steps of One-way Analysis of Variance

Propose hypotheses

H_0: $\mu_1=\mu_2=\ldots=\mu_k$,the independent factor exerts no significant influence on the dependent factor.

H_1:$\mu_1,\mu_2,\mu_k$are not all identical, the independent factor exerts a significant influence on the dependent factor.

Construct test statistics:

Compute the mean value of each level (sample), as shown in Table 4.

Calculate the mean value of all observed values. There are 23 observed values, and 1 mean value.

Table 4. Mean Values of Observed Values

Level	Retail Industry	Tourism	Aviation Industry	Home Appliance Manufacturing Industry
Sample size	7	6	5	5
Sample mean	49	48	35	59
Total mean	47.87			

Calculate the sum of squares and mean square for each error. The total sum of squares (SST, Total Sum of Squares) is, in simple terms, the sum of the squared deviations of all values in the sample from the grand mean (the mean derived by averaging each group's mean, and then averaging those means). The formula is:

$$\mathrm{SST}=\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(x_{ij}-\bar{\bar{x}}\right)^2 \quad (14)$$

The sum of squares between groups (SSA, Sum of Squares Between Groups) denotes the extent of divergence between each group's mean and the grand mean, illustrating the variation across different treatment groups (i.e., discrepancies arising from different factor levels). Simply put, it is the sum of the squares of the differences between each group's mean and the grand mean. The formula is

$$\mathrm{SSA}=\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\bar{x}_i-\bar{\bar{x}}\right)^2=\sum_{i=1}^{k}n_i\left(\bar{x}_i-\bar{\bar{x}}\right)^2 \quad (15)$$

the within-group sum of squares (SSE, Sum of Squares Within Groups), also termed the error sum of squares, indicates the extent of divergence between the observed values in each group and the group mean, and represents the variation stemming from random errors (e.g., individual differences, measurement errors, etc.). The calculation method is to take the deviation between the observed value in the group and the group mean, square it, and then sum these squared deviations.

$$SSE=\sum_{i=1}^{k}\sum_{j=1}^{m_i}\left(x_{ij}-\overline{x}_i\right)^2 \quad (16)$$

the mean square between groups (MSA, Mean Square Between Groups) denotes the average of the sum of squares between groups and is utilized to assess the average magnitude of discrepancies between groups. Among them, $df_A=k-1$ represents the degrees of freedom between groups, with k being the number of groups.

$$MSA=\frac{SSA}{k-1} \quad (17)$$

The within-group mean square (MSE, Mean Square Within Groups) denotes the average of the within-group sum of squares and is utilized to assess the average magnitude of random errors. Among them, $df_E=N-k$ represents the within-group degrees of freedom (N being the total sample size).

$$MSE=\frac{SSE}{n-k} \quad (18)$$

Calculate the test statistic F-value, as shown in Table 5

When H_0 holds true, the ratio of the between-group mean square (MSA) to the within-group mean square (MSE) follows an F-distribution with numerator degrees of freedom $k-1$ and denominator degrees of freedom $n-k$ expressed as:

$$F=\frac{MSA}{MSE}\sim F(k-1,n-k) \quad (19)$$

Make a decision

If $F>F_0$, reject the null hypothesis H_0 , signifying that the discrepancies between the means are significant, and the factor under test exerts a significant influence on the observed values.

If $F<F_0$, fail to reject the null hypothesis H_0 , there is no evidence that the factor being tested has a significant effect on the observed values.

Table 5. One-way Analysis of Variance

Note. of Variation	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Square (MS)	F-value	P-value	Critical F-value
Between Groups	1456.61	3	485.54	3.4066	0.0388	3.1274
Within Groups	2708	19	142.53			
Total	4164.61	22				

Measurement of Relationship Strength

The ratio of the sum of squares of the independent variable to the total sum of squares is referred to as $\boldsymbol{R^2}$, namely:

$$R^2=\frac{SSA(\text{sum of squares between groups})}{SST(\text{ total sum of squares })} \quad (20)$$

The square root R can be utilized to assess the strength of the association between two variables. In the example,

$$R^2=34.98\%,\ R=0.59 \quad (21)$$

5. Multifactor Analysis of Variance

5.1 Multifactor Analysis of Variance

Multifactor analysis of variance is a statistical analysis method used to study the influence of two or more independent variables (factors) on two or more dependent variables, while considering the interaction between factors. It infers whether the influence of independent variables on dependent variables is significant by testing whether there are significant differences in the means of dependent variables among different groups, as shown in Table 6.

Table 6. Concept Comparison

Comparison Content	**One-way Analysis of Variance**	**Multifactor Analysis of Variance**
Definition	A test that allows people to compare the means of three or more groups of data.	A test that allows people to compare the means of three or more groups of data, with two independent variables taken into account.
Number of Independent Variables	One	Two
What is being compared?	The means of the dependent variable for three or more groups of independent variables.	The influence of multiple groups of two independent variables on the dependent variable and the mutual influence between them.

Number of Sample Groups	Three or more.	Each variable should have multiple samples.

Note. Collation of online materials

5.2 Application of One-way and Multifactor Analysis of Variance in Tourism

Lv Xingyang [5] et al. explored the functional relationship between occupational stigma and proactive customer service performance (PCSP), and used effects such as monetary compensation and organizational support to moderate the functional relationship between occupational stigma and proactive customer service performance. Under the condition of monetary compensation, PCSP under the condition of high occupational stigma is significantly lower than that under the condition of low stigma, and the effect size is relatively large (η^2=0.12), Among them, the F-value is the core indicator for measuring the significance of differences between groups in analysis of variance, and its value reflects the degree of differences between groups relative to the within-group error. Its value is much greater than 1 (F(1,178)=35.03), representingThe smaller the degree of change caused by random differences, the stronger the relationship strength, as shown in Table 7. It is worth noting that organizational monetary compensation is ineffective: neither its main effect nor its interaction effect is significant. Moreover, under both high and low stigma conditions, monetary compensation did not significantly improve PCSP, indicating that material compensation is difficult to alleviate the negative impact of occupational stigma on PCSP.

Table 7. Multifactor Analysis of Variance

Variable Type	**Organizational Monetary Compensation**		
Main Effect Analysis	Interaction Effect Analysis	High Stigma Comparison	Low Stigma Comparison
The main effect of occupational stigma is significant:	The interaction effect is not significant:	No compensation group vsMonetary compensation group:	No compensation group vs Monetary compensation group:
F(1,178)=35.03, p<0.001, η^2=0.16	F(1,178)=0.02, p=0.88, η^2<0.001	M=4.67 Vs 4.76	M=5.44 vs 5.49
The main effect of monetary compensation is not significant		F (1,89)=0.22, p=0.64, not significant	F(1,89)=0.11, p=0.74, not significant
F(1,178)=3.33, p=0.57, η^2=0.002			

Note. Literature compilation

Beyond categorical moderators, continuous features such as a streamer's influence and product-matching reliability can be modeled as moderators (centered) interacting with treatments; theory suggests high influence/high matching favors selective product strategies, while low-low favors non-selection, highlighting heterogeneous effects relevant to tourism retail partnerships (Zhen et al., 2024)[8].

Beyond categorical factors, narrative style (trainable) and AI-virtual streamer interaction design are actionable levers: optimize flow via professional/entertaining scripts, and match interaction type to product category-product-focused cues for utilitarian goods (via perceived value) and social-oriented cues for hedonic goods (via social presence). (Zhang et al., 2025; Liu et al., 2025)[10].

6. Conclusion

This study systematically integrates textbooks, online courses, and academic literature. Focusing on four core methods in tourism market research-moderation effect, mediation effect, one-way ANOVA (Analysis of Variance), and multivariate ANOVA-and integrating SPSS and R programming software, it deeply explains the theoretical logic, calculation process of each method, and its practical application in tourism scenarios (e.g., analysis of factors influencing homestay sales, study on differences in industry complaints). This provides methodological support for relevant research in the tourism field. Meanwhile, the study also points out current limitations: the lack of in-depth exploration of the linear regression foundation required for moderating and mediating effects, the application of R language mediating effect visualization tools (e.g., mediationPlot package), and the processing strategy for insignificant data in multi-way ANOVA. Further research will be conducted on these directions in the future.

References

[1] Robinson C , Schumacker R E .Interaction Effects: Centering, Variance Inflation Factor, and Interpretation Issues[J]. 2009.

[2] Chen D , Zhang W , Bi J ,et al.Hosts' online affinities and their impacts on the number of online reviews on peer-to-peer platforms[J].Tourism Management [2025-06-12].

[3] Chen Dongzhi, Qiu Hanqin. A Study on the Influence of Hosts' Self-reported Content on Homestay Sales -- From the Perspective of Social Cognitive Theory[J]. Tourism Science, 2024, 38(10): 100-127.

[4] Wang Caicai, Qiu Yisi, Xu Hong, et al. The Impact of Digital Transformation of Rural Homestays on Survival Resilience -- The Mediating Effect of "Dual-stage Dynamic Capability"[J]. Tourism Tribune, 2025(5).

[5] Lv X , Zhang K , Song Q ,et al.How does occupational stigma and work dirtiness inhibit hotel front-line employees' proactive customer service performance and its solutions[J].Tourism Management, 2024, 104(000).

[6] Qiu, H., Chen,D.*, Lyu, J., He, H. Li, C.(2021). Affinity-seeking strategies of homestay hosts: Scale development and validation, Journal of Hospitality and Tourism Management, 49(2021), 195-203.

[7] Yan, R., Tang, Z., & Liu, D. (2024). Can virtual streamers replace human streamers? The interactive effect of streamer type and product type on purchase intention. Marketing Intelligence & Planning, 43(2).

[8] Zhen, X., Wang, P., & Li, X. (2024). The Streamer’s sales strategy choice considering sales effort. Journal of Retailing and Consumer Services, 78, 103745.

[9] Zhang K, Luo B, Yan J, et al. The art and impact of narrative: How streamer narrative styles shape consumers’ purchase intention[J]. Journal of Retailing and Consumer Services, 2025, 87: 104357.

[10] Liu H, Zhang P, Cheng H, et al. Impact of AI-generated virtual streamer interaction on consumer purchase intention: A focus on social presence and perceived value[J]. Journal of Retailing and Consumer Services, 2025, 85: 104290.