DOI：10.19473/j.cnki.1008-4940.2019.01.007　网络出版时间：2019-03-27 09:53:55
网络出版地址：http://kns.cnki.net/kcms/detail/35.1333.g4.20190326.1610.014.html

基于生存模型的酒店客人住宿时间影响因素分析

——以武汉市酒店为例

徐松华，鲁婉婷

（武汉商学院　武汉旅游研究院，湖北　武汉，430056）

［摘　要］以 2018 年武汉酒店住客为研究对象，运用生存分析中的寿命表法、Kaplan-Meier 法及 Cox 回归法，分析酒店客人住宿时间及影响因素。结果显示：客人在酒店住宿的平均时间为 2.65*d*，且与客人职业、来汉次数、出行目的、花费水平和服务满意度等 5 个因素显著相关，而与客人性别、年龄、来源地和出行方式因素无显著相关；花费水平低、来汉次数少、服务满意度高能显著延长客人在酒店的住宿时间；职业为公务员、企管人员，以及商务会议的客人在酒店住宿时间相对较长；而职业为工人和农民，以及观光游览的客人在酒店住宿时间相对较短。

［关键词］住宿时间；生存分析；Cox 风险模型；影响因素

［中图分类号］K902　［文献标识码］A　［文章编号］2096-3300（2019）01-0044-07

自 2013 年以来，在中国经济增速放缓和酒店消费逐渐回归“社会理性”的大背景下，国内酒店尤其是高端酒店遭遇经营“寒潮”，部分高档酒店收入大幅下滑，经营困难。2014 年“中国五星级酒店品牌的现状及发展趋势”高峰论坛提供的数据显示，2013 年国内五星级酒店营业额普遍下降三成以上。作为酒店经营的重要指标，客人住宿时间在一定程度上反映了酒店的吸引力以及客人对酒店的整体满意度，也直接决定了客人在酒店入住期间的餐饮、住宿等消费支出。

当前关于酒店住宿时间的研究，国内关注的焦点集中在“住宿时间结算”的问题上，对酒店客人住宿时间的分布特征及影响因素的研究还不多见。酒店客人住宿时间是一个持续时间型数据，具有生存数据的基本特征。一是客人住宿时间经常以入住天数来统计，是一个非负的变量；二是在随访期内，部分客人没有出现“结账退房”失效事件，即存在删失数据。因此，适合应用生存分析模型对其进行研究。

一、研究方法

生存分析法是研究所关注事件在什么时间发生的一种事件数据分析方法，是一种既考虑结果又考虑随访时间的处理生存数据的方法。生存数据包含生存时间、观测结果以及相关因素。其中生存时间是一个非负变量，应用经典统计工具如线性回归模型会导致有偏的估计，因而并不适宜做生存时间的模拟；同时，生存数据中一定删失数据的存在，严重违反传统统计模型的假设，也会造成显著的偏

收稿日期：2018-12-13
基金项目：湖北省教育厅科学技术研究项目“武汉国内游客市场空间结构演变及开拓”（B2017282）。
作者简介：徐松华（1975-），男，湖北孝感人，讲师，硕士，研究方向：旅游市场与资源开发；
鲁婉婷（1983-），女，湖北武汉人，讲师，硕士，研究方向：酒店管理及教育。

误[1]。HELSEN 等[2]指出，生存分析方法在处理持续时间型数据时具有优越性，尤其是在处理删失数据时更具有无法替代的作用[3]，最近几年国内外开始将之应用于旅游研究中来分析游客停留时间问题[4-7]。生存分析法主要基于以下基本原理。

1. 生存函数（Survival Function）是反映个体生存时间超过时间 t 的概率，记做 $S(t)$。若无删失数据，则 $S(t)=P(T\geq t)$ = 过了 t 时刻仍存活的个数/观察开始时的总个数，其中 t 为个体的存活时间。但如果资料中含有删失数据，生存率的计算公式应为：

$$S(t_k)=P(T\geq T)=P_1\cdot P_2\cdots P_k \quad (1)$$

其中 P_1、$P_2\cdots P_k$ 表示不同时间段的生存概率，可以看出，这种情况下生存率是多个时段生存概率的累积，故又称为累积生存概率（Cumulative Probability of Survival）。当 $t=0$ 时，生存函数取值为 1，随着时间推移（t 逐渐增大），生存函数的取值逐渐减小。因此，生存函数是时间 t 的单调递减函数。

2. 非参数分析

非参数分析方法不引入任何的外生变量，包括寿命表法（Life Table）和 Kaplan-Meier 估计。其中寿命表法适用于观察例数较多而分组的资料，通过计算落入时间区间 $[t_{k-1}, t_k]$ 内的失效和删失的观察个数来估计该区间上的死亡概率，然后用该区间及之前各区间上的生存概率之积来估计 $S(t_k)$[8]。Kaplan-Meier 估计又称乘积极限法（Product-Limit Method），于 1958 年由卡普兰（Kaplan）与迈耶（Meier）提出，主要用于观察例数较少而未分组的生存资料，是利用条件概率与概率的乘法原理计算生存率及其标准误的。

$$S(t_i)=S(t_{i-1})\ S(t_i/t_{i-1}) \quad (2)$$

其中 $S(t)$ 表示 t 年的生存率，$S(t_i/t_{i-1})$ 表示活过 t_{i-1} 年又活过 t_i 年的条件概率。

3. Cox 比例风险模型。

由于生存分析模型中的参数估计法对生存函数分布有假设限定，若假设限定有误，那么估计的准确性将会下降；而半参数法只规定影响因素和生存状况之间的关系，不对生存函数的分布情况作出限定，是一种研究生存概率影响因素的多因素分析方法。对于一批生存数据，在事先不知道寿命分布的总体趋势，且又不好判断应该用何种模型最合适时，多数学者一般直接采用非参数方法或半参数法。因此，作为半参数分析的代表性方法，Cox 比例风险模型近年来得到了快速的发展。该模型将风险概率 $H_i(t)$ 建模在时间 t 上的基准概率 $h_0(t)$ 和影响因素向量 X 的函数之上，即：

$$H_i(t)=h_0(t)\cdot\exp^{(((\beta_1)(X_{i1}))+((\beta_2)(X_{i2}))+\cdots+((\beta_k)(X_{ik})))} \quad (3)$$

其中，$H_i(t)$ 指 t 时刻风险函数、风险率或瞬时死亡率，$h_0(t)$ 是基准的生存分布危险函数，即所有变量都取 0 时 t 时刻风险函数。X_{i1}，X_{i2}，…，X_{ik} 为预后变量向量，β_1，β_2，$\cdots\beta_k$ 为回归系数向量。

Cox 模型以半参数方程回归方式对风险作出估计，并得到 β 的极大似然估计值，作为各影响因素的风险比系数。通过系数 β 可以得出该因素是保护因素还是危险因素、相对危险度的大小，其中 $RR=\exp^{(\beta)}$。若 $\beta>0$，$RR>1$，说明变量 X 增加时，危险率增加，即 X 是危险因素；$\beta<0$，$RR<1$，说明变量 X 增加时，危险率下降，即 X 是保护因素；$\beta=0$，$RR=1$，说明变量 X 增加时，危险率不变，即 X 是危险无关因素。本文采用生存分析中的 Cox 回归模型进行分析，对酒店客人住宿时间的影响因素进行估计。

二、研究设计

（一）研究假设

根据国内外对游客停留时间的研究结果，假设客人住宿时间与人口学特征、出行特征、消费特征和服务质量等四个维度的解释变量有直接关系。假设之一，客人住宿时间由客人的人口学特征决定，研究测定的人口学特征变量包括来源地、年龄、性别、职业；假设之二，客人住宿时间与客人的出行特征存在关联，研究测定的出行特征变量包括出行目的和出行方式；假设之三，客人住宿时间受客人的消费特征影响，研究测定的消费特征变量包括花费水平（人均天花费额）、来汉次数；假设之四，客人住宿时间受酒店服务质量控制，研究测定的酒

店服务质量变量仅包括客人服务满意度，因为服务满意度是客人对酒店服务质量集中和综合的反映。上述变量测度分为 2 种，住宿天数属连续型变量；性别、年龄、职业、来源地、出行动机、出行方式为分类变量；花费水平、来汉次数和服务满意度属序次变量。观测变量的解释与调查结果基本数据如表 1。

表 1 统计变量说明

Tab. 1 Description of statistical variables

	一级变量	二级变量	描述
统计变量	人口统计特征	来源地	省内=1；华北=2；华东=3；华中=4；其它=5
		性别	男=1；女=2
		年龄	24 岁及以下=1；25~44 岁=2；45 岁及以上=3
		职业	公务员=1；企事业管理人员=2；工人或农民=3；科教人员=4；服务销售人员=5；学生=6；其他=7
	消费特征	来汉次数	1 次=1；2~3 次=2；4 次及以上=3
		花费水平（人均天花费额）	380 元及以下=1；380~1 000 元=2；1 000 元以上=3
	出行特征	出行目的	休闲度假=1；观光游览=2；商务会议=3；其他=4
		出行方式	旅行社随团=1；自驾车=2；其他=3
	服务质量	服务满意度	不满意=1；满意=2；很满意=3

（二）变量设计

生存分析模型设计以 10d 为随访期，在随访期内调查客人是否发生“结账退房”为“失效”事件，客人结账退房（即“失效”事件）时在酒店住宿天数即为生存时间。客人在酒店入住天数超过随访期，即入住天数超过 10d 的住店客人定义为删失数据。在随访期内，客人是否办理结账退房为生存状态变量，该变量有两个水平，变量标记为：1=客人已结账退房；0=删失。

（三）数据来源

在武汉主城区随机选取 20 家不同档次、类型的宾馆、酒店作为调研地点，以各酒店结账退房客人为调研对象，以面对面的方式对客人进行问卷调查。从 2018 年 3 月至 2018 年 8 月，共投放调查问卷 1 450份，回收调查问卷 1 362 份，回收率为 93.93%；经程序录入审核，获得有效问卷为1 271份，有效率 93.32%。

（四）数据处理与分析

本文采用 SPSS17.0 进行统计学处理。首先应用寿命表法分析客人在酒店住宿时间的总体分布规律，表 3 是在 10d 的随访期内武汉酒店客人住宿时间的寿命表分析结果，其中客人在酒店住宿的平均时间为 2.65d，而且 50%武汉客人在酒店住宿时间不会超过 2.79d。

从表 2 可以看出：客人入住酒店 1d 后，就有“结账退房”终点事件发生，依据发生频率可分为 3 个阶段：（1）高发期，集中于［1，2）、［2，3）、［3，4）这 3 个时间区间内，占到完全数据的 83.01%，相应地，期末累积生存比例下降趋势明显，下降速度较快，这就表明大部分住店客人或因行程安排，或对酒店不太满意，而都选择将入住天数控制在 3d 以内；（2）缓和期，集中于［4，5）、［5，6）、［6，7）、［7，8）4 个时期内，占到完全数据的 15.42%，期末累积生存比例下降速度减缓，表明该部分客人多是因为行程原因而选择结账退房，对酒店的满意度在逐渐积累，有在酒店较长住宿的强烈意愿；（3）平滑期，其余剩下时间区间内武汉客人住店生存率下降得更为平缓，3 个时间区间内结账退房的客人仅占完全数据的 1.57%。这表明客人入住酒店超过 7d 后，对酒店的内外环境逐渐熟

悉，服务质量逐渐满意，长住酒店的意愿上升，有很大可能成为酒店的长住客。

表 2　客人住宿天数寿命表

Tab. 2　Life table of guests' accommodation time in Wuhan hotels

期初时间	期初记入数	期内退出数	有效例数	期内终结数	终结比例/%	生存比例/%	期末累积生存比例	期末累积生存比例标准误差	风险率/%	风险率标准误
1	1271	0	1271	266	21	79	0.79	0.01	23	0.01
2	1005	0	1005	465	46	54	0.42	0.01	60	0.03
3	540	0	540	324	60	40	0.17	0.01	86	0.04
4	216	0	216	88	41	59	0.10	0.01	51	0.05
5	128	0	128	72	56	44	0.04	0.01	78	0.08
6	56	0	56	13	23	77	0.03	0.01	26	0.07
7	43	0	43	23	53	47	0.02	0.00	73	0.14
8	20	0	20	2	10	90	0.01	0.00	11	0.07
9	18	0	18	0	00	100	0.01	0.00	00	0.00
10	18	10	13	8	62	38	0.01	0.00	00	0.00

备注：生存时间的中位数为 2.79*d*。

图 1 更为直观地显示：随访期的 1～3*d* 所对应的生存函数降幅较大，从各段生存率之间的高度差可以明显看出；4～7*d* 考察范围内生存函数阶梯状高度差减小，表明降幅变缓；7*d* 之后各时期降幅更小，最终几乎演变为一条直线。

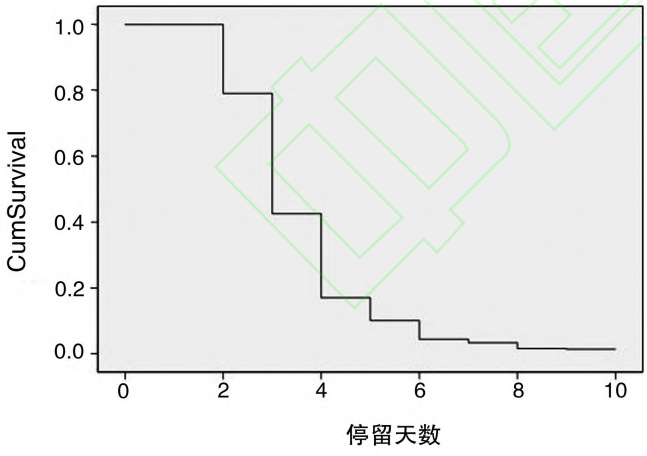

图 1　生存分布函数

Fig. 1　Survival function of guests in Wuhan hotels

三、生存模型的实证分析

（一）影响因素的确定

采用 Kaplan-Meier 法分别检验来源地、职业等 9 个变量对酒店客人住店时间的影响显著性。为了稳健起见，分别使用 LogRank test、Breslow test、Tarone-Ware test 三种检验方式。它们的区别是赋予观测权重的方式不同，其中 LogRank test 各时间点权重一样，此法最常用；Breslow 以各时间点的观察例数为权重；Tarone-Ware 以各时间点观察例数的平方根为权重。显著性检验结果如表 3 所示。三种检验方式所得结果保持一致：客人职业、来汉次数、出行目的、消费水平和服务满意度 5 个变量的统计显著性水平检验值（Sig.）均小于 0.01，达到较高的显著性，表明这 5 个因素是影响客人住宿时间的重要因素；而客人性别、年龄、来源地和出行方式 4 个变量则相反，未通过显著性水平检验，则表明这 4 个因素不是客人住宿时间的影响因素。

（二）Cox 生存模型的构建

采用 Cox 回归分析方法将职业、来汉次数、出行目的、人均天花费和服务满意度 5 个经 Kaplan-Meier 单因素检验呈现出显著性的影响因素变量进行预后预测检验。为消除纳入模型中各指标之间可能具有的共线性影响，采用基于偏最大似然估计的向前法（Forward：LR）选择自变量进入 Cox 回归方程，建立 Cox 比例风险模型。最后，还需从单个参数与模型整体两个方面对 Cox 比例风险模型的显著

性进行检验，其中单个参数检验采用 Wald 检验，整体检验采用 Score 检验，两种检验方式均包括卡方值、自由度、显著性 3 个检验结果。

表 3 Kaplan-Meier 单因素显著性检验结果

Tab. 3 Single factor significance test of Kaplan-Meier

影响因素	LogRank			Breslow			Tarone-Ware		
	卡方	df	*P* 值	卡方	df	*P* 值	卡方	df	*P* 值
来源地	1.762	4	0.779	2.223	4	0.695	2.313	4	0.678
性别	0.911	1	0.340	0.920	1	0.337	0.859	1	0.354
年龄	1.294	2	0.524	1.554	2	0.460	1.518	2	0.468
职业	57.778	6	0.000	55.145	6	0.000	59.994	6	0.000
来汉次数	19.813	2	0.000	36.049	2	0.000	30.735	2	0.000
消费水平	114.22	2	0.000	150.92	2	0.000	138.33	2	0.000
出行目的	106.6	3	0.000	79.429	3	0.000	93.198	3	0.000
出行方式	1.678	2	0.432	1.843	2	0.398	1.844	2	0.398
服务满意度	376.651	2	0.000	287.88	2	0.000	336.15	2	0.000

（三）Cox 生存模型的拟合

纳入偏最大似然估计向前逐步回归的数据共 1 271个，占全部数据的 100%。其中完全事件数为 1 261个，占 99.21%；删失数为 10 个，删失率为 0.79%；无数据在逐步回归过程中被剔除。此外，经过 5 步向前回归，服务满意度、消费水平、旅游目的、旅游次数和职业 5 个变量 Sig. 值均小于 0.05，满足显著性检验标准，依次先后进入 Cox 比例风险模型，最终结果见表 4。从模型参数的整体检验方面来看，Score 检验的卡方值为 304.679，自由度为 12，Sig. 值小于 0.001，检验结果显著；从模型的单个参数检验结果来看，纳入模型的所有协变量参数估计值的 Sig. 值都小于显著水平 0.05。因此，模型参数的显著性无论单变量检验还是总体检验效果都比较理想，在 5%的显著性水平下，Cox 比例风险模型假设成立。

（四）结果分析

Cox 比例风险模型中影响因素的参数估计结果见表 4。从表 4 可以看出：

1. 变量“花费水平”的偏回归系数 *B* 为 0.254 >0，表明“花费水平”是客人住店时间的危险性因素，即客人花费水平越高，倾向于在酒店住宿的时间会越短，而且消费水平每提高一个等级，客人结帐离店的可能性会提高 28.9%。这主要是花费水平较高的客人，受旅行支出预算的限制，在酒店住宿时间就相应缩短。

2. 变量“来汉次数”的偏回归系数 *B* 为 0.139 >0，表明“来汉次数”是客人住店时间的危险性因素，即多次来汉的客人，在汉住宿的天数会越来越少，而且来汉次数每提高一个等级，客人结帐离店的可能性要提高 14.9%。这主要是随着客人来汉次数的增多，武汉的旅游吸引力会相对减弱，在武汉逗留的时间会逐渐缩短，进而在酒店住宿时间也就相应缩短。

3. 变量“服务满意度”的偏回归系数 *B* 为 −0.797<0，表明“服务满意度”是客人住宿时间的保护性因素，即对酒店服务越满意的客人，其在酒店住宿的时间会延长，而且服务满意度每提高一个等级，客人在酒店继续住宿而不结帐退房的机会要提高 54.9%。这显然与酒店经营实际相契合，客人对酒店服务质量越满意，就会把酒店当作自己的家，从而增加客人在酒店住宿的愉悦感，所以在酒店住宿时间也就相应延长。

4. 变量“职业”为 2（企管人员）、3（工人和农民）、4（文教人员）、5（服务销售人员）、6（学生）和 7（其它人员）的偏回归系数 *B* 分别为

0.263、0.482、0.358、0.437、0.300 和 0.384，均大于 0，表明它们是客人住宿时间的危险性因素，即相对于公务人员类客人，以上几类职业的客人在酒店住宿时间短。其中最短的是服务销售人员，其次是工人和农民类客人。这主要是公务人员、企管人员多是公务性花费为主，出行计划性较强；而工人和农民多是自费，出行较自由，所以他们的酒店住宿时间差异比较明显。

表 4　Cox 比例风险模型中影响因素的参数估计结果

Tab. 4　Parameter estimation of influencing factors in Cox proportional hazards model

因素水平	偏回归系数 B	标准误差	Wald 统计量	自由度	检验 P 值	Exp（B）	95%置信区间	
							下限	上限
花费水平	0.254	0.042	37.409	1	0.000	1.289	1.189	1.399
来汉次数	0.139	0.040	11.867	1	0.001	1.149	1.062	1.244
服务满意度	−0.797	0.064	154.138	1	0.000	0.451	0.398	0.511
职业			14.697	6	0.023			
职业（1）	0[a]			1				
职业（2）	0.263	0.124	4.533	1	0.033	1.301	1.021	1.658
职业（3）	0.482	0.167	8.306	1	0.004	1.620	1.167	2.249
职业（4）	0.358	0.144	6.229	1	0.013	1.431	1.080	1.896
职业（5）	0.437	0.133	10.743	1	0.001	1.548	1.192	2.011
职业（6）	0.300	0.148	4.075	1	0.044	1.349	1.009	1.805
职业（7）	0.384	0.129	8.796	1	0.003	1.468	1.139	1.892
出行目的			40.544	3	0.000			
出行目的（1）	0[a]			1				
出行目的（2）	0.247	0.077	10.294	1	0.001	1.280	1.101	1.488
出行目的（3）	−0.264	0.083	10.048	1	0.002	0.768	0.652	0.904
出行目的（4）	0.086	0.085	1.026	1	0.311	1.090	0.922	1.289

注：a 表示哑变量编码方式为 Indicator，并且以最先一个变量值为参照基准（估计系数设为 0）；−2 似然对数 = 15980.363，整体卡方值 = 304.679，自由度 = 12，检验 P 值 = .000。

5. 旅游目的为“观光游览”客人的偏回归系数 B 为 0.247>0，表明这类客人相对于休闲度假客人，在酒店住宿的时间要相应缩短；旅游目的为“商务会议”客人的偏回归系数 B 为−0.269<0，表明该类客人相对于休闲度假客人，其在酒店住宿时间相对延长；旅游目的为“其他”客人的偏回归系数 B 为 0.086>0，未通过显著性检验，表明该类客人相对于休闲度假客人，在酒店住宿时间上没有显著性差异，这主要是商务会议类客人行程计划性强，出行自主性差，而观光游览客人走马观花，武汉只是其旅游目的地之一，在汉逗留时间不可能太长。因此不同旅游目的客人在酒店住宿时间差异比较明显。

四、结论与不足

本文以武汉酒店住宿客人为研究对象，运用生存分析中的寿命表法、Kaplan−Meier 法及 Cox 回归法，分析了酒店客人住宿时间及影响因素。结果显示：客人在酒店住宿，前三天是结帐退房的高峰期，有 83.01%的客人会在此期间选择结帐退房，而且 50%武汉客人在酒店住宿时间不超过 2.79d，所有客人在酒店住宿的平均时间为 2.65d。客人在酒店住宿时间与客人职业、来汉次数、出行目的、花费水平和服务满意度有显著相关性，而与客人性别、年龄、来源地和出行方式的相关性并不显著。花费水平、来汉次数是客人住宿时间的危险性因素；服务满意度是客人住宿时间的保护性因素；职业为公务员、企管人员类客人住宿时间相对较长，工人和农民类客人住宿时间相对较短；旅游目的为商务会议类客人住宿时间相对较长，而观光游览类客人住宿

时间相对较短。

生存分析在旅游科学研究的应用主要集中在游客停留时间方面，在酒店方面的应用研究还不多见。本文虽然采用问卷调查获得了第一手数据，但由于受多种主、客观条件的限制，很难采集各层次人群样本，样本的代表性有待提高，由此可能导致偏差产生；此外，客人在酒店住宿时间受到很多因素影响，一些变量可能没有在本文的分析中得以体现，需通过后续研究，运用不同学科知识与方法进一步挖掘和探寻。

参考文献：

[1] MORITA J G , LEE T W , MOWDAY R T. Introducing survival analysis to organizational researchers: a selected application to turnover research [J]. Journal of Applied Psychology, 1989, 74(2):280-292.

[2] HELSEN K, DAVID C S. Analyzing duration times in marketing: evidence for the effectiveness of hazard rate models [J]. Marketing Science, 1993,11(3):395-414.

[3] LI S. Survival analysis [J]. Marketing Research, 1995, 7(8):17-23.

[4] GOKOVALIA U, BAHARA O, KOZAK M. Determinants of length of stay: a practical use of survival analysis [J]. Tourism Management, 2007,28(3):736-746.

[5] THRANE C. Analyzing tourists' length of stay at destinations with survival models: a constructive critique based on a case study[J]. Tourism Management, 2012, 33(1):0-132.

[6] 方世敏,陈洁. 景区游客停留时间的影响因素研究[J]. 中南林业科技大学学报(社会科学版), 2013, 7(6):1-4.

[7] 斯建培, 钱波, 桂晶晶, 等. 西湖风景区游客逗留时间决定因素研究[J]. 浙江大学学报(理学版), 2012, 39(4):477-483.

[8] 吴冰. 生存分析及其应用——以创业研究为例[J]. 上海交通大学学报(哲学社会科学版), 2006(3):63-65.

A Study on Influencing Factors of Hotel Guests' Accommodation Time Based on the Survival Analysis

——A Case study of the hotels in Wuhan

XU Songhua, LU Wanting

(Institute of Tourism, Wuhan Business University, Wuhan 43005, China)

Abstract: Taking the guests in Wuhan hotels in 2018 as the research object, this paper analyzes the factors influencing the hotel guests' accommodation time by using the life table method, kaplan-meier method and Cox regression method in the survival analysis. The results show that the average length of stay in the hotel was 2.65d, which was significantly correlated with five factors, including the occupation of the guest, the number of visits to China, travel purpose, spending level and service satisfaction, but not significantly correlated with the factors of the guest's gender, age, place of origin and travel style. Low cost, few visits to the hotel, and improved service satisfaction can significantly extend the stay time of guests in the hotel. The accommodation time for civil servants, business managers, and business conference guests in the hotel is relatively long, while workers and farmers, as well as tourists, spend relatively less time in hotels.

Key words: accommodation time; survival analysis; Cox model; influencing factors

(责任编辑：杨成平)