

The Evaluation of Yili Kazak Autonomous State's Tourism Development Potential Based on Factor Analysis Model

Jie Zeng, Xiaoshi Yi

School of Mathematics and Statistic, Yili Normal University, Yining Xinjiang
Email: 1357304986@qq.com

Received: May 3rd, 2020; accepted: May 21st, 2020; published: May 28th, 2020

Abstract

Based on the theory of factor analysis, this paper makes a comprehensive analysis of Yili state's tourism development potential from 2007 to 2017, and explores the trends of Yili state's tourism development, and obtains regression prediction model by means of tourism income and two main factors to make principal component regression analysis. The results can provide a basis for the balanced development and designing of Yili state's tourism.

Keywords

Yili State, The Tourism Industry, Factor Analysis

基于因子分析模型的伊犁哈萨克自治州旅游业发展潜力评估

曾 杰, 易校石

伊犁师范大学, 数学与统计学院, 新疆 伊宁
Email: 1357304986@qq.com

收稿日期: 2020年5月3日; 录用日期: 2020年5月21日; 发布日期: 2020年5月28日

摘 要

本文基于因子分析的原理对伊犁州2007~2017年旅游业发展潜力进行综合分析, 探寻伊犁州旅游业发展趋势, 并借助旅游收入和两个主因子做回归分析, 进而得到主成分回归模型, 研究结果可为伊犁州旅游

业均衡发展和规划提供决策依据。

关键词

伊犁州, 旅游业, 因子分析

Copyright © 2020 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

伊犁哈萨克自治州(简称伊犁州)位于中国的西北边陲,地处中亚内陆腹地,东北与俄罗斯、蒙古国接壤,西与哈萨克斯坦共和国交界,内邻克拉玛依市、阿克苏地区、巴音郭楞蒙古自治州、昌吉回族自治州和石河子市。总面积 35 万平方公里,相当于中国总面积的三十分之一,良好的自然生态、浓郁的民俗、突出的资源优势、深厚的历史文化,奠定了伊犁州旅游业无可比拟的发展优势。伊犁作为“一带一路”上的璀璨明珠,是向西开放的“桥头堡”,它在古代丝绸之路之上就占有相当重要的地位,处于丝绸之路经济带大通道的核心,既是连接我国和这条经济带的枢纽,又是建设丝绸之路经济带的前沿,在新时代的背景下,它更被赋予了伟大的责任与使命,是非常有发展潜力和发展条件的地方。

旅游产业属于综合性产业,它的发展必然带动餐饮、住宿、交通、文化、娱乐等产业的发展。近年来,伊犁州党委深入贯彻落实党中央、自治区党委决策部署,立足伊犁丰富的旅游资源优势,加强旅游基础设施和景区景点建设,积极开发旅游线路、旅游商品,全力推进自然风光游、民俗风情游和冬季冰雪游等,使得旅游产业迅猛发展。

目前,关于旅游产业发展潜力评估的研究取得了一定进展[1] [2],但影响旅游业发展的因素颇多,所以影响旅游业发展的指标选取没有统一的标准与规范[3],因此还需要进一步的深入研究。鉴于此,本文旨在吸收前人优秀成果的基础上[4],先选用因子分析法,对伊犁州旅游产业发展潜力进行综合分析,再利用旅游收入与两个主因子做主成分回归分析,得出回归预测模型,进而评估伊犁哈萨克自治州旅游产业发展潜力。

2. 研究方法

旅游产业综合性强,其发展受众多因素的综合影响,发展潜力评估需要采用多指标综合评价体系,其中因子分析法是一种典型、高效的数据降维方法,在面对诸多具有内在相关性的变量时,因子分析试图使用少数几个随机变量来描述多个变量所体现的一种基本结构。具体表示为:

$$\begin{cases} X_1 = a_{11}F_1 + a_{12}F_2 + \dots + a_{1m}F_m + \varepsilon_1 \\ X_2 = a_{21}F_1 + a_{22}F_2 + \dots + a_{2m}F_m + \varepsilon_2 \\ \vdots \\ X_p = a_{p1}F_1 + a_{p2}F_2 + \dots + a_{pm}F_m + \varepsilon_p \end{cases}$$

用矩阵表示为

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pm} \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix}$$

其中, X_1, X_2, \dots, X_p 为 p 个原始变量, a_{ij} 为第 i 个变量在第 j 个因子上的载荷, F_1, F_2, \dots, F_m ($m < p$) 为因子变量, ε_i ($i=1, 2, \dots, m, m < p$) 为特殊因子, 即

$$X = AF + \varepsilon.$$

本文使用主成分因子提取方法[5], 选取累积方差贡献率大于 85% 时若干公因子用于综合评价, 构造综合评价函数为:

$$F = \sum_{j=1}^m \left(\frac{\lambda_j}{\sum_{i=1}^m \lambda_i} \right) F_i. \quad (1)$$

其中, λ_i 是公因子 F_i 对应的特征值。

3. 实证分析

3.1. 主要指标与数据的选取

本文选取伊犁州旅游产业的 10 个指标, 分别为: X_1 旅游 A 级景区、 X_2 第三产业就业人数、 X_3 旅游星级饭店数、 X_4 旅行社、 X_5 载客汽车、 X_6 旅客周转量、 X_7 社会消费品零售总、 X_8 第三产业投资总额、 X_9 人均生产总值、 X_{10} 互联网户数, 并根据 2007~2017 年伊犁哈萨克自治州统计年鉴[6]、伊犁哈萨克自治州国民经济和社会发展统计公报[7]得到以上 10 个指标的原始数据, 如表 1 所示:

Table 1. Raw data

表 1. 原始数据

年份	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
2007	39	443,065	80	61	51,327	43.9	94.66	115.26	10,352	9.63
2008	45	464,027	94	63	62,897	50.13	109.79	109.2	12,349	12.43
2009	46	484,944	100	63	75,058	54.78	131.05	145.1	15,054	20.64
2010	51	508,509	112	60	92,264	51.64	154.3	182.78	16,221	23.49
2011	55	524,996	124	51	120,734	54.9	172.67	245.06	19,479	31.02
2012	68	548,471	122	52	150,819	60	199.87	318.63	23,354	41.26
2013	63	587,108	124	57	191,995	65.84	232.56	502.28	26,801	54.75
2014	64	640,864	121	59	178,745	71.71	265.29	499.37	30,328	59.12
2015	65	694,619	126	82	183,937	52.51	296.29	729.99	32,824	85.95
2016	79	707,665	125	106	220,523	45.33	330.32	561.43	34,119	99.04
2017	119	767,415	127	124	261,210	40.06	353.44	565.86	33,120	107.19

3.2. 因子分析

首先, 根据所选指标, 搜集伊犁州 2007~2017 年统计数据, 由于指标的单位不一致, 因此运用 SPSS20.0 进行标准化处理。计算各指标相关系数矩阵, 可得出原始指标间具有较强相关性的结论, 各指标间存在线性关系。如表 2 所示:

Table 2. Test of KMO and Bartlett
表 2. KMO 和 Bartlett 的检验

取样足够度的 Kaiser-Meyer-Olkin 度量		0.639
	近似卡方	204.963
Bartlett 的球形度检验	df	45
	Sig.	0.000

样本的 KMO 检验值为 0.639, Bartlett 球形检验的近似卡方值为 204.963, 在自由度为 45 的条件下显著性概率 sig 为 0.000, 表明数据适合做因子分析。

其次, 采用主成分分析法并按照累积方差贡献率大于 85% 的标准选取主因子。如表 3 所示, 前两个公共因子的累积方差贡献率已达到 94.404%, 所以, 只需对前两个主因子进行分析即可。

Table 3. Total variance of interpretation
表 3. 解释的总方差

成份	初始特征值			提取平方和载入			旋转平方和载入		
	合计	方差的%	累积%	合计	方差的%	累积%	合计	方差的%	累积%
1	7.777	77.771	77.771	7.777	77.771	77.771	7.464	74.639	74.639
2	1.663	16.633	94.404	1.663	16.633	94.404	1.976	19.764	94.404
3	0.302	3.022	97.426						
4	0.172	1.725	99.151						
5	0.048	0.482	99.633						
6	0.022	0.216	99.848						
7	0.012	0.117	99.965						
8	0.003	0.032	99.997						
9	0.000	0.002	99.999						
10	0.000	0.001	100.000						

对指标降维得到成份矩阵表(表 4), 由于因子在原始变量上的载荷不易解释, 选用方差最大化正交旋转法进行因子旋转, 可得旋转成份矩阵表如表 5 所示。

Table 4. Component matrix
表 4. 成份矩阵^a

	成份	
	1	2
Zscore (旅游 A 级景区)	0.893	-0.234
Zscore (第三产业就业人数)	0.993	-0.029
Zscore (旅游星级饭店数)	0.817	0.422
Zscore (旅行社)	0.769	-0.619
Zscore (载客汽车)	0.98	0.098
Zscore (旅客周转量)	-0.083	0.969
Zscore (社会消费品零售总额)	0.996	0.033
Zscore (第三产业投资总额)	0.922	0.216
Zscore (人均生产总值)	0.97	0.208
Zscore (互联网户数)	0.991	-0.082

Table 5. Rotation component matrix
表 5. 旋转成份矩阵^a

	成份	
	1	2
Zscore (旅游 A 级景区)	0.817	0.43
Zscore (第三产业就业人数)	0.961	0.253
Zscore (旅游星级饭店数)	0.892	-0.226
Zscore (旅行社)	0.609	0.777
Zscore (载客汽车)	0.977	0.126
Zscore (旅客周转量)	0.139	-0.963
Zscore (社会消费品零售总额)	0.978	0.193
Zscore (第三产业投资总额)	0.947	-0.002
Zscore (人均生产总值)	0.992	0.017
Zscore (互联网户数)	0.947	0.304

表中的数值表示对应指标在公因子上的载荷, 反映了指标与公因子之间的相关程度, 即公因子包含对应指标的信息量, 若载荷值越高表示解释力越强。对旋转成份矩阵分析, 可将指标集分为两个主因子, 第一主因子在第三产业就业人数、社会消费品零售总额、人均生产总值等方面显著, 第二主因子在旅行社、旅客周转量方面显著。

最后, 用回归法估计得到成份得分系数矩阵表如表 6 所示, 即可得出因子分析模型和因子得分函数:

$$\begin{aligned}
 X_1 &= 0.8F_1 + 0.163F_2 + \varepsilon_1, \\
 X_2 &= 0.12F_1 + 0.046F_2 + \varepsilon_2, \\
 &\vdots \\
 X_{10} &= 0.113F_1 + 0.077F_2 + \varepsilon_{10}.
 \end{aligned}$$

$$\begin{aligned}
 F_1 &= 0.08X_1 + 0.12X_2 + 0.16X_3 + 0.012X_4 + 0.136X_5 + 0.122X_6 \\
 &\quad + 0.129X_7 + 0.145X_8 + 0.15X_9 + 0.113X_{10}, \\
 F_2 &= 0.163X_1 + 0.046X_2 - 0.223X_3 + 0.385X_4 - 0.029X_5 - 0.57X_6 \\
 &\quad + 0.01X_7 - 0.1X_8 - 0.094X_9 + 0.077X_{10}.
 \end{aligned} \tag{2}$$

Table 6. Component score coefficient matrix
表 6. 成份得分系数矩阵

	成份	
	1	2
Zscore (旅游 A 级景区)	0.080	0.163
Zscore (第三产业就业人数)	0.120	0.046
Zscore (旅游星级饭店数)	0.160	-0.223
Zscore (旅行社)	0.012	0.385
Zscore (载客汽车)	0.136	-0.029

Continued

Zscore (旅客周转量)	0.122	-0.570
Zscore (社会消费品零售总额)	0.129	0.010
Zscore (第三产业投资总额)	0.145	-0.100
Zscore (人均生产总值)	0.150	-0.094
Zscore (互联网户数)	0.113	0.077

根据(1)式计算每个年份的因子综合得分, 表达式为:

$$F^{(i)} = \frac{0.746}{0.746+0.198} F_1^{(i)} + \frac{0.198}{0.746+0.198} F_2^{(i)}, \quad i = 1, 2, \dots, 11. \quad (3)$$

其中 $F_1^{(i)}$ 表示第 i 年在第一个公因子上的得分, $F_2^{(i)}$ 表示第 i 年在第二个公因子上的得分。

由(2)、(3)式计算可得伊犁州 2007~2017 年的旅游产业发展潜力指数, 通过图 1, 可观测出伊犁州旅游产业发展潜力是一个上升的过程。2007~2010 年伊犁州旅游产业发展潜力上升趋势缓慢; 2013~2014 年产业发展略有停顿; 2014 年后产业发展呈现线性增强态势, 体现出较大的发展潜力空间。发展潜力指数从 2007 年-1.065 持续上升, 在 2012 年由负转正, 到 2017 年已增至 1.439。总体而言, 伊犁旅游发展的总体趋势和规模必然增长和扩大, 若随着全国旅游形式的高涨, 增速有可能进一步提高。

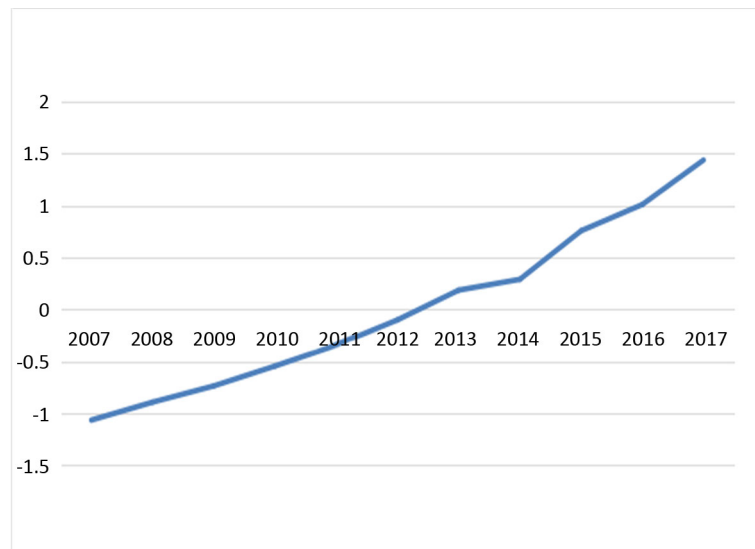


Figure 1. Evolution of the development potential of the tourism industry in Yili State

图 1. 伊犁州旅游产业发展潜力演变

3.3. 回归分析

经过上述的因子分析, 根据两个主因子和旅游业收入(Y) (单位: 亿元), 利用回归分析原理建立回归方程[8] [9]。

假设两个主因子为自变量 F_1, F_2 , 旅游业收入为因变量 Y , 则得出回归直线方程式:

$$Y = \beta_0 + \beta_1 F_1 + \beta_2 F_2.$$

用 SPSS 软件继续对上述数据进行处理, 得出相关数据如下:

Table 7. Coefficient
表 7. 系数

模型	非标准化系数		标准系数		t	Sig.	共线性统计量	
	β	标准误差	试用版				容差	VIF
(常量)	-0.074	0.080			-0.926	0.385		
1	f_1	0.740	0.091	1.034	8.121	0.000	0.844	1.184
	f_2	0.311	0.106	0.376	2.949	0.021	0.844	1.184

由表 7 得: $\beta_1 = 0.74$, $\beta_2 = 0.311$, $\beta_0 = -0.074$ 。则回归方程为:

$$y = 0.74f_1 + 0.311f_2 - 0.074.$$

对上述模型分别进行模型拟合度检验和 F 检验, 具体如表 8、表 9:

Table 8. Model summary
表 8. 模型汇总

模型	R	R 方	调整 R 方	标准估计的误差
1	0.951 ^a	0.904	0.877	0.23816575

Table 9. Analysis of variance
表 9. 方差分析

模型	平方和	df	均方	F	Sig.	
1	回归	3.745	2	1.873	33.012	0.000 ^b
	残差	0.397	7	0.057		
	总计	4.142	9			

回归方程的决定系数 $R^2 = 0.904$, 说明回归方程与观察值之间的拟合优度很好, 另外, 取显著性水平 $\alpha = 0.01$, 查 F 分布表有 $F_{0.01}(2,7) = 9.55$ 且 $F > F_{0.01}(2,7)$, 说明回归方程的回归效果显著, 该回归方程具有较好的实用性。

将原始数据变量 X_1, X_2, \dots, X_{10} 代入回归方程可得出旅游收入关于旅游 A 级景区、第三产业就业人数、旅游星级饭店数等的主成分回归方程:

$$Y = 0.4493X_1 + (8E - 0.5)X_2 + 0.2782X_3 + 0.4876X_4 + 0.0001X_5 - 0.83X_6 \\ + 0.0988X_7 + 0.0314X_8 + 0.0008X_9 + 0.2774X_{10} - 97.45$$

用 2017 和 2018 年的旅游收入做预测如表 10 所示, 从预测值和观测值的比较来看, 模型对数据的预测误差较小, 表明该模型的预测效果较好, 可以用来做实际预测。

Table 10. Prediction and estimation for 2017-2018
表 10. 2017~2018 年预测估计

年份	观测值	预测值	误差率	预测区间上限值	误差率
2017	297.85	224.598	0.2459	336.72	0.131
2018	363.97	270.862	0.2558	382.984	0.052

4. 结语

综上所述, 伊犁州旅游业发展经过 11 年的不断积累, 发展速度由缓慢增长到快速增长, 发展潜力指数从负值增长为正值。得益于中央的大力支持和对口援疆工作积极开展等一系列措施的共同作用。限制伊犁州旅游业发展的因素主要表现为: 资金匮乏, 导致旅游产品开发不力, 缺乏强吸引力; 景区外交通住宿问题; 吸引外资能力等, 这些限制因素的改变将会促使伊犁州旅游业发展进一步扩大。

利应用因子分析方法消除多重共线性后, 用两个综合变量进行回归, 使问题简化, 得到旅游业收入(Y)的回归预测模型, 可为伊犁州拓展旅游市场, 制定旅游发展规划, 促进旅游业发展提供一定的科学依据。

参考文献

- [1] 杨敏. 青海旅游产业的发展潜力评估[J]. 统计与决策, 2006(14): 102-104.
- [2] 倪向丽. 我国省际旅游产业竞争力因子聚类评估[J]. 统计与决策, 2018(21): 54-56.
- [3] 张洪, 孙娟, 朱琼琳. 基于因子与聚类分析的安徽旅游经济发展潜力研究[J]. 资源开发与市场, 2013(11): 1208-1212.
- [4] 高凤伟, 杨芳, 许世泽. 基于因子分析模型的安徽省城市旅游竞争力综合评价[J]. 佳木斯大学学报(自然科学版), 2016(3): 494-496.
- [5] 肖启华, 黄硕琳, 王慰. 多指标面板数据因子分析的分层模型及应用[J]. 数学的实践与认识, 2015(12): 86-93.
- [6] 伊犁州统计局. 伊犁统计年鉴[M]. 伊宁: 伊犁州统计局, 2007-2016.
- [7] 伊犁州统计局. 伊犁哈萨克自治州国民经济和社会发展统计公报[R]. 2017.
- [8] 程毛林. 基于主成分回归模型的经济增长因素分析[J]. 运筹与管理, 2012(1): 175-179.
- [9] 沈菊红. 预测农民收入的主成分回归模型[J]. 安徽农业科学, 2008(24): 10285-10287.