

茶馆“相”遇妙 沧海一“声”笑

相声观众偏好分析与付费意愿调查



团队名称：

研途璀璨

摘 要

相声作为优秀的非物质文化遗产，是中华民族传统文化的瑰宝。近年来，以德云社为代表的一系列相声剧场发展迅猛。借助微博、微信、抖音等文化社区平台传播，相声发展势如破竹，也促进了相声付费市场。与此同时，国家对中华优秀传统文化具有政策支持。2017 年中共中央办公厅印发了《关于实施中华优秀传统文化传承发展工程的意见》，提出打造以相声为代表的传统曲艺集中演出街区。这些政策的加持让我们对相声市场的发展前景持积极态度。

为了解相声观众的偏好与付费意愿，团队将调查对象锁定为国内相声观众。采用“线上+线下”的方式，线上采用配额抽样与方便抽样相结合的方法，在全国范围内七个地区发放调查问卷，实际发放问卷 1600 份，其中有效问卷 1286 份，有效率达 80.375%；线下则是对黑龙江省曲艺团进行深度访谈。问卷采集了相声观众的基本信息、观众对待相声的认知和偏好情况、观众对待相声的行为特征、已付费观看相声的观众态度以及观众对相声发展的前景展望五方面数据。

基于词云统计，通过后裔采集器对知乎、微博、百度贴吧三个大型网站进行文本挖掘，从中筛选出与相声相关的内容生成数据表将其导出，生成词云图。词云图显示观众对德云社、郭德纲、张云雷、天津、北京、春晚、演出等讨论较多。

基于情感词典，对采集到的文本进行分词处理后，通过情感词典对其进行情感倾向分析，最终将词语分成正面情感与负面情感两类，分别统计频率，其中正面情感占 65%。

基于多元有序 Logistic 回归模型，对影响受访者关于相声的了解程度的因素进行回归分析，选取性别、年龄、文化程度等 6 个影响因素。结果显示，男性群体对相声的了解程度普遍高于女性；31-40 岁年龄段以及高收入人群为最了解相声的人群；大众关于相声的了解程度基本不会因为文化程度的不同而变化；

基于 K-means 聚类，对受访的相声观众进行人群划分，最终分为三类。第一类人群为收入稳定的中年女性；第二类人群为收入较高的职场青年男性；第三类人群为对相声略有兴趣的男大学生。各类人群喜欢的相声题材、观看相声的途径、能够接受的付费价格区间和喜爱的相声演员各有不同。

基于 Text-CNN 模型，根据问卷中设置的填空题得到观众喜爱的相声演员特征的文本。经过分词和清洗，利用语义分析系统统计所有词语频数，挑选出词频最高的 30 个词语，利用词语之间的相关性建立模型，将主题分成三类，三个主题从不同角度反映了优秀相声演员的特质。

基于随机森林模型，研究各类要素对观众喜爱相声程度的影响并进行重要性排序。结果显示价格和体验性价比因素会在较大程度上影响观众对相声的喜爱程度；而排在最后两位的分别是幽默和演员功底，即观众无论多大程度上喜欢相声都会认为幽默和演员功底很重要。

基于结构方程模型，设置潜变量和可测变量，研究影响观众对相声付费意愿的因素。分析显示整体体验与氛围感知的影响程度最大，营销感知也具有一定影响。此外，幽默感与演员功底是观众最关注的两个因素，而消极因素中语言地方性太强等会严重打击观众对相声付费的意愿程度。

基于贝叶斯决策模型，从观众对相声的喜爱、了解和满意程度出发，探究出以下三个结论：喜爱程度的提升虽能带来相声付费意愿的大幅增长，但在超过50%时就会难以形成持续的市场增速；后续宣传营销带来的效果对于观众选择的影响程度会呈现越来越小的状态；观众观看相声的频率超过每月1-2次、同时对观看付费相声的满意次数达到2-3次后，付费意愿显著提升。

基于层次分析法，将影响观众选择相声团体的因素分为服务价值和服务产品两类。服务价值包括票价、营销推广以及品牌影响力；服务产品包括内容创新、演员、剧本以及文化内涵。结果显示，在服务产品方面，内容创新指标的权重最高，其次是文化内涵；在服务价值方面，品牌影响力的权重排序第三；票价与营销推广权重相对排名靠后。

基于云模型，将观众在付费观看表演前对各方面的期待值作为指标权重，求出权重系数。根据观众对各指标实际值的打分情况，通过逆向云发生器进行计算，并用正向云发生器生成足够的云滴以生成云图。最终得到观众对于相声内容的幽默有趣以及相声演员的功底比较满意，对于相声题材、现场互动、性价比、环境舒适度等一般满意。

基于以上结论，对相声观众提出如下建议：（1）不断提高艺术审美；（2）客观看待相声演员的表演，切勿带有“粉丝滤镜”；（3）尊重相声表演者，支持相声等传统曲艺的发展；（4）不随波逐流，建立自己的欣赏体系。对相声行业提出如下建议：（1）创新相声表演内容；（2）提高相声表演的文化内涵；（3）延续师徒口传心授的方式进行人才培养；（4）适应当今互联网不断发展的新形势。对相声表演团体提出如下建议：（1）发挥品牌影响力；（2）考虑专场表演的现场环境舒适度；（3）磨练相声演员的表演功底；（4）把给观众带来欢乐作为持续奋斗的目标。

关键词：相声观众；文本挖掘；K-means 聚类；贝叶斯决策模型；云模型

目 录

第一部分 引言..... 1

一、调查背景..... 1

 （一）相声的研究背景..... 1

 （二）调查意义..... 1

二、文献综述..... 2

 （一）相声的基本概念..... 2

 （二）相声发展的政策支持..... 2

 （三）相声传承的必要性..... 3

三、相声发展情况的 PEST 分析..... 3

四、相声市场概述..... 4

 （一）相声市场现状..... 4

 （二）相声市场的 SWOT 分析..... 4

第二部分 调查策划与实施..... 5

一、调查内容..... 5

二、调查对象与范围..... 5

三、调查方式..... 6

 （一）文案调查法..... 6

 （二）深度访谈法..... 6

 （三）问卷调查法..... 6

四、调查实施..... 6

 （一）调研工作流程..... 6

 （二）研究框架..... 7

五、预调查..... 7

 （一）预备访谈..... 7

 （二）委托调查..... 7

 （三）问卷预调查..... 8

六、抽样设计..... 8

 （一）目标总体..... 8

 （二）抽样方法..... 8

 （三）样本容量的计算与分配..... 9

(四) 抽样实施过程.....	9
七、数据处理.....	10
(一) 数据清洗.....	10
(二) 数据检验.....	10
八、数据分析方法.....	11
(一) 描述性分析.....	11
(二) 相关分析.....	12
(三) 多元有序 Logistic 回归模型.....	12
(四) 人群的 K-Means 聚类分析.....	12
(五) 随机森林.....	12
(六) Text-CNN 模型.....	12
(七) 结构方程模型.....	13
(八) 贝叶斯决策模型.....	13
(九) 层次分析法.....	13
(十) 云模型.....	13
第三部分 描述性统计.....	14
一、基本信息.....	14
(一) 性别：女性观众和男性观众比例较均衡.....	14
(二) 年龄：总体集中于青年年龄层次.....	14
(三) 地区观众比例：华北、东北观看相声的人群比重高.....	15
(四) 教育背景：观众受教育水平普遍较高.....	15
(五) 职业：普通职员占比最大.....	16
(六) 月生活开支：中高等月生活费用人群比重大.....	16
二、观众对相声的态度和认知分析.....	17
(一) 观众对相声的了解程度适中.....	17
(二) 绝大多数被调查者对相声倾向喜爱.....	17
(三) 绝大多数观众认为相声演员水平良莠不齐会给相声界带来较大不良影响.....	18
(四) 被调查者对粉丝自发的集资、打榜、控评、举灯牌等行为的态度两极化.....	18
(五) 看好相声等传统曲艺未来发展前景的观众近八成.....	19
(六) 超过七成被调查者认为相声可凭借独特的艺术形式走出国门.....	19
三、观众看相声的途径和观众粘性分析.....	20
(一) 春晚等电视节目以及手机视频软件成为观众观看相声的主要途径.....	20

(二) 观众粘性描述性统计分析：相声的观众黏性较高.....	20
四、观众对相声的偏好.....	21
(一) 多数被调查者偏好历史知识、传奇轶事、工作生活题材.....	21
(二) 相声演员中郭德纲目前最受被调查者的喜欢.....	22
(三) 10~25 分钟的相声表演更受观众喜爱.....	22
五、观众付费情况及付费意愿分析.....	23
(一) 观众付费情况的描述性统计分析.....	23
(二) 年龄段与付费意愿的关系.....	23
(三) 月生活费与愿意支付的最高金额的相关分析.....	24
第四部分 观众对相声的偏好及影响观众喜爱相声的因素分析.....	26
一、观众对相声态度的外部影响因素.....	26
(一) 相声网络评价的数据处理与分词——基于词云统计.....	26
(二) 相声网络评价的情感倾向分析——基于情感词典.....	27
二、影响观众对相声偏好的基本情况因素.....	28
(一) 影响观众对相声了解程度的因素分析——基于多元有序 Logistic 模型.....	28
(二) 相声观众的人物画像——基于 K-Means 聚类.....	30
三、影响观众喜爱相声的因素.....	33
(一) 观众喜爱的相声演员特征主题分类——基于 Text-CNN 模型.....	33
(二) 各类要素对观众喜爱相声程度的影响研究——基于随机森林.....	35
第五部分 观众对相声付费意愿的综合评价分析.....	37
一、影响观众付费观看相声意愿的因素分析——基于结构方程模型.....	37
(一) 结构方程模型介绍.....	37
(二) 潜变量设定与研究假设.....	37
(三) 模型修正与检验.....	38
(四) 模型参数估计.....	39
(五) 模型结果分析.....	40
二、观众选择付费观看相声与否的决策行为分析——基于贝叶斯决策模型.....	41
(一) 贝叶斯模型介绍.....	41
(二) 变量选取.....	42
(三) 贝叶斯计算过程.....	42
(四) 观众对相声付费的预测结果.....	43
三、观众对相声团体选择的影响因素分析——基于 AHP 层次分析法.....	45

（一）AHP 模型介绍.....	45
（二）AHP 评价指标体系的构建.....	46
（三）观众对相声团体选择的影响因素权重设定.....	46
（四）模型结论.....	47
四、观众对付费观看相声的满意度评价分析——基于云模型.....	47
（一）云模型介绍.....	47
（二）满意度评价方法.....	48
（三）模型计算结果与云滴分布情况.....	49
第六部分 结论与建议.....	52
一、结论.....	52
（一）相声观众现状.....	52
（二）观众对相声的偏好情况.....	52
（三）影响观众喜爱相声的因素.....	53
（四）观众对相声付费意愿的综合评价.....	54
二、建议.....	55
（一）相声观众.....	55
（二）相声行业.....	56
（三）相声表演团体.....	56
附录 1 调查问卷.....	59
附录 2 部分访谈记录.....	66
附录 3 代码.....	69
附录 4 委托书.....	77

图目录

图 1	研究框架.....	7
图 2	观众性别分布.....	14
图 3	观众年龄分布.....	14
图 4	各地区观众分布对比.....	15
图 5	观众教育背景分布.....	15
图 6	观众职业分布.....	16
图 7	月生活费用分布.....	16
图 8	观众对相声的了解程度.....	17
图 9	观众对相声的喜爱程度分布.....	17
图 10	造成不良影响的现象.....	18
图 11	被调查者对粉丝行为的态度.....	18
图 12	观众对相声等传统曲艺未来发展前景的态度.....	19
图 13	被调查者认为相声是否可以走出国门的分析图.....	19
图 14	观看相声的途径.....	20
图 15	观看相声的频率.....	20
图 16	观看相声的时长.....	21
图 17	观众偏好的相声题材.....	21
图 18	观众喜欢的相声演员.....	22
图 19	观众喜欢观看相声的时长.....	22
图 20	观众付费情况.....	23
图 21	观众付费意愿.....	23
图 22	年龄与付费意愿分布.....	24
图 23	月生活费与付费意愿分布.....	25
图 24	词云统计图.....	26
图 25	相声网络评价情感分布.....	27
图 26	Text-CNN 模型详细过程原理.....	33
图 27	观众喜爱的相声演员特征关键词提取.....	35
图 28	观众对相声喜爱程度的影响指标随机森林重要性排序图.....	36
图 29	观众相声付费意愿影响因素的结构方程模型初始路径图.....	38
图 30	观众相声付费意愿影响因素的结构方程模型修正后的路径图.....	39
图 31	影响观众相声付费意愿因素的标准化结构方程模型路径系数图.....	40
图 32	影响相声付费意愿程度的因素.....	42
图 33	喜爱程度载率值.....	44
图 34	喜爱程度载率增长速率.....	44
图 35	了解程度载率值.....	44
图 36	了解程度载率增长速率.....	44
图 37	满意程度载率值.....	45
图 38	满意程度载率增长速率.....	45
图 39	评价指标体系图.....	46
图 40	各指标实际值云图.....	50
图 41	云滴分布情况图.....	51

表目录

表 1 PEST 分析..... 3

表 2 SWOT 分析..... 4

表 3 工作流程.....6

表 4 抽样分层与样本人数分配表..... 9

表 5 Cronbach α 系数表..... 11

表 6 KMO 和巴特利特检验表..... 11

表 7 卡方检验(1)..... 24

表 8 卡方检验(2)..... 25

表 9 变量赋值.....28

表 10 回归结果分析.....29

表 11 K-Means 均值聚类分析结果.....31

表 12 观众类型分类表.....32

表 13 Text-CNN 词语主题分类表.....34

表 14 观众对相声喜爱程度的影响指标重要程度排序表..... 35

表 15 潜变量及对应可测变量..... 37

表 16 研究假设.....38

表 17 模型检验与拟合优度结果..... 39

表 18 潜变量之间的标准化路径系数表..... 40

表 19 贝叶斯模型的一步迭代过程..... 43

表 20 一致性检验结果.....46

表 21 各观测指标的权重分配表..... 47

表 22 矩阵结果（期望值） 49

表 23 矩阵结果（实际值） 49

第一部分 引言

一、调查背景

（一）相声的研究背景

德云社迅猛发展掀起全民相声热潮：近些年来，社会上逐渐涌起一股民族主义思潮，传统文化重新被人们关注和重视，对中国的传统文化产生了或多或少的追寻和依恋。在这一大趋势下，德云社高举“传统艺术”大旗，开辟了一条相声的复兴之路。相声艺术要生存、要发展必须回归剧场。复归相声小剧场的表现形式将演员与观众重新带回剧场，实现了剧场中观众与演员的互动，营造了活泼积极的表演氛围，满足了人们当下的文化生活需求和对传统文化回归的渴望。

互联网+产业发展：近年来，始终坚持小剧场表演的“德云社”借助新媒体技术，不仅满足了现场的观众，还满足了屏幕前的大众。哔哩哔哩、优酷、新浪微博……各种网络平台上都可以看到来自现场观众自发录制的视频，供网络用户浏览观看。这种线上线下相呼应的方式深得大众网民们的喜爱与支持。这让“德云社”“让相声回归剧场”、做“真正的相声”的要旨更加深化。并且大范围的互联网传播，让受众的覆盖范围大大超过小剧场实际的观众人数。这样的传播手段，使“郭德纲”“德云社”的品牌形象传播得更加广泛，更加深入人心。

国家对非物质文化遗产的保护：2008年6月7日，国务院关于公布第二批国家级非物质文化遗产名录和第一批国家级非物质文化遗产扩展项目名录正式对外公布。第二批国家级非物质文化遗产名录：“相声”。2018年文化和旅游部确定了第五批国家级非物质文化遗产代表性项目代表性传承人名单，马志明入选。由此可见国家对相声这一非物质文化遗产的大力支持。

（二）调查意义

相声界：相声付费，焕发行业新生机。近年来，相声小剧场开展的如火如荼。以德云社为例，场票常常一票难求。越来越多的年轻人愿意为相声表演付费。一方面，这让相声这一传统艺术拥有更多元的受众群体，为相声发展注入新的生机与活力。另一方面，相声付费市场的火爆反映出相声的市场需求巨大，行业发展潜力不容小觑。相声行业效益的提高，对提高相声从业者的薪酬、鼓励相声界创造出更多更优质的作品具有正向激励作用。另外，付费市场的出现，使相声在观

演生态、人才结构、受众结构、行业价值链重组等方面都发生了前所未有的深刻变化，出现了相声史上少见的飞升与发展。

大众：陶冶情操，为人民群众提供喜闻乐见的多元文化。随着时代发展，人民美好生活需要日益广泛，也对文化生活提出了更高要求。相声是中国传统文化长期积淀的产物，相声的元素也凝聚了中华民族的审美习惯，将中国人所蕴含的精致的审美品格和喜好。相声是中华民族文明传播的载体，它蕴含着中国传统文化的精髓，在演出的过程中，潜移默化的渗透到人民大众中。相声欣欣向荣的发展势态能满足人民群众日益增长的文化需求。

社会：非物质文化遗产继承与弘扬，传统文化大放异彩。中国相声以其浓厚的文化底蕴、独特的审美意境及非凡的艺术价值深深地扎根于普通大众之中，它既是地域文化、民族文化的瑰宝，也是构建世界文化多元化、多样性必不可缺的因素。相声作为非物质文化遗产，它是各族人民世代相承、与群众生活密切相关的各种传统文化表现形式和文化空间。弘扬相声（非物质文化遗产文化）提升其影响力和凝聚力，其实质在于增强中华文化定力，保持民族精神。

二、文献综述

（一）相声的基本概念

王汝冰在《传统相声商业化发展的现状及建议》中写道：最早兴起于清朝的传统相声，形式从最初的街头表演不断发展，近 20 年更是不断突破自身条件限制，在回归传统的同时，利用新媒体的发展，扩大受众范围，以电视相声、商业巡演等多种形式，打破地域限制，走入更多观众的视野^[1]。郭静欣等人在《数字化媒体背景下相声曲艺创新发展研究》中写道：相声艺术开始于明清年间，起源于华北地区，流行在京津冀，普及与全国及海内外。主要采用口头方式进行表演。表演形式有单口相声、对口相声、群口相声等，表演内容大多数都来源于民间传统文化、源于日常生活琐事，是一门深受群众欢迎的传统曲艺表演^[2]。

（二）相声发展的政策支持

2017 年中共中央办公厅、国务院办公厅印发了《关于实施中华优秀传统文化传承发展工程的意见》（以下简称《意见》），并发出通知，要求各地区各部门结合实际认真贯彻落实。

关于基本原则，《意见》指出：要牢牢把握社会主义先进文化前进方向；坚

持以人民为中心的工作导向；坚持创造性转化和创新性发展；坚持交流互鉴、开放包容；坚持统筹协调、形成合力。关于总体目标，《意见》提出：到 2025 年，中华优秀传统文化传承发展体系基本形成，研究阐发、教育普及、保护传承、创新发展、传播交流等方面协同推进并取得重要成果，具有中国特色、中国风格、中国气派的文化产品更加丰富，文化自觉和文化自信显著增强，国家文化软实力的根基更为坚实，中华文化的国际影响力明显提升。

（三）相声传承的必要性

薛丽婷在《相声文化对现代的意义》中讲到：相声艺术已经进入了一个“泛”相声的时代，这种“泛”主要表现在几个方面，一是表演空间的“泛”，现在的相声已经进入了演播大厅甚至是体育馆，而不像以往那样，相声表演只是天桥卖艺或者在舞台上进行。二是艺术内容上的“泛”，现在在的相声内容更加健康，在时代发展的背景下，取材更加广泛，现代相声正朝着幽默化的方向发展。三是传播手段的“泛”，随着科技的发展，即使不能亲临现场观看，我们也可以从各大短视频 app 上看到^[3]。因此，泛相声时代，相声的内容更加丰富且具内涵，相声载体更加多元，这为相声的传承提供了极大的可能性。

三、相声发展情况的 PEST 分析

PEST 分析是指宏观环境的分析，其中 P 表示政治(politics)，E 表示经济(economy)，S 表示社会(society)，T 表示技术(technology)。通过上述四个因素，我们可以对相声行业发展的宏观环境进行初步探究。

表 1 PEST 分析

政 治 环 境	《国家“十二五”时期文化体制改革和发展规划纲要》提出：文化体制改革深入推进，市场体系进一步完善。整合开发三岔河口、大悲院、吕祖堂、文庙、桃花堤等资源，发展具有历史文化特色的项目。打造以相声为代表的传统曲艺集中演出街区。
经 济 环 境	2003 年更名的北京“德云社”，将传统的“父子师徒”关系进行商业经纪公司的签约式改造，充分利用新媒体，根据社会的主流审美，不断进行自身转变，以新的形式对相声演员进行招募和培养，开创了相声商业巡演这一新形式。新形势的出现，实现在大流量盈利的模式下“雅俗共赏”，在受众需求与艺术传承之间实现相互平衡。
社 会 环 境	至今为止，相声艺术一直肩负传播、延展各地文化的作用与使命，大部分传统相声中都有方言、小调等鲜明的地域文化特征，而相声本身就是集各地区历史文化、百姓智慧为一体的艺术产物。但前段时间，一位青年相声演员的“荧光棒相声”成为热议的文化现象，这看似流行与传统文化的碰撞，实则触及了相声发展路线的敏感神经。

技术环境	在大多数人看来，传统的相声节目的传播渠道是剧场内面对面的人际传播以及通过电视、广播这样的大众传播媒体进行的大众传播。然而随着互联网的发展，出现了越来越多的媒体平台，线上与线下传播相辅相成，这不仅给予观众多种听相声的选择途径，也促进了相声行业的发展。
------	--

四、相声市场概述

（一）相声市场现状

目前相声受众群体日益年轻化，商业巡演等相声发展新模式也深受年轻人的喜爱。总体来看，大众对于相声付费持积极态度。但是受制于人民收入水平、对文化产品投入偏好等多方面因素影响，人们整体的付费意愿在百元左右浮动，相声观众具有付费意愿但付费金额不高是当前市场的普遍现状。

（二）相声市场的SWOT分析

SWOT 分析，基于内外部竞争环境和竞争条件下的态势分析，就是将与研究对象密切相关的各种主要内部优势、劣势和外部的机会和威胁等，通过调查列举出来，并依照矩阵形式排列，然后用系统分析的思想，把各种因素相互匹配起来加以分析，从中得出一系列相应的结论，而结论通常带有一定的决策性。

运用这种方法，我们可以对相声市场所处的情景进行全面、系统、准确的研究。

表 2 SWOT 分析

S 优势	W 劣势
<p>1.走进大众文化视野 相声经过商业化的包装和粉丝文化的推动，逐渐走入大众的文化视野。</p> <p>2.相声从业者薪资 近几年，越来越多的年轻人喜爱相声，相声专场表演增多，高价票使相声行业效益迅速提高。</p> <p>3.鼓励相声创新性发展 相声行业经济效益的提高，对相声作品创作具有正向激励作用。</p>	<p>1.相声的适应性不好 互联网时代对相声而言，既是机遇又是挑战，新媒介平台常常伴随着质疑，可能对相声在一定程度造成负面影响。</p> <p>2.过度商业化 过度商业化的包装，会使相声失去“雅俗共赏”的特质。</p> <p>3.市场乱象层出不穷 具有品牌特色的相声剧场经常出现投机者哄抬票价、扰乱市场经济规律的现象。</p>
O 机会	T 威胁

1.文化社区平台 微博、微信、抖音、哔哩哔哩网站等是国内具有代表性的文化社区，并且受众大，有利于相声的传播与发展。 2.粉丝红利 粉丝文化的盛行是当代普遍存在的社会现象，相声与粉丝文化的耦合使相声进入到新的发展阶段。	1.粉丝文化双刃剑 粉丝文化会带来巨大效益，同时粉丝群体也具有复杂性、多层次性和流动性。 2.互联网直播困境 相声的互联网直播需要密集的包袱和笑点，同时要保证迅速和精准化，这对相声演员的基本功底是一种极大的考验。
---	---

第二部分 调查策划与实施

一、调查内容

本次调查主要包括五个部分：

第一部分为被调查者的基本信息。主要包括相声观众的性别、年龄、职业、所在地区、学历、月收入/生活费等情况以及对相声的喜爱程度、观看相声的频率等情况。

第二部分为观众对待相声的认知和偏好情况。主要包括观看相声的途径、对相声的了解，关心和满意程度以及喜欢相声表演的内容、题材类型和演员等。

第三部分为观众对待相声的行为特征调查。主要包括影响相声观看的因素的重要程度、能够接受相声演出的最高价格范围、喜欢的相声表演的时长、认为相声在发展过程中存在的问题等情况。

第四部分为已付费观看相声的观众态度调查。主要包括喜欢的相声团体及各因素的影响程度、付费观看相声表演前后各项指标的期望值和实际值、各因素对不再付费观看相声的影响程度等情况。

第五部分为观众对相声发展的前景展望。主要包括观众认为应该如何发展相声、对粉丝集资，打榜，举灯牌等的态度、对相声是否可以走向国门的看法以及对相声等传统曲艺未来发展的态度。

二、调查对象与范围

调查对象：国内相声观众。

调查范围：线上在全国范围内七个地区（东北、华东、华北、华中、华南、西南、西北）发放调查问卷。

三、调查方式

（一）文案调查法

通过中国知网 CKNI 数据库、万方数据库、谷歌学术等学术网站检索目前有关相声的文献，对其中的文献综述、实证分析、结论与建议等部分进行详细阅读与研究；通过搜索引擎如百度、搜狐等收集统计数据、相关政策、调研报告等资料，提取有助于本次调查的信息。

（二）深度访谈法

实地走访黑龙江省曲艺团，通过与相声演员进行无结构的、直接的、个人的访问，掌握相声市场发展的现状，并针对目前相声界存在“粉丝经济”的社会现象、当代观众对相声的选择情况以及相声的传承方式等进行访谈。

（三）问卷调查法

调查问卷采用线上调查的形式，委托问卷星平台进行发放，调查对象为国内广大相声观众，其中有一部分问题专门针对已付费观看过相声的观众。

四、调查实施

（一）调研工作流程

表 3 工作流程

项目	开始时间	截止时间
确定调查内容和调查目的	12 月 16 日	1 月 10 日
联系委托方	1 月 16 日	1 月 19 日
设计调查方案与调查问卷	1 月 20 日	2 月 26 日
进行预调查	2 月 28 日	3 月 3 日
问卷分析与调整	3 月 3 日	3 月 6 日
正式调查	3 月 7 日	3 月 16 日
质量控制	3 月 7 日	3 月 16 日
数据处理与分析	3 月 17 日	4 月 28 日
撰写与修改比赛报告	4 月 29 日	6 月 19 日

（二）研究框架

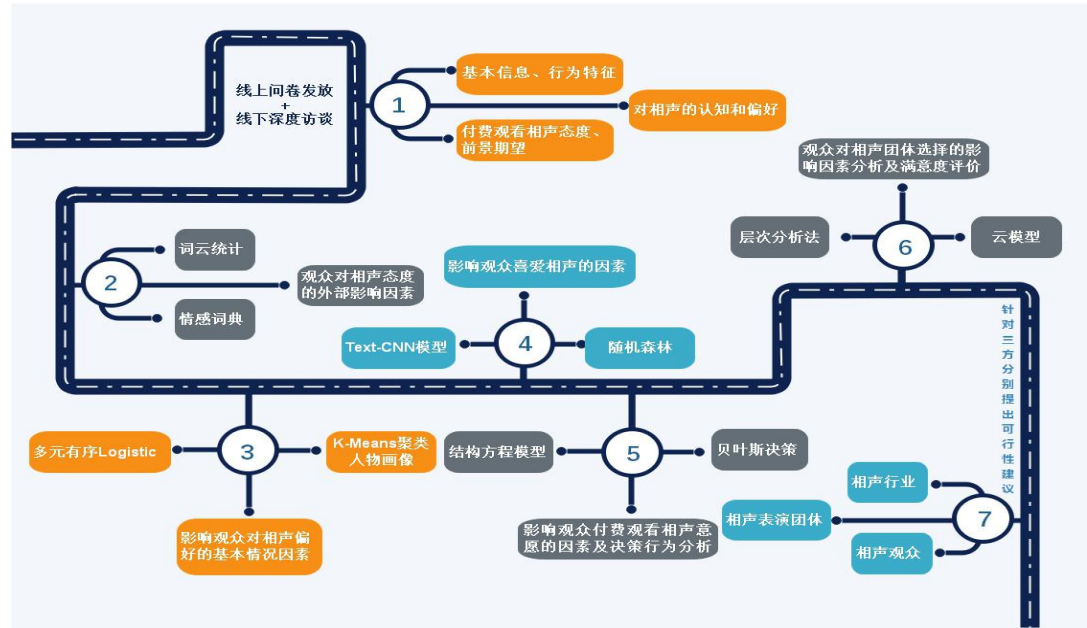


图 1 研究框架

五、预调查

（一）预备访谈

在进行问卷设计之前，为详细了解相声的发展现状，团队成员对黑龙江省曲艺团的相声演员李大宇先生进行了访谈。李大宇先生是知名的相声演员，十多年来，李大宇先生活跃在各大舞台上，具有《超级英雄》、《学评书》等作品，给观众带来了许多欢乐。同时热心于相声传承事业，甚至创立哈尔滨市大学生相声联盟，致力于在高校范围内促进相声的发展。在预备访谈的过程中，李大宇先生对目前国内相声的发展情况进行了详细的介绍。基于此，团队决定将调查方向确定为相声观众偏好分析与付费意愿调查，从而进行了问卷设计。

在进行问卷发放前，团队再一次对李大宇先生进行了预备访谈，并将问卷设计进行了展示与说明。发现部分选项的内容脱离实际市场情形，且为观众无法回答的问题。因此，团队在此基础上，对问卷内容进行了修改并加以完善。

（二）委托调查

在团队的调研活动开始之前，衡阳市雁峰区三缘文化艺术中心的负责人孙庆宪先生对团队的本次调研活动产生了浓厚的兴趣，为了优化艺术中心在相声表演方面的内容设置和题材选择等，丰富表演形式，孙庆宪先生委托团队对相声市场进行系统的调查，并与团队签订了市场调查项目委托书。

（三）问卷预调查

为了检验问卷的有效性，在正式调查之前，团队在衡阳市内进行“拦截式”实地问卷发放，回收了 50 份问卷。通过与被调查者交流以及数据检验，对问卷进行了如下改进：

- 1.将第 6 题“您的月收入为”改成“您的月收入/生活费为”并调整金额区间。原因：被调查者有一部分为无收入的学生群体，如果只问月收入，该群体大多数会选择最低金额区间，不利于后续各项分析。
- 2.增加了一道关于观众对相声表演联系的内容偏好的题目。原因：在原来的题目中，关于观众对相声的偏好主要体现于对题材的选择，但是题材和内容是两个不同的概念，可以增加调查的全面性。
- 3.删除对“相声类型”、“相声著作时代”的偏好的题目。原因：在这两道题中，调查对象有超过八成选择“对口相声”和“当代相声”，由于选择过于集中，经慎重讨论，将这两道题删除。

针对修改后的问卷，团队通过相同步骤进行了第二次预调查，发现问卷有效性有所提高。

六、抽样设计

（一）目标总体

目标总体为全国范围内的相声观众，既包括已付费观看过相声的观众，也包括未付费观看过相声的观众。

（二）抽样方法

问卷采用配额抽样与方便抽样相结合的方法，由于团队旨在调查全国范围内所有的观众对相声的偏好及付费意愿等，因此将全国按照行政区划分为七大区，

并各自作为一层，再根据每层中所包含的常住人口比例确定各地区的问卷发放数量。

（三）样本容量的计算与分配

本次调查的样本容量可根据以下公式结合起来进行计算：

$$n_1 = \frac{Z_{\alpha/2}^2 \pi (1 - \pi)}{E^2} \quad \text{公式（1）}$$

$$n_2 = Deff \times n_1 \quad \text{公式（2）}$$

$$n = \frac{n_2}{r} \quad \text{公式（3）}$$

（其中 n_1 为样本量， $Z_{\alpha/2}$ 为当置信度为 $1-\alpha$ 时 Z 统计量的值， E 为极限误差， π 为总体比例， $Deff$ 为设计效应 r 为预计回答率）

步骤①：在 95% 的置信度（ $\alpha=0.05$ ）下， $Z_{\alpha/2}^2$ 取固定值 1.96，设定极限误差 E 为 4%，总体比例 π 取 0.5 的前提下，计算得样本容量 $n_1 = 601$ 。

步骤②：拟定设计效应 $Deff = 2$ ，计算得样本容量 $n_2 = 1202$ 。

步骤③：拟定预计回答率 $r = 80\%$ ，计算得最终需要调查的样本容量 $n = 1503$ 。

为保证有效调查问卷的数量与调查数据的可靠性，本次调查将样本容量扩大到 1600 份。

（四）抽样实施过程

1. 样本量确定

将全国按照行政区划分为七大区，分别是东北地区、华东地区、华北地区、华中地区、华南地区、西南地区、西北地区，因此根据不同地区常住人口比例进行样本容量分配，得到表 4，并在各层内进行方便抽样。

表 4 抽样分层与样本人数分配表

地区	常住人口数	总体层权	发放问卷数量（份）
东北地区	10793 万人	7.69%	123
华东地区	41423 万人	29.50%	472

华北地区	17575 万人	12.52%	201
华中地区	22485 万人	16.02%	256
华南地区	17426 万人	12.42%	198
西南地区	20331 万人	14.48%	232
西北地区	10349 万人	7.37%	118

2. 问卷发放与质量控制

委托问卷星平台进行有偿问卷发放，同时严格按照表 4 的抽样比例对全国范围内 7 个地区的相声观众进行发放。此外，为保证问卷的有效性与稳定性，第一题设置为【您是否观看过相声】单选题，只有选项中选择“是”的被调查者才可继续作答。此外，我们发放问卷的时间集中在每天的 9 点、13 点、18 点三个时间点，这些时间一般是受访者的空闲时间。这时向受访者推送问卷的答题时间更充裕、回答效果更好。调查过程中，在问卷发放平台后台进行严格的质量控制，设置“同一设备仅能填写一次”的要求，严格筛掉无效问卷，从而提高调查问卷的有效性和可靠性。

本次调查实际发放线上问卷 1600 份，其中有效问卷 1286 份，问卷有效率达 80.375%。

七、数据处理

（一）数据清洗

对所收集的问卷进行数据录入后，查找并修改超出范围、有极端值、逻辑上不一致的数据，进行数据净化；另外，对于缺失值的处理，若出现少量缺失值，则可以使用变量取值的均值来替代，而对于出现大部分数据缺失的问卷，则将整个问卷作废。

（二）数据检验

1. 信度检验

信度分析在于研究数据是否真实可靠，即研究样本是否真实回答了问题。通常情况下，克隆巴赫系数高于 0.7 即可。利用 SPSS 统计软件计算问卷中四个量表题的 Cronbach α 系数，结果如表 5 所示：

表 5 Cronbach α 系数表

	总体	对影响相声观看的因素的重要程度进行打分 (Q17)	各因素对不再付费观看相声的影响程度 (Q21)	各因素对喜爱相声团体的影响程度 (Q23)	对付费观看相声表演前后各项指标的期望值和实际值打分 (Q25)
克隆巴赫 Alpha	0.947	0.839	0.771	0.747	0.926
项数	48	14	7	7	20

由表 5 可知，问卷总体的 Cronbach α 系数大于 0.9，其余各部分也都在 0.7 以上，因此问卷具有较高的可信度。

2.效度检验

效度分析用于分析研究题项是否可以有效地表达对应变量的概念信息，即分析量表题项设计是否合理。借助 SPSS 统计分析软件，对问卷结构效度进行检测，结果如表 6 所示：

表 6 KMO 和巴特利特检验表

		总体	对影响相声观看的因素的重要程度进行打分 (Q17)	各因素对不再付费观看相声的影响程度 (Q21)	各因素对喜爱相声团体的影响程度 (Q23)	对付费观看相声表演前后各项指标的期望值和实际值打分 (Q25)
KMO		0.946	0.870	0.821	0.800	0.947
巴特利特球形度检验	近似卡方	14000.314	2759.773	1030.790	971.192	5846.834
	自由度	1128	91	21	21	190
	显著性	.000	.000	.000	.000	.000

数据问卷总体及各量表的 KMO 值均大于 0.8，另外，Bartlett 检验的 P 值均为 0.000，远小于 0.05，说明问卷整体效度较高。

八、数据分析方法

(一) 描述性分析

描述性分析是对调查所得的大量数据资料进行初步的整理和归纳，以找出这些资料的内在规律——集中趋势和分散趋势。主要借助各种数据所表示的统计量，如均数、百分比等，进行单因素分析。

（二）相关分析

相关分析是研究现象之间是否存在某种依存关系，并对具体有依存关系的现象探讨其相关方向以及相关程度，是研究随机变量之间的相关关系的一种统计方法。

（三）多元有序Logistic回归模型

该模型是逻辑回归的一种特殊情况，其中因变量（ y ）有 n 种不同的选择，并且不同选择按照一定程度递进，函数表达式为：

$$\ln \left[\frac{p(y \leq n)}{1 - p(y \leq n)} \right] = \alpha_n + \sum_{m=1}^k \beta_m x_m$$

y 为因变量， n 表示因变量的 n 个层级；

X_m 表示自变量的第 m 个变量；

α_n 表示截距项；

β_m 表示回归系数。

（四）人群的K-Means聚类分析

K-means 算法是典型的基于距离的聚类算法，采用距离作为相似性的评价指标，即认为两个对象的距离越近，其相似度就越大，该算法认为簇是由距离靠近的对象组成的，因此把得到紧凑且独立的簇作为最终目标。

（五）随机森林

随机森林是一种集成算法（Ensemble Learning），它属于 Bagging 类型，通过组合多个弱分类器，最终结果通过投票或取均值，使得整体模型的结果具有较高的精确度和泛化性能。随机森林采用的 CART 决策树就是基于基尼系数选择特征的。基尼系数的选择的标准就是每个子节点达到最高的纯度，即落在子节点中的所有观察都属于同一个分类，此时基尼系数最小，纯度最高，不确定度最小。

（六）Text-CNN模型

Text-CNN 模型是由 Yoon Kim 在论文(2014 EMNLP) Convolutional Neural Networks for Sentence Classification 中提出的，其主要思想是将不同长度的短文作为矩阵输入，使用多个不同规模的筛选器去提取句子中的关键信息，并用于最终的分类。

（七）结构方程模型

结构方程模型整合了因子分析、路径分析和多重线性回归分析的思想和方法，包括测量模型与结构模型。测量模型求出观察指标与潜变量之间的关系，结构模型求出潜在变量与潜在变量之间的关系，并通过各种拟合指数检验它是否吻合数据。

（八）贝叶斯决策模型

在不完全情报下，对部分未知的状态用主观概率估计，然后用贝叶斯公式对发生概率进行修正，最后再利用期望值和修正概率做出最优决策。贝叶斯决策模型是一种基本的统计决策模型，对于描述复杂市场环境下的决策行为具有重要意义。贝叶斯决策属于风险型决策，决策者虽不能控制客观因素的变化，但却掌握其变化的可能状况及各状况的分布概率，并利用期望值即未来可能出现的平均状况作为决策准则。

（九）层次分析法

层次分析法是一种解决多目标的复杂问题的定性与定量相结合的决策分析方法。该方法是将一个复杂的多目标决策问题作为一个系统，将目标分解为多个目标或准则，进而分解为多指标（或准则、约束）的若干层次，通过定性指标模糊量化方法算出层次单排序（权数）和总排序，以作为目标（多指标）、多方案优化决策的系统方法。

（十）云模型

云模型属于不确定性人工智能范畴，主要用于定性与定量之间的相互转换，自然界中的不确定性从属性角度来说主要有随机性和模糊性。“云”或者“云滴”是云模型的基本单元，“云”是指其在论域上的一个分部，可以用联合概率的形式 (x, μ) 来类比。

第三部分 描述性统计

一、基本信息

（一）性别：女性观众和男性观众比例较均衡

如图 2 所示，在所有调查样本中，女性和男性被调查者占总体比例分别为 57.66%、42.34%，说明女性占比稍微多于男性，观众的性别比例较均衡。

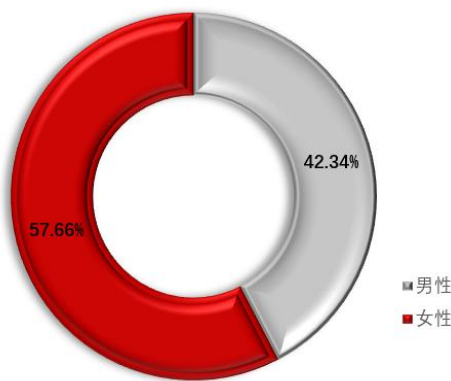


图 2 观众性别分布

（二）年龄：总体集中于青年年龄层次

如图 3 所示，在本次调查回收的有效问卷中，21-30 岁的被调查者占比最大，达到 45.84%；其次是 31-40 岁的被调查者，占比 26.13%；20 岁及以下和 51 岁及以上的被调查者最少，分别占比 9.49%、6.13%。由此可见被调查的观众年龄集中在 21~40 岁，青年占主导地位。

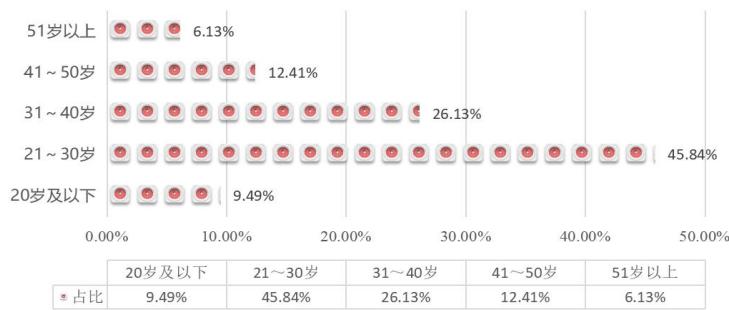


图 3 观众年龄分布

（三）地区观众比例：华北、东北观看相声的人群比重高

如图 4 所示，由于本次调查问卷发放采取按照常住人口比例分配的方式，因此选取对比法来进行分析。可以看出：东北地区和华北地区的问卷回收比例明显大于问卷发放比例，说明这两个地区观看相声的人群比重高于其他地区。

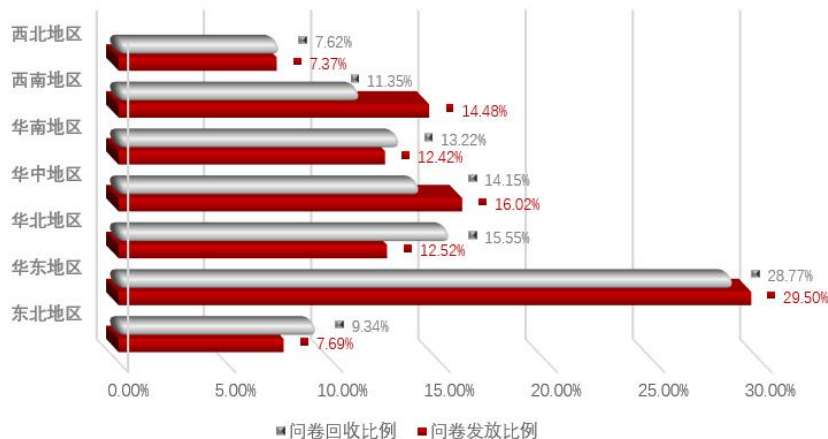


图 4 各地区观众分布对比

（四）教育背景：观众受教育水平普遍较高

如图 5 所示，在被调查者的教育背景调查中，专科和大学本科学历占比最高，达到 70.51%；其次是高中和中专学历，占比为 14.74%；研究生及以上学历占比 9.78%。从总体上看，高中和中专及以上学历被调查者占比 95.03%，说明观众受教育水平普遍较高。



图 5 观众教育背景分布

（五）职业：普通职员占比最大

如图 6 所示，被调查者中普通职员人员最多，占总比接近三成；企业管理者次之，占比 19%；政府/机关干部/公务员以及在校学生占比例基本相同，且相对较多。可见相声的观众职业分布广泛，其中普通职员的比例最高，是最主要的观众群体。

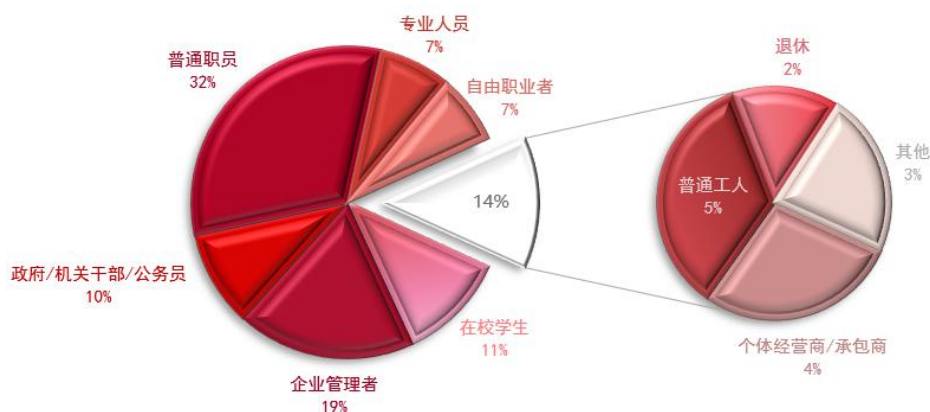


图 6 观众职业分布

（六）月生活开支：中高等月生活费用人群比重大

如图 7 所示，月生活开支为 3000-6000 元的被调查者占比最高，比例为 31.97%；其次是月收入在 6000-9000 元以下的被调查者，占比为 18.56%；月收入在 1500-3000 元的被调查者占比为 17.52%，说明被调查的观众中中高等月生活费用人群比重大。

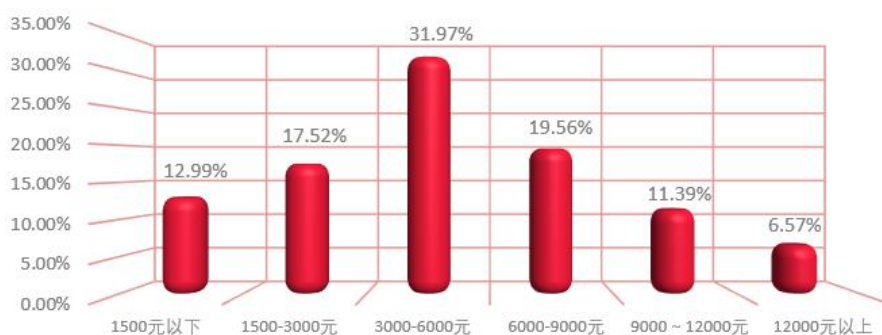


图 7 月生活费用分布

二、观众对相声的态度和认知分析

（一）观众对相声的了解程度适中

如图 8 所示，9.34%的观众表示非常了解相声，38.39% 的观众表示比较了解相声，31.53%的观众表示一般，14.74%的观众表示不太了解，5.99%的观众表示完全不了解。由此可知，观众对相声的了解程度主要集中在比较了解和一般，因此，总体上观众对相声的了解程度适中。

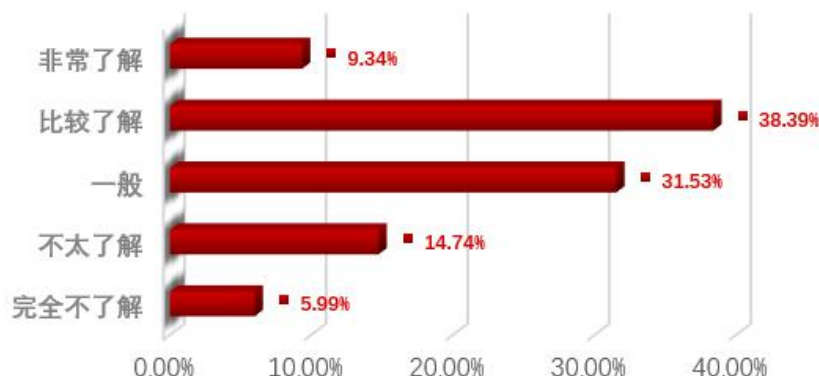


图 8 观众对相声的了解程度

（二）绝大多数被调查者对相声倾向喜爱

如图 9 所示，在收回的有效问卷中，超过七成的被调查者非常喜欢和挺喜欢相声，24.53%的被调查者表示喜欢程度一般，只有 2.84%觉得不喜欢。由此可见，观众对相声的喜爱程度大多倾向积极。

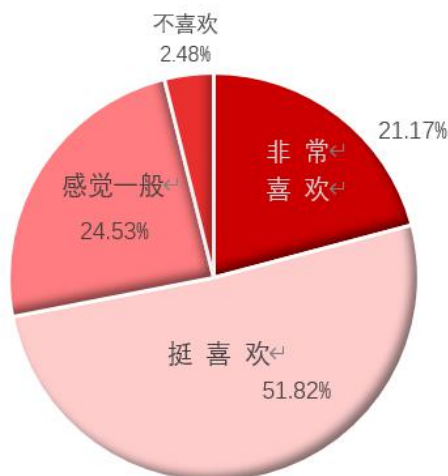


图 9 观众对相声的喜爱程度分布

（三）绝大多数观众认为相声演员水平良莠不齐会给相声界带来较大不良影响

如图 10 所示，关于哪些现象会给相声界带来较大的不良影响的问题上，70.8%的观众选择了“相声演员水平良莠不齐”，54.16%的观众选择“过于依赖偶像经济”，39.56%选择了“主流媒体的批判”，38.69%选择“界内过于浮躁”。由此可见，观众认为相声演员水平良莠不齐带来不良影响的可能性最大，相声行业应该注重演员水平的培养。

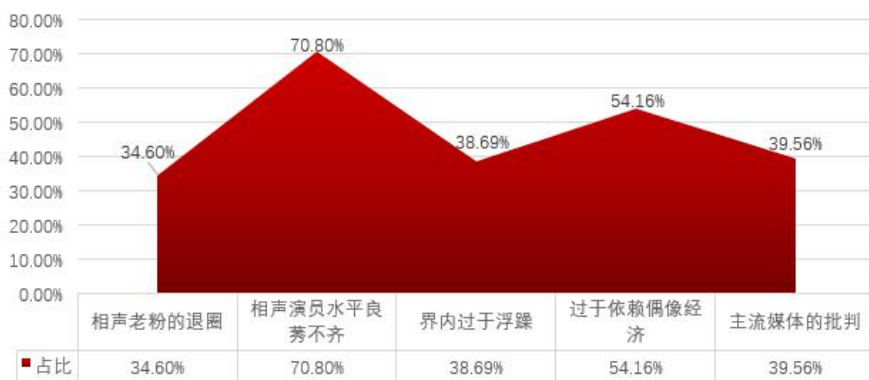


图 10 造成不良影响的现象

（四）被调查者对粉丝自发的集资、打榜、控评、举灯牌等行为的态度两极化

如图 11 所示，调查结果显示：赞同粉丝自发的集资、打榜、控评、举灯牌等行为，认为如此能让更多人关注到相声的被调查者占比 41.31%；态度中立，认为该行为不会造成影响的被调查者占比 13.14%；不赞同该行为，认为传统曲艺本应是纯粹的占比 45.55%。显而易见，被调查者的态度两极化，粉丝自发的集资、打榜、控评、举灯牌等行为影响较大。

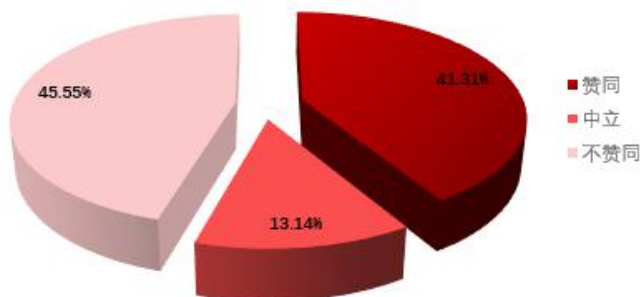


图 11 被调查者对粉丝行为的态度

（五）看好相声等传统曲艺未来发展前景的观众近八成

如图 12 所示,在对相声等传统曲艺未来发展前景持何态度的问题上,75.33%的观众持积极态度,并且认为将会有越来越多的人愿意了解传统曲艺文化;23.5%的观众持中立态度,未深度考虑过,觉得走一步看一步;1.17%的观众持消极态度,认为“相声是虚假繁荣”。说明大部分观众对于相声等传统曲艺的前景偏向积极。

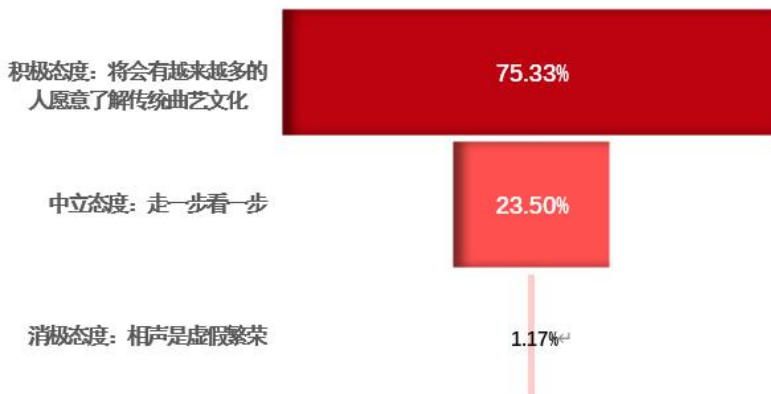


图 12 观众对相声等传统曲艺未来发展前景的态度

（六）超过七成被调查者认为相声可凭借独特的艺术形式走向国门

如图 13 所示,关于相声是否可以走向国门调查中,超过七成的被调查者认为相声可以凭借独特的艺术形式和幽默的内涵走向国门,近二成被调查者认为相声不能够走向国门是由于语言文化的差异导致外国观众可能难以理解其中的笑点,剩余的被调查者则表示对这个问题不太清楚,没有思考过。由此可知,被调查者中大多数支持相声走向国门。

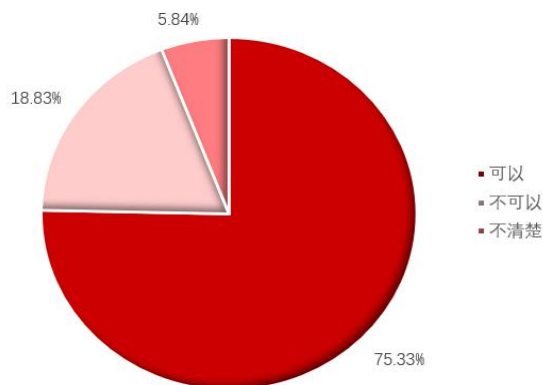


图 13 被调查者认为相声是否可以走向国门的分析图

三、观众看相声的途径和观众粘性分析

（一）春晚等电视节目以及手机视频软件成为观众观看相声的主要途径

如图 14 所示，在被调查者主要通过什么途径观看相声的问题上，选择春晚等电视节目的被调查者数量最多，比例为 71.82%，选择手机视频软件的被调查者比例为 69.64%。因此，观众主要通过电视和手机视频软件来观看相声。

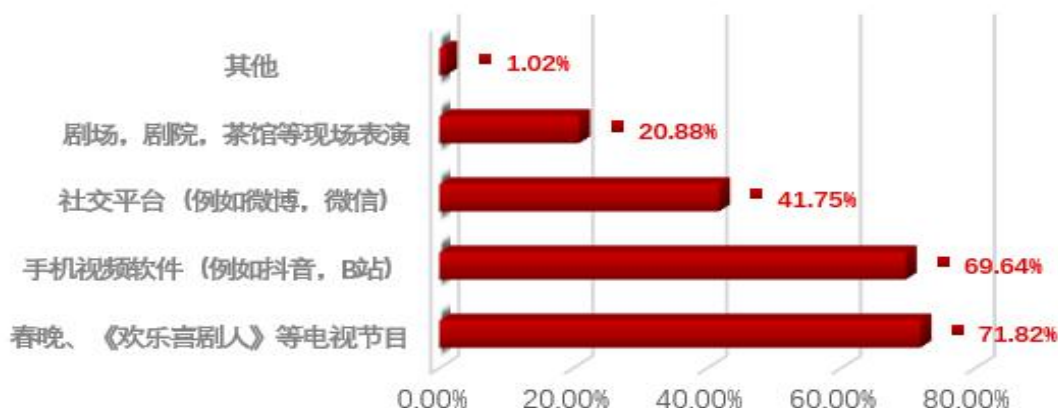


图 14 观看相声的途径

（二）观众粘性描述性统计分析：相声的观众黏性较高

如图 15 所示，每天都看的被调查者占 9.2%，每周至少观看一次的被调查者占比 44.96%，每月至少观看一次的被调查者占 24.96%，每季度至少观看一次的被调查者占 11.82%，几乎不看的占 9.05%。

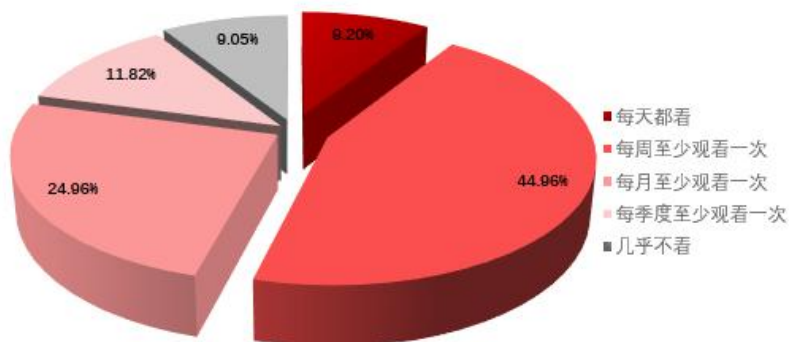


图 15 观看相声的频率

如图 16 所示，在春晚时会看多久的相声问题上，39.71%的观众表示愿意看完所有相声的部分，39.85%的观众会看绝大部分，14.89%的观众则选择看一半，仅 3.5%和 2.04%的观众只看个开头或完全不看。

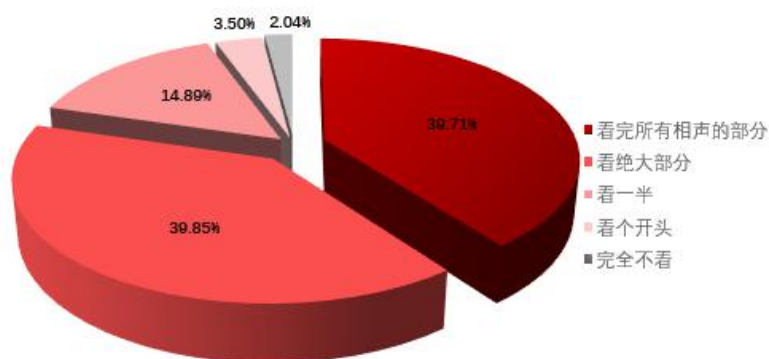


图 16 观看相声的时长

因此，大部分观众高频率、长时段观看相声，观众黏性较高。

四、观众对相声的偏好

（一）多数被调查者偏好历史知识、传奇轶事、工作生活题材

如图 17 所示，在感兴趣的题材内容中，选择“历史知识”、“传奇轶事”、“工作生活”题材的被调查者比例分别为 56.35%、54.45%、54.31%，均超过五成；选择“国家发展”、“校园趣事”的被调查者比例均近半数；选择“其他”的被调查者最少，占比 0.44%，而偏好“江湖野史”和“长篇故事连载”题材的被调查者比例为 34.01%和 10.80%。由此可见，偏好历史知识、传奇轶事和工作生活题材的被调查者占比较高。

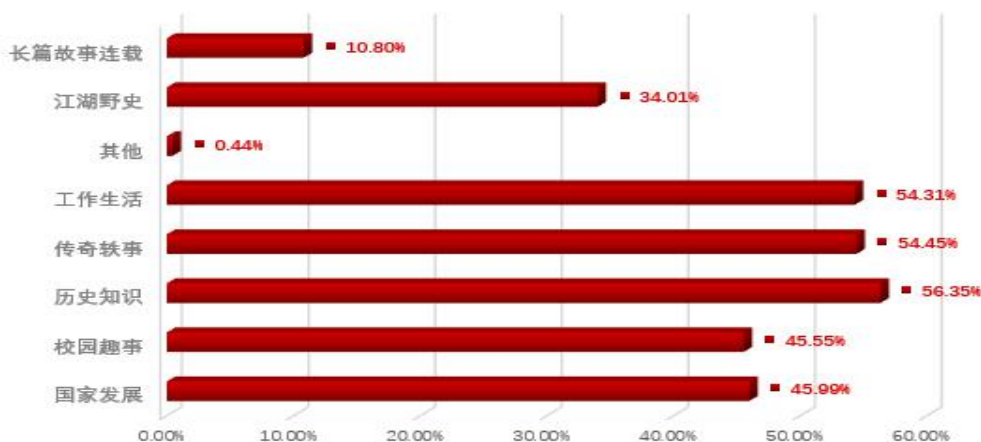


图 17 观众偏好的相声题材

（二）相声演员中郭德纲目前最受被调查者的喜欢

如图 18 所示，关于喜欢的相声演员，67.88%的被调查者选择了郭德纲，其次 63.07%选择了岳云鹏，选择侯宝林和冯巩的被调查者均超过四成，说明相声演员中郭德纲目前最受被调查者的喜欢。

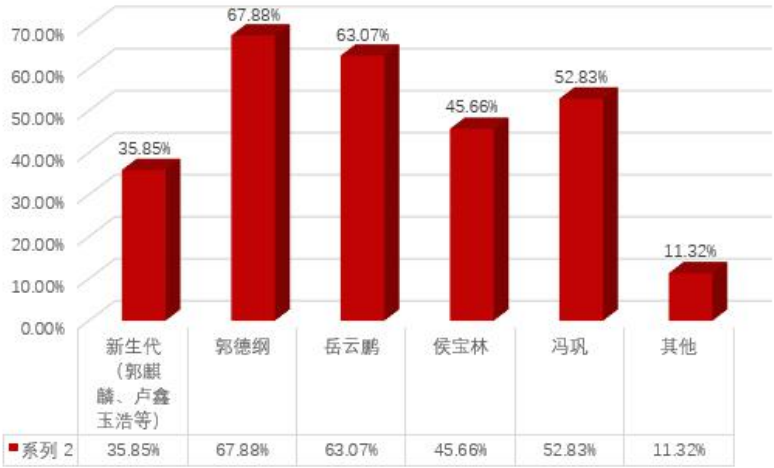


图 18 观众喜欢的相声演员

（三）10~25分钟的相声表演更受观众喜爱

如图 19 所示，喜欢观看一段相声时长为 10 分钟以下的观众占比 11.53%，10~25 分钟的观众占比 56.06%，25~40 分钟的占比 28.47%，40 分钟以上的占比最少为 3.94%。可见 10~25 分钟的相声更受观众喜爱。

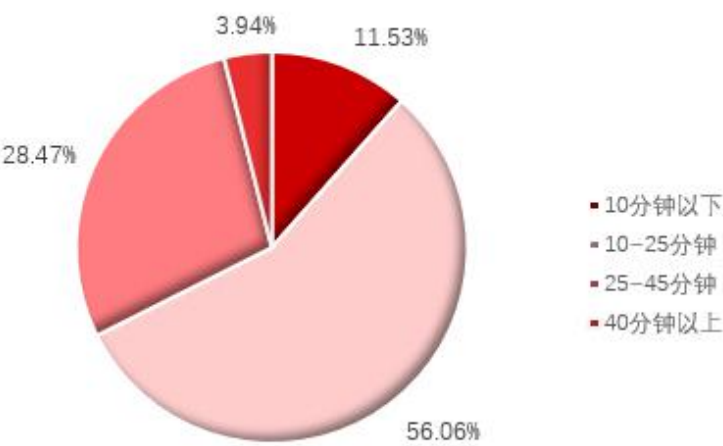


图 19 观众喜欢观看相声的时长

五、观众付费情况及付费意愿分析

（一）观众付费情况的描述性统计分析

从整体上看，付费观看过相声表演的观众数量未过半，但根据观众愿意为相声表演支付的金额，如图 21 可见当相声表演门票低于 300 元时，绝大多数观众愿意付费观看，因此观众付费意愿较强烈，存在大量潜在消费者。

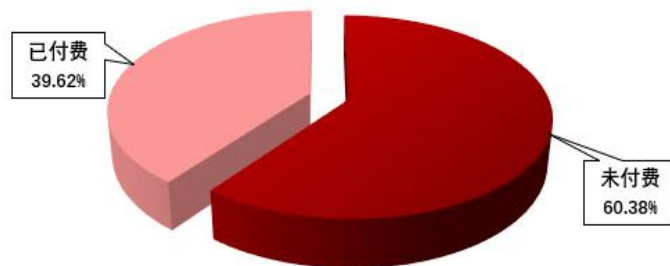


图 20 观众付费情况

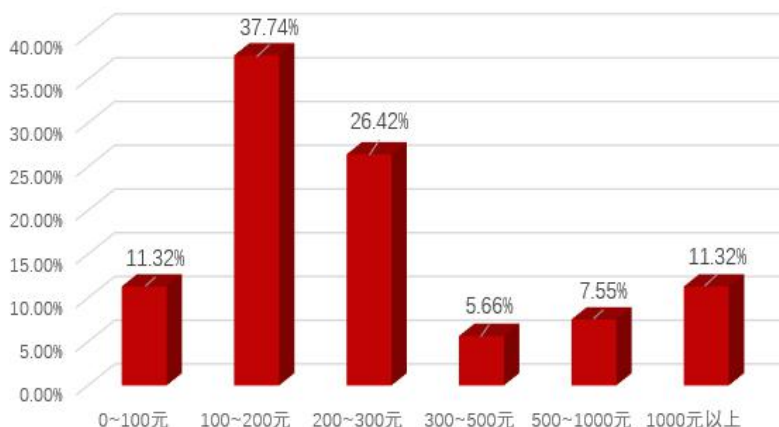


图 21 观众付费意愿

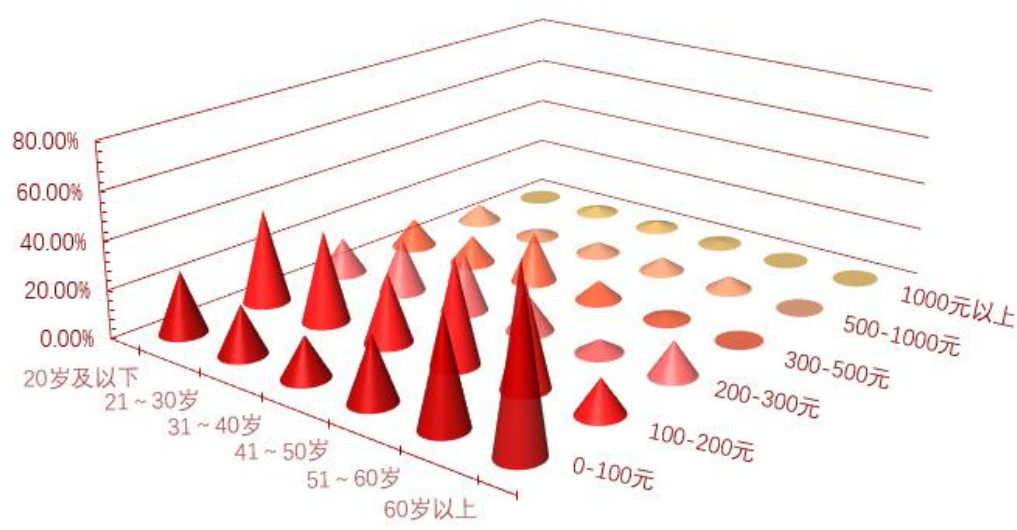
（二）年龄段与付费意愿的关系

如表 7，年龄段与付费意愿之间相关性显著。愿意付高价观看相声的年龄集中在 21—50 岁之间，该年龄段具有一定消费能力；所调查 60 岁以上的观众中无人愿意花费超过 300 元的价格现场观看相声，他们更倾向于低价的消费或者通过其他免费方式看相声。总体来看，大众普遍能接受的价格区间在 0—300 元。

表 7 卡方检验(1)

	值	自由度	渐进显著性（双侧）
皮尔逊卡方	54.538 ^a	25	.001
似然比	56.690	25	.000
线性关联	4.910	1	.027
有效个案数	690		

a.14 个单元格 (38.9%) 的期望计数小于 5。最小期望计数为 0.15。



X\Y	0-100 元	100-200 元	200-300 元	300-500 元	500-1000 元	1000 元以上
20 岁及以下	25.76%	39.39%	15.15%	12.12%	7.58%	0.00%
21~30 岁	20.57%	37.97%	23.10%	12.03%	3.48%	2.85%
31~40 岁	17.22%	28.89%	25%	21.11%	5.00%	2.78%
41~50 岁	27.06%	43.53%	14.12%	8.24%	5.88%	1.18%
51~60 岁	44.44%	41.67%	5.56%	2.78%	5.56%	0.00%
60 岁以上	71.43%	14.29%	14.29%	0.00%	0.00%	0.00%

图 22 年龄与付费意愿分布

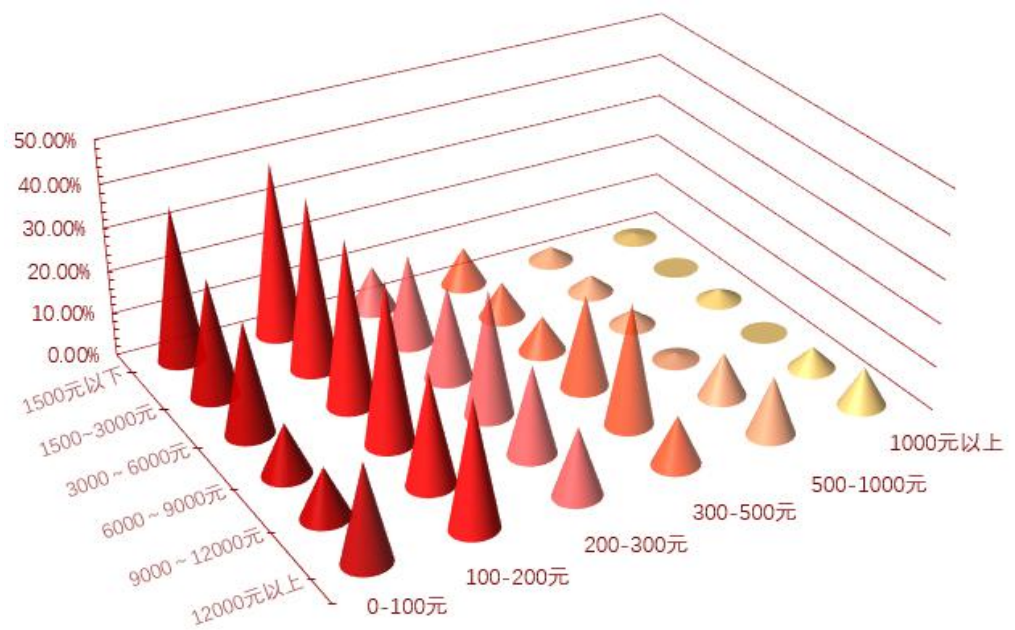
（三）月生活费与愿意支付的最高金额的相关分析

根据调查结果，月生活费与愿意支付的最高金额相关性显著。如图 23 可以看出随着月生活费的提高，观众愿意支付的最高金额越大，当月生活费以 9000 元为界时最为明显，超过 9000 元观众愿意支付 500 元以上的比例迅速增多，但所有月生活费区间愿意支付的最高金额集中在 0—300 元。

表 8 卡方检验(2)

	值	自由度	渐进显著性（双侧）
皮尔逊卡方	96.616 ^a	25	.000
似然比	94.445	25	.000
线性关联	42.931	1	.000
有效个案数	690		

a. 9 个单元格 (25.0%) 的期望计数小于 5。最小期望计数为 0.98。



X\Y	0-100元	100-200元	200-300元	300-500元	500-1000元	1000元以上
1500元以下	35.96%	40.45%	10.11%	8.99%	3.37%	1.12%
1500~3000元	27.27%	39.67%	20.66%	8.26%	4.13%	0.00%
3000~6000元	25.91%	38.18%	21.36%	8.18%	3.64%	2.73%
6000~9000元	11.76%	36.76%	28.68%	21.32%	1.47%	0.00%
9000~12000元	11.39%	25.32%	20.25%	27.85%	10.13%	5.06%
12000元以上	22.22%	28.89%	15.56%	11.11%	13.33%	8.89%

图 23 月生活费与付费意愿分布

两名相声演员比较受关注。相声热门地点天津、北京等都为北方地区，侧面说明相声演出在北方地区更受欢迎。相声观看渠道包括春晚与演出等，说明大众对于这两种观看形式讨论较多。

（二）相声网络评价的情感倾向分析——基于情感词典

我们对采集到的文本进行分词处理后，通过情感词典对其进行情感倾向分析，最终将词语分成正面情感与负面情感两类，并从中提取了七种情绪。为了得到更加完整的情感词典，我们收集了包括 BosonNLP、知网在内的若干情感词典，并且对它们进行了整合去重，同时对部分词语进行了调整，使最后的结果尽可能准确。最终结果如下图所示：

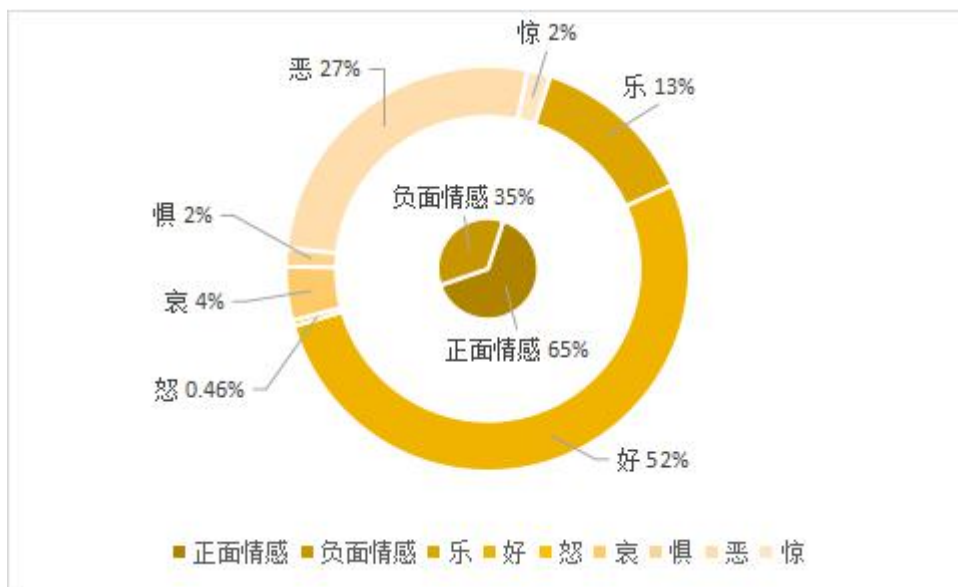


图 25 相声网络评价情感分布

从整体来看，正面情感占 65%，负面情感占 35%，说明相声的网络评价大多数都为正面的，但仍存在部分负面评价。其中，“哀”、“恶”情绪在负面情感中占比较大。以上负面情绪产生的原因可能为目前相声表演内容中粗俗的字眼使得一些观众难以接受，相声表演内容没有做到与时俱进等，我们应重视这些负面情绪产生的原因。

二、影响观众对相声偏好的基本情况因素

（一）影响观众对相声了解程度的因素分析——基于多元有序 Logistic 模型

1.多元有序 Logistic 模型介绍

本文选用多元有序 Logistic 模型来对影响受访者关于相声的了解程度的因素进行回归分析。其中受访者对相声的了解程度（y）有五种不同的选择，并且不同选择按照一定程度递进，函数表达式为：

$$\ln\left[\frac{p(y \leq n)}{1 - p(y \leq n)}\right] = \alpha_n + \sum_{m=1}^k \beta_m x_m$$
公式（4）

公式（4）中，n 表示了解程度的 5 个层级，y 为受访者对相声的了解程度，X_m 表示与受访者自身情况有关的第 m 个变量，α_n 表示截距项，β_m 表示回归系数。

2.模型自变量赋值

由于模型因变量为受访者对相声的了解程度，因此根据里克特量表设置 5 个依次递进的等级（完全不了解，不太了解，一般了解，比较了解，非常了解），分别分析性别、年龄、文化程度、月收入/生活费、对相声的喜爱程度、观看相声的频率 6 个影响因素与对相声的了解程度的关系，下列为 6 个主要的影响因素，变量赋值详见表 9。

表 9 变量赋值

变量名称	变量定义及赋值	变量名称	变量定义及赋值
性别		1500~3000 元	其他=0；1500~3000 元=1
男	其他=0；男=1	3000~6000 元	其他=0；3000~6000 元=1
女	其他=0；女=1	6000~9000 元	其他=0；6000~9000 元=1
年龄		9000~12000 元	其他=0；9000~12000 元=1
20 岁以下	其他=0；20 岁以下=1	12000 元以上	其他=0；12000 元以上=1
21~30 岁	其他=0；21~30 岁=1	喜爱程度	
31~40 岁	其他=0；31~40 岁=1	非常喜欢	其他=0；非常喜欢=1
41~50 岁	其他=0；41~50 岁=1	挺喜欢的	其他=0；挺喜欢的=1
51 岁以上	其他=0；51 岁以上=1	一般	其他=0；一般=1
文化程度		不喜欢	其他=0；不喜欢=1
小学及以下	其他=0；小学及以下=1	观看频率	
初中	其他=0；初中=1	每天都看	其他=0；每天都看=1

高中/中专	其他=0; 高中/中专=1	每周 1~2 次	其他=0; 每周 1~2 次=1
大学本科	其他=0; 大学本科=1	每月 1~2 次	其他=0; 每月 1~2 次=1
研究生及以上	其他=0; 研究生及以上=1	半年 1~2 次	其他=0; 半年 1~2 次=1
月收入/生活费		几乎不看	其他=0; 几乎不看=1
1500 元以下	其他=0; 1500 元以下=1		

3.模型的因素分析结果

本文根据 SPSS 23.0 软件进行多元有序 Logistic 回归分析的结果，确定了各因素对大众关于相声了解度的影响程度，并对所有变量根据公式（4）进行回归，将各种变量控制在 10% 的统计水平下，最后得出回归分析的结论，详见表 10。

表 10 回归结果分析

变量名称	系数	标准误	变量名称	系数	标准误
性别			1500~3000 元	-0.613***	0.182
男	0.558***	0.142	3000~6000 元	0.002	0.148
女	-0.558***	0.142	6000~9000 元	0.638***	0.178
年龄			9000~12000 元	0.841***	0.224
20 岁以下	-1.065***	0.236	12000 元以上	0.774***	0.287
21~30 岁	-0.173	0.139	喜爱程度		
31~40 岁	0.618***	0.161	非常喜欢	1.714***	0.190
41~50 岁	0.137	0.211	挺喜欢的	0.347**	0.139
51 岁以上	0.200	0.313	一般	-1.718***	0.173
文化程度			不喜欢	-1.999***	0.439
小学及以下	-0.419	0.605	观看频率		
初中	-1.707***	0.373	每天都看	1.990***	0.264
高中/中专	0.195	0.197	每周 1~2 次	0.955***	0.144
大学本科	0.150	0.152	每月 1~2 次	-0.021	0.160
研究生及以上	0.019	0.234	半年 1~2 次	-1.492***	0.219
月收入/生活费			几乎不看	-3.089***	0.281
1500 元以下	-1.408***	0.211			

注：*、**、*** 分别表示在 10%、5%、1% 的水平上统计显著

4.模型结论

性别的两个变量均通过了 1% 统计水平的显著性检验，且男性为正向影响，女性为负向影响，说明男性群体对相声的了解程度普遍高于女性。

在年龄方面我们可以看到 21~30 岁、41~50 岁和 51 岁以上均未通过显著性检验，说明这三个年龄段的人对相声了解程度的影响不显著。其余两个年龄段均通过了 1% 统计水平的显著性检验，其中 20 岁以下为负向影响，说明青少年群体普遍对相声不太了解；而 31~40 岁为正向影响，因此我们可以得到该年龄段的人为最了解相声的群体。

在文化程度方面，除了初中之外，其余四个变量均未通过显著性检验，并且初中也对相声的了解程度呈负向影响，故大众对于相声的了解程度不会因为文化程度的不同而产生变化。

在月收入/生活费方面，可以得到低收入（3000 元以下）的受访者对相声的了解程度普遍偏低，而收入 6000 元以上的受访者对相声的了解程度随着收入增多而逐渐增加，说明高收入人群为更了解相声的群体。

在受访者对于相声的喜爱程度方面，可以看出随着喜爱程度的增大，其对相声的了解程度也逐渐增加，生活中人们往往热衷于为自己所喜爱的事物付出，而第一步就是要了解它，因此这是和实际相符合的一个结论。

在观看相声的频率方面，观看相声每周 1~2 次甚至更频繁的受访者对相声的了解程度较高，而半年 1~2 次或者几乎不看的受访者对相声的了解程度较低，可以得出大众对相声的了解程度也随着观看频率的增大而增加。

（二）相声观众的人物画像——基于K-Means聚类

观众对相声的了解和关心程度以及喜欢的题材、付费意愿等不尽相同，若要更好的了解相声观众的结构特点，需要建立与观众的对话，描述并勾勒观众的行为、价值观及需求。对此，本章利用聚类分析的方法将受访的相声观众进行人群的划分，将每一类观众的需求和关注点相结合，旨在细化目标人群，从而更好地为每一类人群制定表演方案。

1.观众分类的变量选择

为探析相声观众的人群特征，团队利用 K-means 聚类分析方法对被调查者进行聚类，将关注的特征变量进行选型独立拆分和 0-1 重新编码，具体包括：

- （1）受访者基本信息（A1 性别、A2 年龄、A3 月收入/生活费、A4 职业类型）；
- （2）受访者认知度（B1 对相声的了解程度、B2 对相声的喜爱程度、B3 观看相声的频率）；
- （3）受访者行为特征（C1 喜欢的相声题材、C2 观看相声的途径、C3 对于观看相声的付费意愿、C4 喜欢的相声演员）。

2.基于 K-均值聚类分析法对相声观众进行聚类

K-均值聚类是一种快速样本聚类方法，在聚类个数已知的情况下，特别适合对大样本数据进行分析，具有计算量大、对系统要求低、占用内存小、处理速度快

的优点，但是这种方法只适合对样本进行聚类，不适合变量聚类。由于是对相声观众样本进行聚类，所以可以采用这种方法。

在任命 K-均值聚类分析的类别个数为 3 的条件下，SPSS 软件输出结果如下表 11 所示：

表 11 K-Means 均值聚类分析结果

变量	聚类			变量	聚类		
	1	2	3		1	2	3
男 (A1)	0	1	1	几乎不看 (B3)	0	0	0
女 (A1)	1	0	0	国家发展 (C1)	1	0	0
20 岁及以下 (A2)	0	0	0	校园趣事 (C1)	0	0	0
21~30 岁 (A2)	0	0	1	历史知识 (C1)	1	1	0
31~40 岁 (A2)	0	1	0	传奇轶事 (C1)	1	0	1
41~50 岁 (A2)	1	0	0	工作生活 (C1)	1	0	1
51 岁以上 (A2)	0	0	0	江湖野史 (C1)	0	0	1
1500 元以下 (A3)	0	0	0	长篇故事连载 (C1)	0	0	0
1500~3000 元 (A3)	0	0	1	春晚等电视节目 (C2)	1	1	1
3000~6000 元 (A3)	1	0	0	手机视频软件 (C2)	0	1	1
6000~9000 元 (A3)	0	0	0	社交平台 (C2)	1	1	0
9000~12000 元 (A3)	0	1	0	茶馆等现场表演 (C2)	1	1	0
12000 元以上 (A3)	0	0	0	0-100 元 (C3)	0	0	1
在校学生 (A4)	0	0	1	100-200 元 (C3)	1	0	0
机关干部/公务员 (A4)	1	0	0	200-300 元 (C3)	0	0	0
企业管理者 (A4)	0	1	0	300-500 元 (C3)	0	0	0
普通职员 (A4)	0	0	0	500-1000 元 (C3)	0	1	0
专业人员 (A4)	0	0	0	1000 元以上 (C3)	0	0	0
自由职业者 (A4)	0	0	0	马三立 (C4)	1	1	0
了解程度为 1 (B1)	0	0	0	侯宝林 (C4)	1	1	0
了解程度为 2 (B1)	0	0	1	刘宝瑞 (C4)	1	1	0
了解程度为 3 (B1)	0	0	0	马季 (C4)	1	1	0
了解程度为 4 (B1)	1	0	0	郭德纲 (C4)	0	1	1
了解程度为 5 (B1)	0	1	0	于谦 (C4)	0	1	1
不喜欢 (B2)	0	0	0	岳云鹏 (C4)	0	1	1
一般 (B2)	0	0	1	孙越 (C4)	0	1	0
挺喜欢的 (B2)	1	0	0	冯巩 (C4)	1	1	0

非常喜欢 (B2)	0	1	0	相声新生代 (C4)	0	1	0
每天都看 (B3)	0	1	0	大兵 (C4)	0	1	0
每周 1~2 次 (B3)	0	0	0	高晓攀 (C4)	0	1	0
每月 1~2 次 (B3)	1	0	0	曹云金 (C4)	0	1	0
半年 1~2 次 (B3)	0	0	1				

3. 聚类结论

基于相声观众的基本信息、认知度和行为特征，我们将人群分为以下三类：

表 12 观众类型分类表

类别	基本信息	认知度	行为特征	观众类型
第一类	性别：女性 年龄：41~50 岁 月收入/生活费： 3000~6000 元 职业：机关干部/公务员	对相声了解程度为 4 喜爱程度为 3 每月观看 1~2 次	①喜欢国家发展、历史知识、 传奇轶事、工作生活等题材 ②通过春晚等电视节目、社交 平台以及茶馆等现场表演的 方式观看相声 ③能够接受付费价格区间为 100-200 元 ④喜欢马三立、侯宝林、刘宝 瑞、马季等相声演员	收入稳定的中年 女性
第二类	性别：男性 年龄：31~40 岁 月收入/生活费： 9000~12000 元 职业：企业管理者	对相声了解程度为 5 喜爱程度为 4 几乎每天都观看相声	①喜欢历史知识题材的相声 作品 ②通过几乎所有途径观看相 声 ③能够接受付费价格区间为 500-1000 元 ④对题项所列的相声演员几 乎都喜欢	收入较高的职场 青年男性
第三类	性别：男性 年龄：21~30 岁 月收入/生活费： 1500~3000 元 职业：在校学生	对相声了解程度为 2 喜爱程度为 2 半年观看 1~2 次	①喜欢传奇轶事、工作生活、 江湖野史等题材 ②通过春晚等电视节目、手机 视频软件的方式观看相声 ③能够接受付费价格区间为 100 元以下 ④喜欢郭德纲、于谦、岳云鹏 等相声演员	对相声略有兴趣 的男大学生

三、影响观众喜爱相声的因素

（一）观众喜爱的相声演员特征主题分类——基于Text-CNN模型

1.Text-CNN 模型介绍

Text-CNN 模型是由 Yoon Kim 在论文(2014 EMNLP) Convolutional Neural Networks for Sentence Classification 中提出的，其主要思想是将不同长度的短文作为矩阵输入，使用多个不同规模的筛选器去提取句子中的关键信息，并用于最终的分类。模型的详细过程原理图如下：

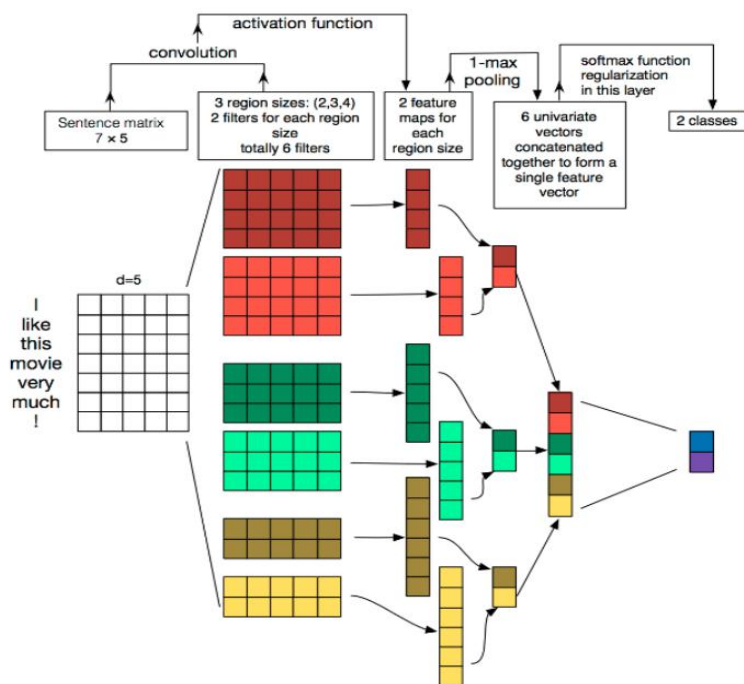


图 26 Text-CNN 模型详细过程原理

2.模型构建过程

根据问卷第 16 题填空题“用几个词形容自己喜欢的相声演员的特点”，我们得到观众喜爱的相声演员特征的文本。经过分词和清洗，对文本进行 Text-CNN 主题分类。利用中科院计算所的语义分析系统 NLPIR 统计所有词语频数，挑选出词频最高的 30 个词语，利用词语之间的相关性建立模型，将主题分成三类，每一类对应着若干词语及其频数。

3.相声演员特征主题分类结论

得出的 Text-CNN 词语主题分类如下表 13 所示：

表 13 Text-CNN 词语主题分类表

主题 1	主题 2	主题 3
段子	多才多艺	好笑
形象	喜庆	活泼
特点	敬业	憨厚
语言	表演	生动
专业	创新	丰富
功底	夸张	诙谐
表情	到位	可爱
动作	喜欢	有趣
内涵	逗乐	风趣
包袱	生活	幽默

通过上表可以得出，主题一大多为名词，反映了观众在观看相声表演时对于演员比较关注的点，例如包袱、内涵、动作、表情等。说明观众喜爱的相声演员需要具备扎实的基本功，表演的时候表情、动作收放自如，语言生动形象且具有特点，包袱、段子引人发笑，表演的内容有文化内涵。

主题二体现了观众对相声演员表演中细节的偏好，“多才多艺”、“夸张”等为观众对相声演员表演形式的偏好，即希望相声演员不拘泥于“说”这一种形式，可以结合“唱”等形式，在必要的时候表现夸张，从而给观众带来不一样的观感；“逗乐”、“生活”、“创新”等为观众对相声演员表演内容的偏好，观众喜爱的相声表演内容应该是贴近生活，能够引起共鸣，并且要不断创新，达到逗乐的效果的。

主题三多为形容词，体现了相声演员吸引观众喜欢的自身人格魅力，其中“幽默”、“风趣”、“诙谐”等词出现上百次，说明观众喜爱的相声演员是能够发挥相声的喜剧效果，给大家带来快乐的；此外，“憨厚”、“活泼”、“可爱”等同样为观众喜爱的相声演员的特点。

这三个主题从不同角度反映了观众喜爱的相声演员的主要特征。下图为观众喜爱的相声演员特征的部分关键词提取：



图 27 观众喜爱的相声演员特征关键词提取

(二) 各类要素对观众喜爱相声程度的影响研究——基于随机森林

1.随机森林模型介绍

随机森林模型可以解释若干自变量(X_1 、 X_2 、 \cdots 、 X_k)对因变量 Y 的作用。随机森林模型可以应用在分类和回归问题上。当因变量 Y 是分类变量时,为分类树,当因变量 Y 是连续变量时,则为回归树。由于本次模型中的因变量——观众对相声的喜爱程度是连续变量,旨在探究内容创新、幽默、演员功底等 14 个因素(自变量 X_i)对观众对相声的喜爱程度(因变量 Y)的影响程度,因此我们选择构建回归树模型。

2.模型的重要程度排序结果分析

表 14 观众对相声喜爱程度的影响指标重要程度排序表

影响指标	IncNodePurity	排序
价格	0.089963	1
相声形式	0.082110	2
观看场所	0.081983	3
营销推广	0.080575	4
体验性价比	0.080242	5
品牌影响力	0.076382	6
政府支持力度	0.075741	7
表演形式	0.075455	8
社会氛围	0.073770	9

剧本	0.070391	10
内容创新	0.062778	11
文化内涵	0.059075	12
幽默	0.046093	13
演员功底	0.045444	14

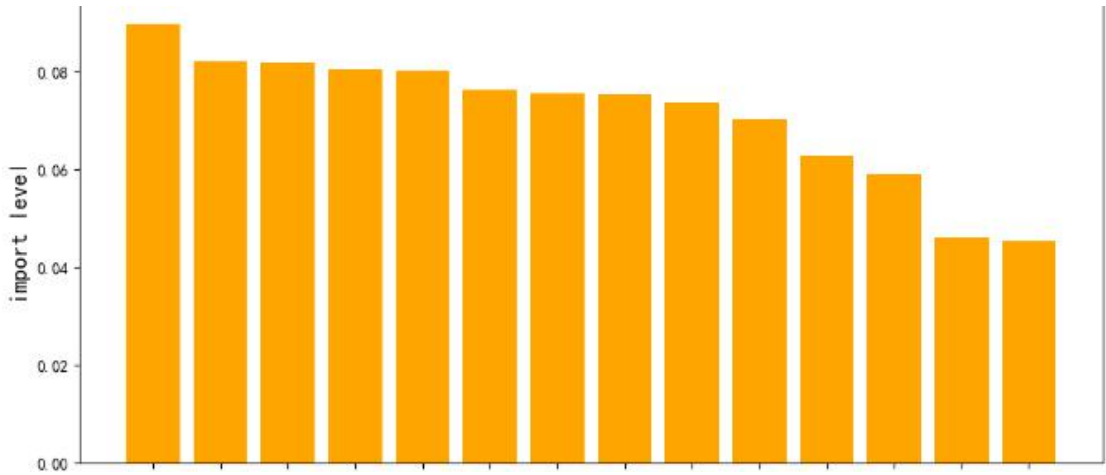


图 28 观众对相声喜爱程度的影响指标随机森林重要性排序图

3.模型结论

从上表与上图中可以看出，影响观众对相声喜爱程度排在最后两位的分别是幽默和演员功底。相声中各种包袱的幽默度在比较小的程度上决定了观众对于相声的喜爱程度并引导着他们成为付费者。演员功底是相声表演的基础，一个优秀的演员往往能让观众看的更开心，更有收获感。而从排序结果可以看出，观众对于相声的喜爱程度和幽默本身无太大的关系，即**观众无论多大程度上喜欢相声都会认为幽默和演员功底很重要。**

而排名第一位——价格这个因素的得分与喜爱程度关联最大，也就是价格这个因素的重视程度对用户的喜爱程度影响是 14 个因素里最大的，其次是场所，而场所的环境是否卫生干净，座位角度设置是否合理，服务态度等等影响着观众的体验满意度。故对观众影响最大的因素其实还是外部因素，换言之，决定用户是否会去听相声的核心标准，仍然是那些外部经济因素，其中价格和场所是最关键的。故我们可以得出一个与平时认知不太相符的结论：**价格和场所带来的体验舒适度与性价比竟然才是真正使消费者对于相声喜爱程度产生差异的原因！**

第五部分 观众对相声付费意愿的综合评价分析

一、影响观众付费观看相声意愿的因素分析——基于结构方程模型

（一）结构方程模型介绍

结构方程模型是基于变量的协方差矩阵来分析变量之间关系的一种统计分析方法，主要用于处理多个原因、多个结果的关系，以及含有潜变量的问题。

（二）潜变量设定与研究假设

- 1.付费意愿：付费意愿是指消费者想要购买商品或服务的可能性。付费意愿越高，消费者购买相应商品的可能性越高。在本模型中，付费意愿是研究的核心因素，会受到其他因素的影响。
- 2.整体体验：整体体验是消费者在衡量自身感觉上对产品整体效用的一种评估。消费者整体体验越好，对相应的商品或服务的购买意愿越强。
- 3.氛围感知：在本文的研究情境下，氛围感知是指观众在购买相声服务后相较于免费的服务对其具有更大的效用。如果观众购买之前的氛围感知不强，即品牌影响力、政府支持力度、社会氛围影响很小，则观众对于相声服务的感知价值就低，付费意愿也就越低。
- 4.营销感知：在本文的研究情境下，营销感知是观众在价格、营销推广、体验性价比等方面感受到的价值效应水平，以此判断相声是否对自己有用、服务是否值得购买。
- 5.消极因素：本文中的消极因素是基于预调查中观众反感的因素总结而成，对相声付费意愿有消极影响。

表 15 潜变量及对应可测变量

潜变量	可测变量
营销感知 A	价格 A1
	营销推广 A2
	体验性价比 A3
整体体验 B	文化内涵 B1
	幽默 B2
	演员功底 B3
	剧本 B4

	内容创新 B5
氛围感知 C	品牌影响力 C1
	政府支持力度 C2
	社会氛围 C3
付费意愿 D	能接受的一场相声演出的最高价格范围 D1
	是否付费观看过相声表演 D2
消极因素 E	没有产生共鸣 E1
	语言地方性太强 E2
	表演形式过于单一 E3
	内容无趣 E4
	新时代娱乐方式的冲击 E5
	身边听相声的氛围不浓 E6
	现场服务考虑不周 E7

根据 5 个潜变量之间的影响情况，可以建立如下假设：

表 16 研究假设

编号	研究假设
H1	营销感知对于整体体验有显著正向影响
H2	氛围感知对于整体体验有显著正向影响
H3	营销感知对于付费意愿有显著正向影响
H4	整体体验对于付费意愿有显著正向影响
H5	氛围感知对于付费意愿有显著正向影响
H6	消极因素对于付费意愿有显著负向影响
H7	营销感知与氛围感知有显著相关关系

根据以上分析，现采用 AMOS 软件，按照结构方程模型路径图的符号规则，设置好因果关系路径图。

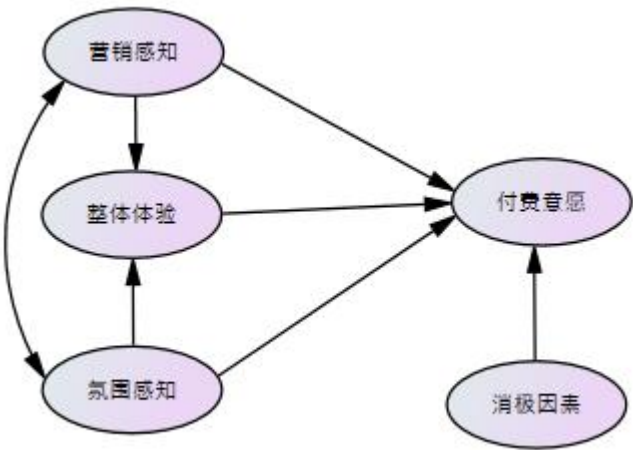


图 29 观众相声付费意愿影响因素的结构方程模型初始路径图

（三）模型修正与检验

初始模型部分变量拟合后出现路径不显著或部分关系无法得出确定值，分析得出原因在于结构方程模型设计存在缺陷。虽然此模型结构理论上可行，但在数据上经验证不可执行，因此进行调整，删除 H7 路径及部分不显著可测变量路径后得到新模型，如下图所示。后又根据 MI 系数大小进行模型修正，添加了 e21 与 e23 相关、e20 与 e21 相关、e15 与 e16 相关、e18 与 e19 相关、e20 与 e23 相关、e18 与 e23 相关的路径。

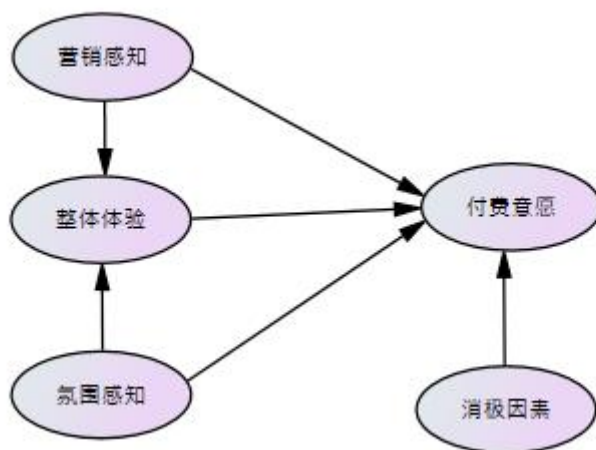


图 30 观众相声付费意愿影响因素的结构方程模型修正后的路径图

在 AMOS 中绘制模型，代入数据，进行模型检验与拟合优度结果见下表。

表 17 模型检验与拟合优度结果

拟合指标	CMIN/DF	GFI	IFI	CFI	RMSEA
建议值	<2	>0.9	>0.9	>0.9	<0.1
结果数据	1.898	0.888	0.872	0.870	0.058

模型的拟合优度是检验理论模型是否科学的重要依据。将本研究理论模型的拟合优度与常规的指标建议取值范围作对比，发现表中绝大多数指标都接近或落在了取值区间内：GFI、IFI、CFI 基本符合要求，其他指标均完全符合要求。最终认为模型的总体拟合情况和适配度均符合要求，拟合效度良好。

（四）模型参数估计

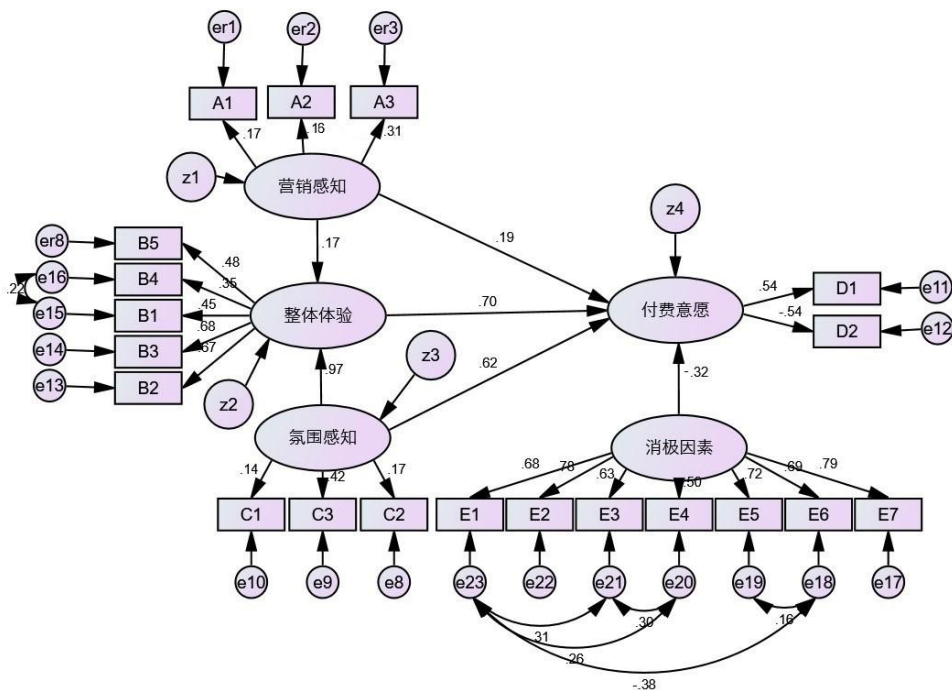


图 31 影响观众相声付费意愿因素的标准化结构方程模型路径系数图

表 18 潜变量之间的标准化路径系数表

			Estimate	S.E.	C.R.	P	对应假设	假设是否通过
整体体验	<---	氛围感知	.974	.052	9.165	.016	H2	通过
整体体验	<---	营销感知	.168	.085	5.324	.000	H1	通过
付费意愿	<---	消极因素	-.328	.050	2.312	.003	H6	通过
付费意愿	<---	氛围感知	.602	.106	11.790	.027	H5	通过
付费意愿	<---	整体体验	.684	.076	-2.701	.007	H4	通过
付费意愿	<---	营销感知	.092	.079	2.735	.006	H3	通过

表中，所有 P 值均 <0.05 ，具有显著统计意义，可测变量对潜在变量具有显著性影响，表示每个问题对于潜在变量都能较好的衡量。

从上表可以看出所有的量表和因子之间的负载系数在 0.01 的显著性水平上具有统计属性。因此修改后的测量模型有效。

（五）模型结果分析

第一，整体体验对于观众相声付费意愿的影响程度最大，且影响方向为正。根据模型给出的路径系数可以看出，整体体验影响付费意愿的标准化路径系数为 0.70，影响程度极大且影响方向为正，符合本结构方程模型的假定条件，故可判

断出相声的观众最为看重的是观看相声时的整体体验感。其中，幽默感与演员的功底是观众最为关注的两个因素。

第二，氛围感知对于观众相声付费意愿的影响程度排名第二，且影响方向为正。氛围感知影响付费意愿的标准化路径系数为 0.62，影响方向为正，符合结构方程模型的假定条件，即品牌影响力、政府支持力度、社会氛围都是影响观众相声付费意愿的影响因素，路径系数分别为 0.04、0.17 和 0.42。其中社会氛围的影响程度最大，可以看出人们在一定程度上有追赶相声热点潮流的意识。

第三，消极因素同样是观众相声付费意愿的重点考虑因素。消极因素影响观众相声付费意愿的路径系数为-0.32，影响力中等且方向为负。即没有产生共鸣、语言地方性太强、表演形式过于单一、内容无趣、新时代娱乐方式的冲击、身边听相声的氛围不浓、现场服务考虑不周这些因素都会在一定程度上对观众相声付费意愿产生消极影响，特别是语言地方性太强、现场服务考虑不周会严重打击观众对相声付费的意愿程度。因此充分考虑观众的各方面需求以提供相应服务非常必要。

第四，营销感知也对观众相声付费意愿有一定影响但影响程度较小。其中，体验性价比是观众较为看重、影响程度最大的一个因素，符合研究对人们付费心态的基本预期。因此，如何改进服务、如何让观众感受到付费后带来的高性价比感知是整个营销环节中最重要的一环。

二、观众选择付费观看相声与否的决策行为分析

——基于贝叶斯决策模型

（一）贝叶斯模型介绍

贝叶斯决策理论是在不完全情报下，对部分未知的状态用主观概率估计，然后用贝叶斯公式对发生概率进行修正，最后再利用期望值和修正概率做出最优决策。贝叶斯决策模型是一种基本的统计决策模型，对于描述复杂市场环境下的决策行为具有重要意义。

贝叶斯决策属于风险型决策，决策者虽不能控制客观因素的变化，但却掌握其变化的可能状况及各状况的分布概率，并利用期望值即未来可能出现的平均状况作为决策准则。其基本思想是：

1.已知条件概率密度参数表达式和先验概率

$$P(A) \sum_{i=1}^n P(B_i) P(A|B_i)$$

2.利用贝叶斯公式转换成后验概率

$$P(A|B_i) = \frac{P(B_i)P(A|B_i)}{\sum_{i=1}^n P(B_i)P(A|B_i)}$$

3.根据后验概率与先验概率差值比率进行决策分析

$$P = \frac{P(B_i|A) - P(A)}{t}$$

（二）变量选取

观众对相声的喜爱程度、对相声的了解程度、对目前相声发展的满意程度都是密切影响观众对相声付费的意愿程度的主要因素。因而综合考虑变量变化的连贯性和类推性，我们最终选择了这三个变量进行观众对相声付费意愿程度的贝叶斯预测。

（三）贝叶斯计算过程

依据关联分析筛选出具有显著相关关系的几组变量，画出计算路径，并依照此路径将两两变量之间的先验概率与后验概率进行比较。

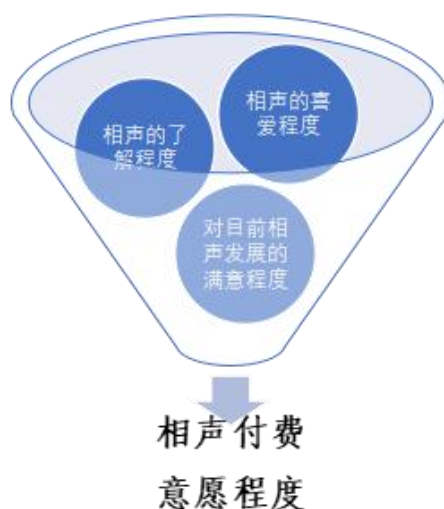


图 32 影响相声付费意愿程度的因素

按照上述计算路线，利用贝叶斯概率公式计算两两变量之间的后验概率，探究观众对于相声付费意愿程度的影响因素，进而给出未来相声观众付费意愿程度可能随之发生的变化方向，即市场前景预测。利用卡方检验，筛选通过检验并和相声付费意愿之间具有显著差异的变量，再对通过检验的影响因素与观众相声付费意愿之间计算得到的先验概率、后验概率进行比较，获得贝叶斯模型一步迭代过程如下表所示：

表 19 贝叶斯模型的一步迭代过程

			相声付费意愿程度 B			检验		
			不愿意	一般	愿意	卡方值	P 值	是否显著
喜爱程度 A1		$p_{\text{后}}(B A3_2)$	0.330	0.330	0.330	67.846	0.000	显著
	不喜爱	$p_{\text{后}}(B A1_1)$	0.174					
	喜爱	$p_{\text{后}}(B A1_2)$		0.452	0.363			
		p	0.021	0.017	0.11			
了解程度 A2		$p_{\text{先}}(B)$	0.330	0.330	0.330	144.263	0.000	显著
	不了解	$p_{\text{后}}(B A2_1)$	0.165					
	了解	$p_{\text{后}}(B A2_2)$		0.479	0.383			
		p	0.07	0.120	0.036			
满意程度 A3		$p_{\text{先}}(B)$	0.330	0.330	0.330	30.401	0.000	显著
	不满意	$p_{\text{后}}(B A3_1)$	0.126					
	满意	$p_{\text{后}}(B A3_2)$		0.537	0.398			
		p	0.007	0.06	0.04			

如上表可知，观众对相声的喜爱程度、对相声的了解程度、对目前相声发展的满意程度这三种变量在不同的条件下能够对相声付费意愿程度的因变量产生不同程度的影响。以第一步迭代为参照，在阈值达到 0.5 即观众对相声付费意愿程度达到 50% 时停止迭代。按照此步骤，以上述三个变量为主要研究节点，对观众对相声付费的意愿程度进行统计预测。

（四）观众对相声付费的预测结果

1. 喜爱程度的提升带来相声付费意愿的大幅增长，但难以形成持续的市场增速，需进一步从其他方面刺激观众的付费意愿。

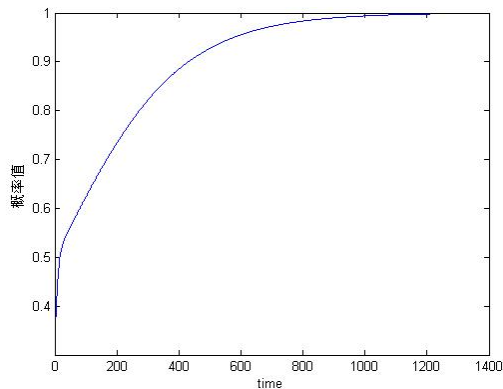


图 33 喜爱程度载率值

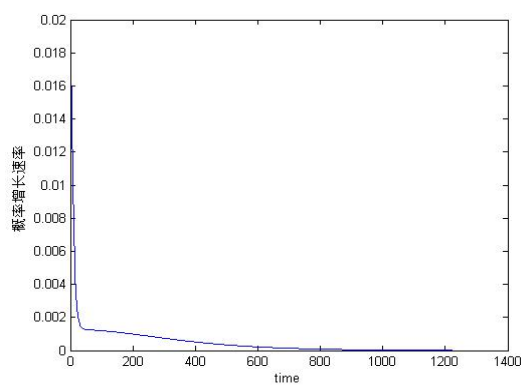


图 34 喜爱程度载率增长速率

随着观众对相声喜爱程度的增加，观众对于相声付费的意愿程度也会明显提高。但是图像的拐点显示，喜爱程度所能形成的相声付费的意愿程度在超过 50% 时就会不断下降直到最终饱和。只有不到 50% 的观众愿意看完春晚的全部相声表演，且大部分观众觉得一段相声表演最适合的时长为 20-25 分钟。一旦观众对于相声形成了初步喜爱之后，相声的喜爱程度的增加就逐渐不是引导观众付费的关键性因素了，如何进一步从其他方面刺激观众的付费意愿则显得十分重要。如开发创新更多的相声表演形式、培养更多广受欢迎的相声演员，实施一些奖励机制鼓励观众观看付费相声，尽可能多的满足观众需求，观众才会更愿意为相声付费。

2.通过微信、微博等社交平台与抖音、快手等直播平台加大观众对相声的了解程度与自己团体的宣传力度，发展一些忠实付费观众，利用一些免费服务吸引潜在观众转发友情链接，以此增加付费观众流。

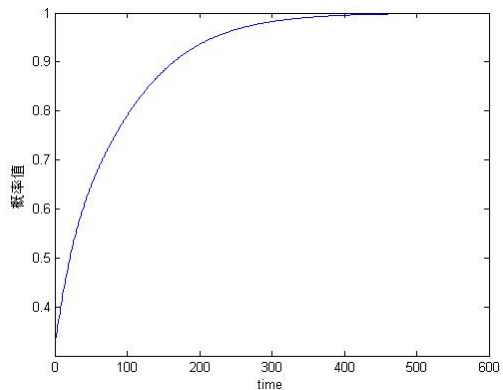


图 35 了解程度载率值

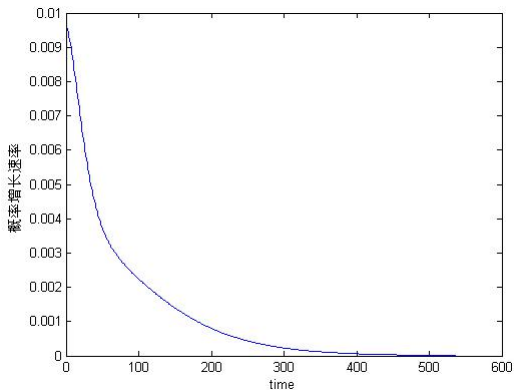


图 36 了解程度载率增长速率

观众对相声了解程度的不断提升会带来其对于相声付费的选择概率的增加。总体上，这种由了解程度带来的选择倾向的增大是不断放缓的。以一般了解——比较了解的了解程度为临界点，观众的相声付费意愿概率将超过 50%，但根据在

此之后增速减缓可知，后续宣传营销带来的效果对于个人选择的影响程度会呈现越来越小的状态，因此后期的重点关注方向为如何发展忠实观众、寻找潜在观众。

3. 观众观看相声的频率超过每月 1~2 次、同时对观看付费相声的满意次数达到 2~3 次后，付费意愿显著提升。

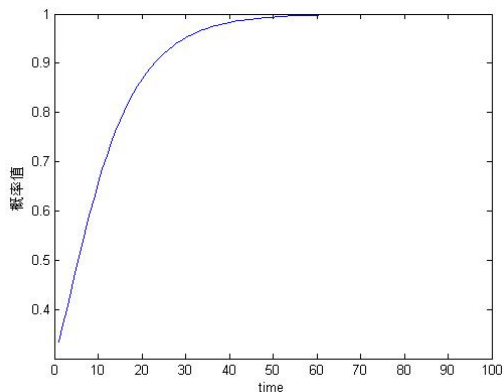


图 37 满意程度载率值

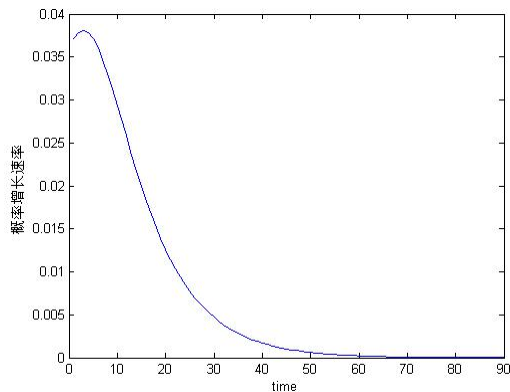


图 38 满意程度载率增长速率

随着观众对相声满意度的不断提升，观众选择相声付费的概率也会不断增大，并且增长的速率呈现出前期快后期慢的特点。在满意程度达到比较满意后达到峰值，其增长速率放缓慢慢慢达到饱和的趋势。说明了观众的满意程度达到比较满意后，其对相声付费意愿逐渐不再增加。因此，不定期的推出观众福利、开展优惠活动等提高观众的感知性价比是一个很好的方式。

三、观众对相声团体选择的影响因素分析

——基于 AHP 层次分析法

（一）AHP 模型介绍

层次分析法是一种解决多目标的复杂问题的定性与定量相结合的决策分析方法。该方法是将一个复杂的多目标决策问题作为一个系统，将目标分解为多个目标或准则，进而分解为多指标（或准则、约束）的若干层次，通过定性指标模糊量化方法算出层次单排序（权数）和总排序，以作为目标（多指标）、多方案优化决策的系统方法。为探究观众对相声团体选择的影响因素，本文构建了以相声团体选择为核心指标的三层次多指标评价体系，求解判断矩阵得到各指标的影响程度大小。

（二）AHP 评价指标体系的构建

通过查阅文献结合问卷调查，本文将影响观众选择相声团体的因素分为两类，分别是服务价值和服务产品。服务价值包括票价、营销推广以及品牌影响力；服务产品包括内容创新、演员、剧本以及文化内涵。从而构建的评价指标体系如下图所示：

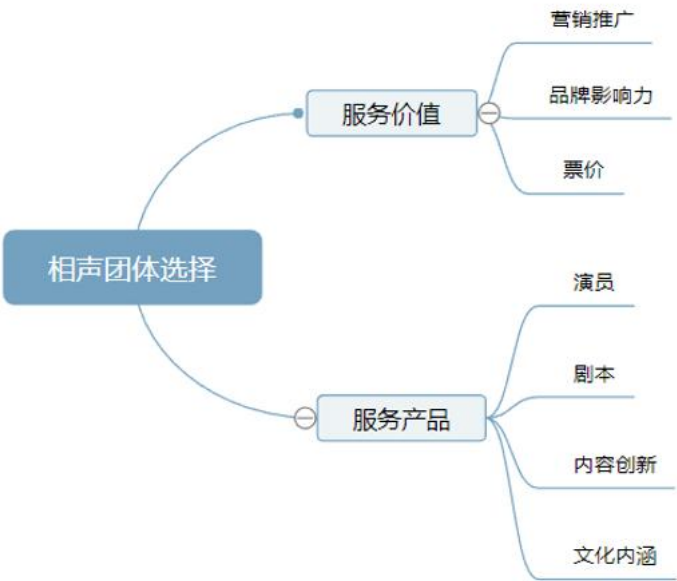


图 39 评价指标体系图

（三）观众对相声团体选择的影响因素权重设定

通过查阅相关文献，并结合相声发展实际情况，确定以上影响因素并由观众对以上因素进行 1—5 级打分，最终确定观众对相声团体选择的影响因素评价指标共包括 3 个评价层，7 个决策层的赋权情况。

通过调查数据，运用和积法，构建判断矩阵并求解，所得结果如下表所示：

表 20 一致性检验结果

	λ_{max}	$C.I$	$R.I$	$C.R$	一致性检验
B1	3.0183	0.0091	0.5800	0.0158	通过
B2	4.0413	0.0138	0.9000	0.0153	通过

表 21 各观测指标的权重分配表

核心指标	结构指标	观测指标	权重	排序
相声团体选择 A	服务价值 B1	票价 C1	0.1222	5
		营销推广 C2	0.0868	7
		品牌影响力 C3	0.1622	3
	服务产品 B2	内容创新 C4	0.2091	1
		演员 C5	0.1122	6
		剧本 C6	0.1377	4
		文化内涵 C7	0.1699	2

（四）模型结论

由表中结果，我们可以得到：在服务产品方面，大多数指标权重较大。其中内容创新指标的权重最高，其次是文化内涵，其权重分别为 0.2091、0.1699，说明内容创新对观众选择相声团体的影响最大，文化内涵的影响程度次之；在服务价值方面，品牌影响力的权重为 0.1622，排序第三，说明品牌影响力对于观众的选择偏好也具有一定影响。票价与营销推广相对排名靠后，这二者影响程度较低。

综上所述可知：内容创新与文化内涵这两个因素对观众选择相声团体的影响程度最大，因此相声团体要适时对表演的相声进行内容创新，以符合当下观众的审美，相声的文化内涵也是相声团体应注重的方面；品牌影响力对于观众选择相声团体也有一定影响，因此各相声团体应适当关注其品牌影响力的大小，尽可能地加大其品牌影响力输出，以期吸引更多观众。

四、观众对付费观看相声的满意度评价分析——基于云模型

（一）云模型介绍

云模型属于不确定性人工智能范畴，其应用领域包括智能控制、数据挖掘，尤其是在多属性决策和分析评价领域占有重要的地位。基本的定义如下：设 X 是一个普通集合， $X=\{x\}$ ，称为论域。关于论域 X 中的模糊集合 A ，是指对于任意元素 x 都存在一个有稳定倾向的随机数 $U_a(x)$ ，叫做 x 对 A 的隶属度。如果论域中的元素是简单有序的，则可以看作是基础变量，隶属度在 X 上的分布叫做隶属云；如果论域中的元素不是简单有序的，而根据某个法则 f ，可将 X 映射到另一个有序的论域 X 上， X 中的一个且只有一个 x 和 x 对应，则 X 为基础变量，隶属度在 X 上的分布叫做隶属云。云模型用三个数据来表示其特征：期望，一般用符号 E_x 表示；熵，一般用符号 En 表示；超熵，一般用符号 He 表示。云有

两种发生器，正向云发生器和逆向云发生器，分别用来生成足够的云滴和计算云滴的云数字特征（Ex，En，He）。观众付费观看相声的满意度是付费变现的直接动因，因此本文运用云模型将观众付费观看相声的满意程度这个定性概念转换成定量概念，并用云图加以表现。

（二）满意度评价方法

根据问卷数据，我们将付费观看相声表演前后各项指标的期望值与实际值的数据进行整理，期望值是指观众在付费观看表演前对相声表演各方面的期待值，将其作为各个指标权重，即每个指标所对应的权重系数设为 $A=\{A_1, A_2, \dots, A_{10}\}$ ，根据问卷数据，通过云模型的逆向云发生器对每个指标的得分情况进行计算，分别求出各个指标的权重系数。权重系数的分值分布在一定的范围，体现了一定的模糊性和随机性。采用权重矩阵 A 表示为：

$$A = \begin{bmatrix} A_1 \\ \vdots \\ A_{10} \end{bmatrix} = \begin{bmatrix} Ex_{A1} & En_{A1} & He_{A1} \\ \vdots & \vdots & \vdots \\ Ex_{A10} & En_{A10} & He_{A10} \end{bmatrix}$$

根据观众对各指标实际值的打分情况，通过云模型的逆向云发生器进行计算，并用正向云发生器生成足够的云滴，以此生成各指标的云图。各指标实际值设为 $P=\{P_1, P_2, \dots, P_{10}\}$ ，实际值的得分分布在一定范围，其结果表示为：

$$P = [P_1, \dots, P_{10}] = \begin{bmatrix} Ex_1 & En_1 & He_1 \\ \vdots & \vdots & \vdots \\ Ex_{10} & En_{10} & He_{10} \end{bmatrix}^T$$

根据矩阵 P 和权重矩阵 A ，利用模糊合成算子求得综合评价云模型为：

$$W = P \circ A = \begin{bmatrix} Ex_1 & En_1 & He_1 \\ \vdots & \vdots & \vdots \\ Ex_{10} & En_{10} & He_{10} \end{bmatrix}^T \circ \begin{bmatrix} Ex_{A1} & En_{A1} & He_{A1} \\ \vdots & \vdots & \vdots \\ Ex_{A10} & En_{A10} & He_{A10} \end{bmatrix}$$

$$= \begin{bmatrix} Ex_1 * Ex_{A1} + Ex_2 * Ex_{A2} + \dots + Ex_{10} * Ex_{A10} \\ \sqrt{\left(|Ex_1 Ex_{A1}| \sqrt{\left(\frac{En_1}{Ex_1} \right)^2 + \left(\frac{En_{A1}}{Ex_{A1}} \right)^2} \right)^2 + \dots + \left(|Ex_{10} Ex_{A10}| \sqrt{\left(\frac{En_{10}}{Ex_{10}} \right)^2 + \left(\frac{En_{A10}}{Ex_{A10}} \right)^2} \right)^2} \\ \sqrt{\left(|Ex_1 Ex_{A1}| \sqrt{\left(\frac{He_1}{Ex_1} \right)^2 + \left(\frac{He_{A1}}{Ex_{A1}} \right)^2} \right)^2 + \dots + \left(|Ex_{10} Ex_{A10}| \sqrt{\left(\frac{He_{10}}{Ex_{10}} \right)^2 + \left(\frac{He_{A10}}{Ex_{A10}} \right)^2} \right)^2} \end{bmatrix}$$

$$= (Ex \quad En \quad He)$$

（三）模型计算结果与云滴分布情况

根据上述评价方法计算得到的矩阵结果以及各指标实际值的云图如下所示：

表 22 矩阵结果（期望值）

指标	Ex_{An}	En_{An}	He_{An}
演员配合默契度期望值	4.0321	0.7084	0.3748
演员功底期望值	4.1226	0.7770	0.1979
现场互动期望值	3.9212	0.7927	0.2960
文化内涵期望值	4.0336	0.8134	0.2585
题材期望值	4.0774	0.7630	0.2879
创新期望值	4.0540	0.8170	0.2372
与当下热点结合期望值	3.9810	0.7057	0.4085
幽默有趣期望值	4.1328	0.8155	0.1711
性价比期望值	3.9796	0.7478	0.3529
环境舒适度期望值	4.0000	0.7355	0.3823

表 23 矩阵结果（实际值）

指标	Ex_n	En_n	He_n
演员配合默契度实际值	3.9197	0.8025	0.2360
演员功底实际值	4.0073	0.7338	0.3641
现场互动实际值	3.8380	0.8314	0.1542
文化内涵实际值	3.8818	0.8074	0.2523
题材实际值	3.8234	0.8109	0.1521
创新实际值	3.8715	0.8380	0.1727
与当下热点结合实际值	3.8482	0.8414	0.0936
幽默有趣实际值	4.0350	0.7557	0.3475
性价比实际值	3.7927	0.8439	0.0000
环境舒适度实际值	3.8467	0.8030	0.1617

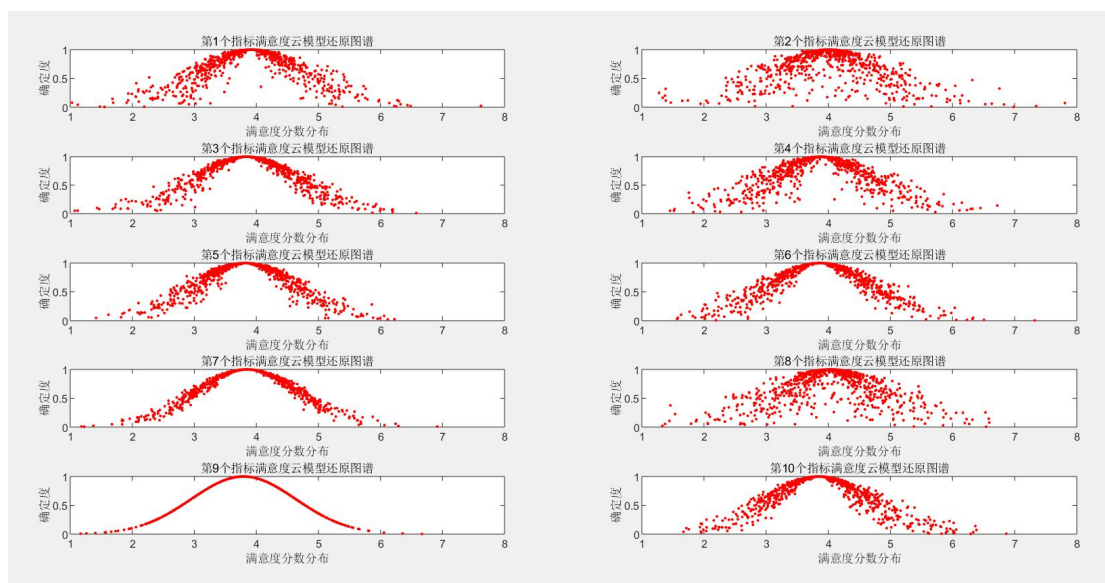


图 40 各指标实际值云图

(四) 满意度综合评价结果

在以上云图中，可以看到第一、三、四、五、六、七、十个指标云滴分布情况大致相同，其指标分别为演员配合默契度实际值、现场互动实际值、文化内涵实际值、题材实际值、创新实际值、当下热点结合实际值、环境舒适度实际值，其云图绝大部分云滴落在评价区间[3,5]之中，落在中间值4的云滴相对较少，且这些指标的期望都在3.8~4之间。说明观众对以上指标的评价结果为“比较满意”。

第二与第八个指标分别为演员功底实际值，幽默有趣实际值。两者的期望值都大于4，云图中云滴分布都比较分散，没有很明显的集中性，说明不同观众对指标的评价具有一定差别，但总体来说评价结果为“满意”。

第九个指标为性价比实际值，其期望为3.7927，在十个指标中数值最小。观察其云图，云图中云滴几乎集中在一条线上，说明观众对该指标的评价较为一致。

综合上述分析，我们可以看到观众对于相声内容的幽默有趣以及相声演员的功底比较满意，对于相声题材、现场互动、性价比、环境舒适度等一般满意。因此，各相声团体应重视演出给观众带来的综合性价比。在氛围方面，相声演员应适时地与观众进行现场互动，通过砸现挂等方式来拉近与观众之间的距离，以增加观众的参与感与体验感；题材方面也应受到相声演员的关注，及时了解观众对于题材的偏好，有助于演员更好的发挥；演出环境更是相声团体应注意的方面，现场的舒适度将直接影响到观众对于相声团体的印象。即使演员在台上的表演很出色，观众也有可能因为受到环境的影响而错过表演的精彩之处，得到较差的体验感。

最后，根据矩阵 P 和权重矩阵 A ，得到综合评价云模型参数的计算结果为：

$$W = P \circ A = (3.8873, 0.5669, 0.5412)$$

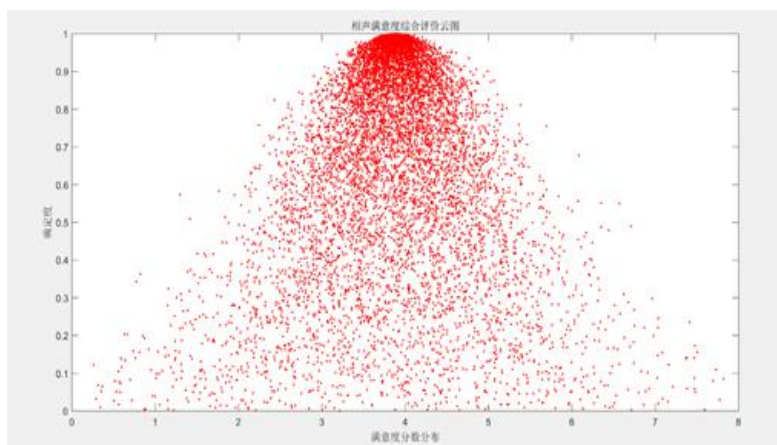


图 41 云滴分布情况图

综合评价云模型参数的期望为 3.8873，表明观众对付费观看的相声表演的评价为较满意。利用综合评价云模型数字特征值和正向云发生器做出满意度综合评价云图，如上图所示。很显然，通过上图的云滴的分布情况来看，绝大部分云滴都落在综合评价分值 3——5 之间，因此观众对于付费观看的相声表演综合评价为比较满意。

第六部分 结论与建议

一、结论

（一）相声观众现状

1.相声观众主体分布集中

相声观众主要集中在青年年龄层次、中高学历、中高收入、且收入较稳定的群体；而在地区观众比例方面，华北、东北地区观看相声的人群比重较高。

2.相声观众的部分偏好具有倾向性

在感兴趣的题材内容中，选择“历史知识”、“传奇轶事”、“工作生活”题材的被调查者均超过五成，因此这三个题材为观众普遍感兴趣的；在观众喜爱的相声演员中，郭德纲、岳云鹏、侯宝林、冯巩为所列题项中被选择比例最高的四人；此外，10~25 分钟的相声更受观众喜爱。

3.相声观众的付费意愿与年龄段、月收入相关

愿意付高价观看相声的年龄集中在 21~50 岁之间，该年龄段具有一定消费能力；总体来看，大众普遍能接受的价格区间在 0~300 元。此外，随着月生活费/月收入的提高，观众愿意支付的最高金额越大。

4.相声观众对相声未来发展普遍持积极态度

在对相声等传统曲艺未来发展前景持何态度的问题上，75.33%的观众持积极态度，并且认为将会有越来越多的人愿意了解传统曲艺文化；同时也有超过七成的观众认为相声可以凭借独特的艺术形式和幽默的内涵走向国门。

（二）观众对相声的偏好情况

1.外部因素对观众关于相声态度的影响

（1）观众对相声团体、相声演员、相声热门地点、相声观看渠道等具有普遍偏好

网络评价词云统计图显示：对于相声团体，字体最大的为德云社，说明在一众相声团体中，网络上对于德云社的讨论最为热烈；在相声演员方面，郭德纲、

张云雷等词频较高，说明这两名相声演员比较受到关注；相声热门地点天津、北京等都为北方地区，侧面说明相声演出在北方地区更受欢迎；相声观看渠道包括春晚与演出等，说明大众对于这两种观看形式讨论较多。

（2）相声网络评价的情感分布具有倾向性

基于情感词典的情感倾向分析显示，正面情感占 65%，负面情感占 35%。说明相声的网络评价大多数都为正面的，但还是存在一部分负面的评价。在负面情感中，“哀”、“恶”情绪占比较大。

2.影响观众对相声偏好的基本情况因素

（1）各群体对相声的了解程度不同

多元有序 Logistic 模型显示，男性群体对相声的了解程度普遍高于女性，甚至大部分女性是不了解相声的；其次，31~40 岁年龄段的人以及高收入人群分别为各类中最了解相声的人群，大众关于相声的了解程度基本不会因为文化程度的不同而产生变化；此外，随着喜爱程度的增大，其对相声的了解程度也逐渐增加；大众对相声的了解程度也随着观看频率的增大而增加。

（2）相声观众的人物画像分为三类

K-means 聚类分析显示，相声观众可以分为三个群体：

第一类人群为收入稳定的中年女性，喜欢国家发展、历史知识、传奇轶事、工作生活等题材，通过春晚等电视节目、社交平台以及茶馆等现场表演的方式观看相声，能够接受的付费价格区间为 100-200 元，同时喜欢马三立、侯宝林、刘宝瑞、马季等相声演员；

第二类人群为收入较高的职场青年男性，喜欢历史知识题材的相声作品，通过几乎所有途径观看相声，能够接受的付费价格区间为 500-1000 元，同时对题项所列的相声演员几乎都感兴趣；

第三类人群为对相声略有兴趣的男大学生，喜欢传奇轶事、工作生活、江湖野史等题材，通过春晚等电视节目、手机视频软件的方式观看相声，能够接受的付费价格区间为 100 元以下，同时喜欢郭德纲、于谦、岳云鹏等相声演员。

（三）影响观众喜爱相声的因素

1.观众喜爱的相声演员特征：分为三个主题

Text-CNN 模型将被调查者喜爱的相声演员的特征分成了三类。主题一大多

为名词，反映了观众在观看相声演员表演时对于演员比较关注的点。主题二体现了观众对相声表演中细节的偏好。主题三多为形容词，体现了相声演员吸引观众喜欢的自身人格魅力。

2.观众喜爱相声程度的影响因素具有重要程度排序：价格和体验性价比排名前二

随机森林模型的排序结果显示观众对价格这个因素的重视程度会最大程度上影响用户对相声的喜爱程度，其次排名第二的是体验性价比，而影响观众对相声喜爱程度排在最后两位的分别是幽默和演员功底，即观众对于相声的喜爱程度和幽默本身无太大的关系，观众无论多大程度上喜欢相声都会认为幽默和演员功底很重要，而价格和体验性价比才是真正使消费者对于相声喜爱程度产生差异的原因。

（四）观众对相声付费意愿的综合评价

1.观众付费观看相声意愿的影响因素：整体体验和氛围感知影响程度最大

结构方程模型结果显示整体体验与氛围感知这两个影响因素对观众相声付费意愿的影响程度最大，标准化路径系数分别为 0.70 和 0.62；营销感知也对观众相声付费意愿有一定影响但影响程度较小。此外，幽默感与演员的功底是观众最为关注的两个因素，而消极因素中特别是语言地方性太强、现场服务考虑不周会严重打击观众对相声付费的意愿程度。

2.观众选择付费观看相声与否的决策行为：与喜爱、了解和满意程度有关，且随时间变化

喜爱程度的提升虽能带来相声付费意愿的大幅增长，但在超过 50%时就会难以形成持续的市场增速；后续宣传营销带来的效果对于观众选择的影响程度会呈现越来越小的状态；观众观看相声的频率超过每月 1-2 次、同时对观看付费相声的满意次数达到 2-3 次后，付费意愿显著提升。

3.观众对相声团体选择的影响因素：以内容创新与文化内涵为主导

AHP 层次分析法显示，内容创新与文化内涵对观众选择相声团体的影响程度最大，其权重分别为 0.2091、0.1699。其次，品牌影响力对观众选择相声团体也有较大影响，其权重为 0.1622，排序第三。另外，票价与营销推广对观众的选择偏好也具有一定影响，但影响相对较小。

4.付费观众对相声现场表演较为满意，少数指标亟待改善

基于云模型的付费观众对表演满意度评价显示，观众对于现场表演的相声幽

默有趣程度以及相声演员的功底比较满意，其云滴的期望分别为 4.0350、4.0073，处于较高水平。观众对于演员配合默契程度、现场互动、性价比、与当下热点结合等方面一般满意，其中演员配合默契程度与现场互动实际值与期望值差距较小，性价比与结合当下热点实际值与期望值存在一定差距。在文化内涵、创新、题材和环境舒适度方面，其实际值与期望值差距较大，未达到观众预期，仍需努力改进不足。

二、建议

（一）相声观众

1. 不断提高艺术审美

相声界关于“俗”的问题以及“俗”的尺度，并没有一个明确的界限，容易被演员曲解为“低俗”，在相声作品中利用敏感词汇来吸引眼球的现象较为普遍。而作为观众应不断提高自身艺术审美，拒绝低俗作品，免于被低俗作品所传递的价值观所误导。

2. 客观看待相声演员的表演，切勿带有“粉丝滤镜”

近年来，有些忠实粉丝在微博中为自己喜爱的相声演员建立“超话”，为其集资打榜，通过删除或举报对演员的负面评价来操纵网络评论，对演员们带有强烈的“粉丝滤镜”。显然，这十分不利于相声演员的进步与长足发展。相声观众应客观的看待相声演员在舞台上的表演，并对演出进行客观的评价，给予演员一定的进步空间。

3. 尊重相声表演者，支持相声等传统曲艺的发展

相声的发展与观众的支持密不可分，相声表演者在一定程度上也是中华优秀传统文化的传承者，因此观众应该尊重相声表演者的劳动成果。平时可以多多走进茶馆、剧院等地点观看相声，与此同时，付费观看相声应该通过官方渠道，自觉抵制黄牛票。

4. 不随波逐流，建立自己的欣赏体系

观众对相声的偏好会引导相声行业的发展趋势，现如今不少人盲目跟风，对相声没有自己的理解，这样不利于相声行业的进一步发展，因此观众在相声表演内容、相声演员的甄别上要树立自己的欣赏体系。

（二）相声行业

1.创新相声表演内容

相声观众逐渐趋于年轻化，因此相声行业应该与时俱进，深入了解当今相声观众喜欢的题材和表演形式，不局限于传统相声作品。同时要充分结合时事热点，表演内容要能够引起观众的共情，引发观众思考。

2.提高相声表演的文化内涵

现如今观众对于相声表演的要求并不只是获得快乐，更希望能够感受到中国传统艺术的魅力。因此相声表演内容应该抵制段子中庸俗、低俗的内容，保证相声的纯洁性，也不应该出现映射党和政府、宗教、残疾人等此类作品，发挥相声“雅俗共赏”的特质。

3.延续师徒口传心授的方式进行人才培养

相声观众普遍认为目前相声演员水平良莠不齐，新老演员水平差距过大，这对相声未来的发展将会是一个很大的限制。因此相声行业应该延续老一辈师徒口传心授的方式进行人才培养，浓厚的文化、艺术氛围可以刺激相声新人拥有不一样的艺术发展和思维方式；师傅同时也应该教授徒弟“做人在先，做艺在后”的道理，徒弟不仅要传承相声艺术，还应该对做人、做艺的品德进行传承。

4.适应当今互联网不断发展的新形势

观众认为观看相声的社会氛围会对其关于相声的付费意愿产生影响，同时也能够提高其对于相声的了解度。当今互联网+产业不断发展，微博、微信、抖音、哔哩哔哩网站等是国内具有代表性的文化社区，并且受众较大，因此相声行业应该充分利用这些文化社区对相声进行宣传，让大众感受到相声独特的艺术形式和幽默的内涵，争取能够实现“全民观看相声”的盛况，让相声不再是郭德纲口中所说的“虚假繁荣”。

（三）相声表演团体

1.发挥品牌影响力

在观众关于相声团体选择的问题中，品牌影响力是比较重要的一项。如今德云社、嘻哈包袱铺、青曲社、大逗相声、星夜相声会馆、听云轩为影响力较大的几大相声表演团体，这些团体应该顺势发展，在每一次演出中总结经验，争取给

观众带来更好的观感；对于普通的相声团体，也应该认真对待每一次演出，不断提高自己的品牌影响力，从而吸引更多的相声观众。

2.考虑专场表演的现场环境舒适度

大多数已付费观看过相声表演的观众对现场环境的舒适度不太满意，并且广大相声观众也认为现场服务考虑不周会严重打击对相声付费的意愿程度。因此，各相声表演团体在保证表演质量的同时，应该着重提高环境舒适度。在表演前要对场馆进行充分的考察，做好卫生清洁等服务，在表演过程中也要向观众提供最佳的观看体验。

3.磨练相声演员的表演功底

观众无论多大程度上喜欢相声都会认为演员功底很重要，然而任何一个优秀的相声演员都需要付出无数的努力，因此相声演员在跟随师傅学艺的过程中，也应该不断培养自己对待艺术的感觉，创造属于自己的特点。因为无论在何种情况，相声演员的基本功都是最重要的资质。

4.把给观众带来欢乐作为持续奋斗的目标

一提到喜爱的相声演员的特征，“幽默”、“风趣”、“喜感”、“包袱”成为相声观众口中最常出现的词汇，同时，观众认为相声演员夸张的动作、表情，多才多艺的表演形式和敬业的态度在表演过程中都是非常重要的。因此相声表演团体应该精准提炼笑点，充分发挥相声的喜剧效果，不断地给观众带来欢乐。

参考文献

- [1]王汝冰.传统相声商业化发展的现状及建议[J].视听,2020(02):142-145.
- [2]郭静欣,刘文华.数字化媒体背景下相声曲艺创新发展研究[J].中外企业家,2020(06):239.
- [3]薛丽婷.相声文化对现代的意义.<https://wenku.baidu.com/view/c4d2a581680203d8ce2f2438.html>.2012年12月7日发布.
- [4]郭瑶,方金.农村基本公共服务农户满意度影响因素研究——基于 Logistic-ISM 模型[J].科技和产业,2019,19(07):100-107.
- [5]成娅辉. k-means 算法改进及其在通信行业客户细分中的应用[D].湖南大学,2010.
- [6]李伟,胡云飞,李澎林.基于多视角二分 k-means 的高校图书馆用户画像研究[J].浙江工业大学学报,2020,48(02):141-147.
- [7]李家樑.基于贝叶斯决策和大数据的雾霾预测系统研究[J].通讯世界,2019,26(07):128-129.
- [8]李欣海.随机森林模型在分类与回归分析中的应用[J].应用昆虫学报,2013,50(04):1190-1197.
- [9]侯杰泰,成子娟.结构方程模型的应用及分析策略[J].心理学探新,1999(01):54-59.
- [10]郭学品,罗自强,石春.基于云模型的网购满意度综合评价[J].统计与决策,2018,34(23):60-62.
- [11]宗利永,林雨萧,陈群.基于文本挖掘的在线电影短评情感分析——以《哪吒之魔童降世》为例[J].现代电影技术,2020(05):42-48.
- [12]刘葛泓,李金泽,李卞婷,邵南青,窦万峰.基于 Text-CNN 联合分类与匹配的合同法律智能问答系统研究[J].软件工程,2020,23(06):8-12+4.

附录1 调查问卷

茶馆“相”遇妙，沧海一“声”笑 ——相声观众偏好分析与付费意愿调查

您好！我们是_____的学生，正在参加“全国大学生市场调查与分析大赛”，希望能够了解到目前相声的观众偏好以及市场化发展前景的情况，特进行此次调查。请放心，我们会对您提供的所有信息进行严格保密，所有数据仅用于学术研究，不记名。您只需根据自己的实际情况填写本调查问卷，请选择符合您情况的选项。衷心感谢您的支持与配合！

您是否观看过相声表演？ [单选题] *

- A.是（开始作答）
- B.否（结束）

第一部分 基本情况

Q1. 您的性别是： [单选题] *

- A.男
- B.女

Q2. 您的年龄是： [单选题] *

- A.20岁及以下
- B.21~30岁
- C.31~40岁
- D.41~50岁
- E.51岁以上

Q3. 您的职业类型为： [单选题] *

- A.在校学生
- B.政府/机关干部/公务员
- C.企业管理者（包括基层及中高层管理者）
- D.普通职员（办公室/写字楼工作人员）
- E.专业人员(如医生/律师/老师等)
- F.普通工人（如工厂工人）
- G.商业服务业职工
- H.个体经营商/承包商
- I.自由职业者
- J.退休
- K.其他

Q4. 您所在的地区为： [单选题] *

- A.东北（黑龙江省、吉林省、辽宁省）
- B.华东（上海市、江苏省、浙江省、安徽省、福建省、江西省、山东省、台湾省）
- C.华北（北京市、天津市、山西省、河北省、内蒙古自治区）
- D.华中（河南省、湖北省、湖南省）

- E.华南（广东省、广西壮族自治区、海南省、香港特别行政区、澳门特别行政区）
 F.西南（四川省、贵州省、云南省、重庆市、西藏自治区）
 G.西北（陕西省、甘肃省、青海省、宁夏回族自治区、新疆维吾尔自治区）

Q5. 您的学历为： [单选题] *

- A.小学及以下
 B.初中
 C.高中/中专
 D.大学本科/专科
 E.研究生及以上

Q6. 您的月收入/生活费为： [单选题] *

- A.0~1500 元
 B.1500 元~3000 元以下
 C.3000~6000 元
 D.6000~9000 元
 E.9000~12000 元
 F.12000 元以上

Q7. 您对相声的喜爱程度？ [单选题] *

- A.非常喜欢
 B.挺喜欢的
 C.一般
 D.不喜欢

Q8. 您观看相声的频率大概为？ [单选题] *

- A.每天都看
 B.每周 1~2 次
 C.每月 1~2 次
 D.半年 1~2 次
 E.几乎不看

第二部分 观众对待相声的认知和偏好情况

Q9. 您主要通过什么途径观看相声？ [多选题] *

- A.春晚、《欢乐喜剧人》等电视节目
 B.手机视频软件（例如抖音，B 站）
 C.社交平台（例如微博，微信）
 D.剧场，剧院，茶馆等现场表演
 E.其他 _____ *

Q10. 程度题

	1	2	3	4	5
您对相声的了解程度：					
您对相声艺术的发展和					

传承的关心程度：					
您对目前相声发展的满意程度：					

Q11. 春晚时您大概愿意看多久的相声节目？ [单选题] *

- A.愿意看完所有相声的部分
- B.看绝大部分
- C.看一半
- D.看个开头
- E.完全不看

Q12. 您希望相声表演联系以下哪些内容？ [多选题] *

- A.中外影视作品
- B.生活趣事
- C.历史人物与事件
- D.新闻大事件
- E.年轻潮流文化
- F.歌颂时代英雄
- G.讽刺不良现象
- H.其他 _____ *

Q13. 您觉得相声的笑点在哪里？ [多选题] *

- A.口音
- B.动作
- C.掺插在其中的小包袱
- D.主题升华
- E.演员喜感
- F.现场互动
- G.从未发现过的笑点
- H.其他 _____ *

Q14. 您对以下哪些题材的内容感兴趣？ [多选题] *

- A.国家发展
- B.校园趣事
- C.历史知识
- D.传奇轶事
- E.工作生活
- F.江湖野史
- G.长篇故事连载
- H.其他 _____ *

Q15. 您喜欢下列哪些相声演员？ [多选题] *

- A.马三立
- B.侯宝林
- C.刘宝瑞

- D.马季
E.郭德纲
F.于谦
G.岳云鹏
H.孙越
I.冯巩
J.相声新生代（郭麒麟，卢鑫玉浩等）
K.大兵
L.高晓攀
M.曹云金
N.其他 _____ *

Q16. 用几个词形容自己喜欢的相声演员的特点 [填空题] *

第三部分 观众对待相声的行为特征调查

Q17. 请您对下列影响相声观看的因素的重要程度进行打分：[矩阵量表题] *

	非常不重要	不太重要	一般	比较重要	非常重要
内容创新					
营销推广					
价格					
品牌影响力					
体验性价比					
社会氛围					
政府支持力度					
观看场所					
幽默					
演员功底					
剧本					
文化内涵					
相声形式（单口、对口、群口等）					
表演形式（文哏、文哏、柳活、腿子活、贯口活）					

Q18. 您认为以下哪些现象会给相声界带来较大的不良影响？ [多选题] *

- A.过于依赖偶像经济
B.相声演员水平良莠不齐
C.主流媒体的批判

D.相声老粉的退圈

E.界内过于浮躁

Q19. 您能接受的一场相声演出的最高价格范围是？ [单选题] *

A.0-100 元

B.100-200 元

C.200-300 元

D.300-500 元

E.500-1000 元

F.1000 元以上

Q20. 您觉得一段相声最适合的时长是多少？ [单选题] *

A.10 分钟以下

B.10-25 分钟

C.25-40 分钟

D.40 分钟以上

Q21. 您认为相声在发展过程中存在的问题有： [多选题] *

A.传承方式有问题，导致新老演员水平差距过大

B.不能带来可观的利益，不适应市场经济

C.生活节奏加快，人们没有时间了解和接触相声

D.没有与时代挂钩，缺乏创新，新作品匮乏且低级无聊

E.如今大众娱乐生活丰富，相声的替代品太多

F.其他 _____ *

Q22. 您认为观看相声给您带来的最显著的影响为？ [单选题] *

A.心情更加愉悦

B.减压

C.增长见识

D.心态更加乐观

Q23. 您是否付费观看过相声表演？ [单选题] *

A.是（开始第四部分作答）

B.否（跳转至第五部分）

第四部分 已付费观看相声的观众态度调查

Q24. 以下各因素对您不再付费观看相声的影响程度是： [矩阵量表题] *

	非常不重要	不太重要	一般	比较重要	非常重要
没有产生共鸣					
语言地方性太强					
表演形式过于单一					
内容无趣					
新时代娱乐方式的冲击					
身边听相声的氛围不浓					

现场服务考虑不周					
----------	--	--	--	--	--

Q25. 您喜欢的相声团体包括： [多选题] *

A.德云社

B.嘻哈包袱铺

C.青曲社

D.大逗相声

E.星夜相声会馆

F.听云轩

G.其他 _____ *

Q26. 以下各因素对您喜爱相声团体的影响程度是： [矩阵量表题] *

	非常不重要	不太重要	一般	比较重要	非常重要
内容创新					
营销推广					
票价					
品牌影响力					
演员					
剧本					
文化内涵					

Q27. 更喜欢哪种座位形式？

[输入 0(圆桌茶馆式)到 100(剧院式)的数字]*

Q28. 请您对付费观看相声表演前后各项指标的期望值和实际值打分（1-5 分）：

[矩阵量表题] *

	1	2	3	4	5
演员配合默契度期望值					
演员配合默契度实际值					
演员功底期望值					
演员功底实际值					
现场互动期望值					
现场互动实际值					
文化内涵期望值					
文化内涵实际值					
题材期望值					
题材实际值					
创新期望值					
创新实际值					

与当下热点结合期望值					
与当下热点结合实际值					
幽默有趣期望值					
幽默有趣实际值					
性价比期望值					
性价比实际值					
环境舒适度期望值					
环境舒适度实际值					

Q29. 您认为专场相声表演有什么亟待解决的问题？ [多选题] *

- A. 剧场相声底线过低
- B. 人才管理制度问题
- C. 相声段子创新不足
- D. 观众刨活搭茬严重
- E. 现场服务考虑不周
- F. 部分相声团体商演门票一票难求

第五部分 观众对相声发展的前景展望

Q30. 您认为我们该怎样发展相声？ [多选题] *

- A. 将传统技巧和时代元素相结合
- B. 抵制相声段子中的庸俗、低俗的内容，保证相声的纯洁性
- C. 与其他产业（如旅游业）进行商业性结合以取得利润及拓宽生存空间
- D. 政府大力支持，进行公益演出
- E. 利用新型媒体加大对相声的宣传力度
- F. 加快相声剧本的更新速度，鼓励原创
- G. 其他 _____ *

Q31. 对于粉丝自发的集资、打榜、控评、举灯牌等行为，您如何看待？ [单选题] *

- A. 赞同，能让更多人关注到相声，是好事
- B. 中立，没什么影响
- C. 不赞同，传统曲艺本应很纯粹

Q32. 您认为相声是否可以走出国门？ [单选题] *

- A. 可以，以其独特的艺术形式和幽默的内核
- B. 不可以，由于语言文化的差异观众可能 get 不到其中的笑点
- C. 不清楚

Q33. 您对相声等传统曲艺未来发展前景持何态度？ [单选题] *

- A. 积极态度，相声得以传承，越来越多的人愿意了解传统曲艺文化
- B. 中立，走一步看一步
- C. 消极态度，认为“相声是虚假繁荣”

附录2 部分访谈记录

嘉宾介绍：我们联系到黑龙江省曲艺团的相声演员李大宇先生，进行了关于当前相声市场的深入访谈。李大宇，黑龙江省曲艺团相声演员。十多年来，李大宇先生活跃在各大舞台上，具有《超级英雄》、《学评书》等作品，给观众带来了许多欢乐。同时热心于相声传承事业，甚至创立哈尔滨市大学生相声联盟，致力于在高校范围内促进相声的发展。

问：您认为现在越来越多的年轻人逐渐了解并喜爱相声的原因有哪些？

答：首先现在年轻人喜爱相声的前提是喜欢喜剧，喜剧效果是相声的表达形式。其次，相声是与时俱进的语言艺术，随着时代发展，内容不断创新，这是其他艺术形式所不能比拟的。

问：您觉得好的相声是什么样的，怎样定义这种好？

答：我认为好的相声要能够代表广大人民群众的内心里，受广大人民群众喜欢，且能反映相声的特性。应该是能为时代流传下去的，能歌颂、讽刺一些事物，映射时代现象的相声；不仅仅为博得观众的“乐”，也能带给人思考，引人发省的作品才是好相声。此外，好相声要具有共鸣感，从实际生活中切入，内容要与观众产生共情。

问：您认为目前相声发展的困境有哪些？

答：相声发展总体来讲与其他艺术形式相比较是蒸蒸日上的，发展势头较好。困境主要集中在相声从业者与观众对“好相声”的认知存在分歧，好相声、好包袱在行业内的认知也不尽相同。总之，相声从业者内部认知分歧与相声从业者与观众对“好相声”认知的分歧是当前面临的主要困境。

问：黑龙江省曲艺团表演的相声题材大概是什么样的？

答：我们表演的题材主要是传统作品、在传统作品上加以改进的继承性作品以及独立创作的全新作品。曲艺团在选取题材时会结合时代发展背景，例如映射党和政府、宗教、残疾人等此类作品，我们会将其予以剔除。对传统作品则取其精华去其糟粕，我们力求观众能喜爱作品并引发深思。同时，为了让年轻人更容易接受作品的包袱，我们将传统经典人物与现代人物相结合，并引导年轻人树立正确的价值观。

问：您对目前相声粉丝自发的集资、打榜、举灯牌等应援行为如何看待？

答：我认为粉丝自发行为，属于一种正常的社会现象。如同年轻人会喜欢追星、听偶像演唱会等，我们要合理看待这种现象。只要符合法律法规、不对他人造成伤害，我们就对此行为持积极态度。

问：您是怎样成为一名相声演员的？

答：我是天津人，从小耳濡目染，受爷爷喜爱相声的影响。小学开始就学习、表演相声；在高中时，报名相声补习班进行系统的训练；在大学阶段，我常常思考为什么相声在大学社团不火，于是联系周边学校社团成立哈尔滨市大学生相声联盟。我起初是笼络了六所高校，然后直到现在发展到了12所高校，几乎遍布了哈尔滨市所有拥有相声社的大学，我们联系到了哈尔滨市曲协、黑龙江省曲团的专业老师给我们提供帮助，帮助包括：免费的培训、助演等。我们所需要的一切关于相声的帮助，他们都给我们提供无条件的支持。然后也是通过这个机会，我才和黑龙江省曲艺团的各位老师，包括我们的团长刘彤老师有过一些联系。因为我本身也有些基础，再加上大学几年的培训、练习和演出，慢慢地有了那么一些资质，所以我毕业后考入了黑龙江省曲艺团从事专业。值得一提的是，近五年来哈尔滨市大学生相声联盟发展十分迅猛，在全国的大学生相声领域处于一个领跑阶段。

问：在人才培养方面，您觉得师徒这种口传心授的方式还能适应现代文化市场竞争的节奏吗？

答：我觉得是必要且一定要坚持下去的。首先，像很多人说的师傅带徒弟的口传心授，我解释这么一个：所有的艺术行业不是能够写下来的教材，所有能写下来的教材都不是能体现的，所有能体现的教材适用于物理、化学这些科目，但是，所有的艺术门类、创作性的门类都不可能以课本形式去教授。师傅带徒弟式的口传心授，他教授的是对待艺术的感觉，这个感觉是非常奇妙的东西，你需要长期跟在你师傅身边，跟在你周围的老师生活在一个圈子里，长期的文化氛围、艺术氛围可以刺激你的思维让你拥有不一样的艺术发展、思维方式，然后在这种方式中，如果你能够创造属于你自己的东西，那么你就能够成为下一个艺术家，这是艺术的传承方式，所以说，上课这种事情对艺术来说没有意义，它只是打基础，是在整个基础生涯中的第一步。

第二个，我认为最重要的，师傅带徒弟做人在先，做艺在后，拜名师不是拜有名的师傅而是拜一个德行、品行、名望都优的师傅。师傅教你的是在这行做人做艺应该怎么去做，我们相声行业的所有人都知道相声行业的师徒如父子，你的“师父”是父亲的父，他所教授你的是他的为人处世，是他的世界观、价值观。所以我认为时代无论到任何一个地步永远不可能摒弃师徒的这种传承，它所传承的东西是不可比拟的，如果把这种师徒的传承，做人、做艺品德的传承泯灭的话，那将是一件非常可怕的事情。

附录3 代码

1.随机森林

```
import numpy as np

import pandas as pd

from sklearn.ensemble import RandomForestClassifier

url1 = pd.read_csv(r'2.txt', header=None)

url1.columns = ['相声的喜爱程度', '内容创新', '营销推广', '价格',
                '品牌影响力', '体验性价比', '社会氛围',
                '政府支持力度', '观看场所', '幽默',
                '演员功底', '剧本', '文化内涵', '相声形式', '表演形式']

# print(url1)

# 查看几个标签
Class_label = np.unique(url1['相声的喜爱程度'])

print(Class_label)

# 查看数据信息
info_url = url1.info()

print(info_url)


from sklearn.model_selection import train_test_split

print(type(url1))

# url1 = url1.values

# x = url1[:,0]

# y = url1[:,1:]

x, y = url1.iloc[:, 1:].values, url1.iloc[:, 0].values

x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.3, random_state=0)

feat_labels = url1.columns[1:]

# n_estimators: 森林中树的数量

# n_jobs 整数 可选（默认=1） 适合和预测并行运行的作业数，如果为-1，则将作业数设置
为核心数
```

```

forest = RandomForestClassifier(n_estimators=10000,random_state=0,
n_jobs=-1,criterion='entropy')

forest.fit(x_train, y_train)

# 下面对训练好的随机森林，完成重要性评估

# feature_importances_ 可以调取关于特征重要程度

importances = forest.feature_importances_

print("重要性：", importances)

x_columns = url1.columns[1:]

indices = np.argsort(importances)[::-1]

for f in range(x_train.shape[1]):

    # 对于最后需要逆序排序，我认为是做了类似决策树回溯的取值，从叶子收敛

    # 到根，根部重要程度高于叶子。

    print("%2d) %-*s %f" % (f + 1, 30, feat_labels[indices[f]], importances[indices[f]]))

# 筛选变量（选择重要性比较高的变量）

threshold = 0.15

x_selected = x_train[:, importances > threshold]

# 可视化

import matplotlib.pyplot as plt

plt.figure(figsize=(10, 6))

plt.title("相声喜爱程度中各个特征的重要程度", fontsize=18)

plt.ylabel("import level", fontsize=15, rotation=90)

plt.rcParams['font.sans-serif'] = ["SimHei"]

plt.rcParams['axes.unicode_minus'] = False

for i in range(x_columns.shape[0]):

    plt.bar(i, importances[indices[i]], color='orange', align='center')

    plt.xticks(np.arange(x_columns.shape[0]), x_columns, rotation=90, fontsize=15)

plt.show()

2.贝叶斯决策

import math

```

```

import time

import numpy as np

from collections import Counter

from sklearn import preprocessing

from sklearn.model_selection import train_test_split


def calcProbDensity(meanLabel, stdLabel, test_X):

    numAttributes = len(test_X)

    MultiProbDensity = 1.0

    # print "this is calcPD"

    # print meanLabel

    # print stdLabel

    for i in range(numAttributes):

        MultiProbDensity *= np.exp(-np.square(test_X[i] - meanLabel[i]) / (2.0 *
np.square(stdLabel[i]))) / (

            np.sqrt(2.0 * math.pi) * stdLabel[i])

    #print MultiProbDensity

    return MultiProbDensity


def calcPriorProb(Y_train):

    i, j = 0, 0

    global labelValue, classNum

    numSamples = Y_train.shape[0]

    labelValue = np.zeros((numSamples, 1))

    # print "laV:",labelValue

    Y_train_counter = sum(Y_train.tolist(), [])

    cnt = Counter(Y_train_counter)

    # print "cnt",cnt

    for key in cnt:

        labelValue[i] = key

        i += 1

    classNum = i

```

```

# print "laV2:",labelValue

# print classNum

Pc = np.zeros((classNum, 1))

eachLabelNum = np.zeros((classNum, 1))

for key in cnt:

    Pc[j] = float(cnt[key]) / numSamples

    eachLabelNum[j] = cnt[key]

    j += 1

return labelValue, eachLabelNum, classNum, Pc


def trainBayes(X_train, Y_train):

    startTime = time.time()

    numTrainSamples, numAttributes = X_train.shape

    print "trainBayes"

    print numAttributes,numTrainSamples

    labelValue, eachLabelNum, classNum, Pc = calcPriorProb(Y_train)

    meanlabelX, stdlabelX = [], []

    for i in range(classNum):

        k = 0

        labelXMaxtrix = np.zeros((int(eachLabelNum[i]), numAttributes))

        for j in range(numTrainSamples):

            if Y_train[j] == labelValue[i]:

                labelXMaxtrix[k] = X_train[j, :]

                k += 1

        meanlabelX.append(np.mean(labelXMaxtrix, axis=0).tolist())

        stdlabelX.append(np.std(labelXMaxtrix, ddof=1, axis=0).tolist())

    meanlabelX = np.array(meanlabelX).reshape(classNum, numAttributes)

    stdlabelX = np.array(stdlabelX).reshape(classNum, numAttributes)

    print meanlabelX

    print "###"

    print stdlabelX

    print('---Train completed.Took %f s.' % ((time.time() - startTime)))

```

```
return meanlabelX, stdlabelX, Pc
```

```
def predict(X_test, Y_test, meanlabelX, stdlabelX, Pc):

    numTestSamples = X_test.shape[0]

    matchCount = 0

    for m in range(X_test.shape[0]):

        x_test = X_test[m, :]

        pred = np.zeros((classNum, 1))

        #print "Pc:", Pc

        for i in range(classNum):

            pred[i] = calcProbDensity(meanlabelX[i, :], stdlabelX[i, :], x_test) * Pc[i]

            print i

        print pred

        predict1 = labelValue[np.argmax(pred)]

        print predict1, "####", Y_test[m]

        if predict1 == Y_test[m]:

            matchCount += 1

        print "matchCount", matchCount, numTestSamples

    accuracy = float(matchCount) / numTestSamples

    print accuracy

    return accuracy
```

```
if __name__ == '__main__':

    print('Step 1.Loading data...')

    data = np.loadtxt("wine.txt", delimiter=',')

    print('---Loading completed.')

    x = data[:, 1:14]

    y = data[:, 0].reshape(690, 1)

    # print x

    # print "!!!!"

    # print y

    print('Step 2.Splitting and preprocessing data...')
```

```

X_train, X_test, Y_train, Y_test = train_test_split(x, y, test_size=0.4)

scaler = preprocessing.StandardScaler().fit(X_train)

X_train = scaler.transform(X_train)

X_test = scaler.transform(X_test)

print('---Splitting completed.\n---Number of training samples:%d\n---Number of testing
samples:%d' \

      % (X_train.shape[0], X_test.shape[0]))

print('Step 3.Training...')

meanlabelX, stdlabelX, Pc = trainBayes(X_train, Y_train)

print('Step 4.Testing...')

accuracy = predict(X_test, Y_test, meanlabelX, stdlabelX, Pc)

print('---Testing completed.Accuracy:%.3f%%' % (accuracy * 100))

```

3.AHP 层次分析法

1. %AHP 法计算过程
2. function[wi,w,lamda,CI,RI,CR]=Weight(A)
3. [wi,lamda]=eigs(A,1); %求最大特征值及对应的特征向量 [特征向量, 特征值]
4. n=size(A,1); %获取矩阵 c 的行数
5. CI = (lamda-n)/(n-1) ; % 计算一致性指标
6. R=[0 0 0.58 0.90 1.12 1.24 1.32 1.41 1.45];%平均随机一致性指标 RI
7. RI=R(1,n);%取该矩阵对应的 RI 值
8. CR = CI /RI ; % 计算一致性比例
9. w = wi/sum(wi) ; % 特征向量归一化
10. P=[1 1/2;2 1];
11. [wi_M,w_M,lamda_M,CI_M,RI_M,CR_M]=Weight(P);

4.云模型

1. %云模型参数计算过程
2. function [x,y,Ex,En,He]=cloud_compute(y,n)
3. Ex=real(mean(y));
4. En=real(mean(abs(y-Ex)).*sqrt(pi./2));
5. He=real(sqrt(var(y)-En.^2));
6. for q=1:n
7. Enn=real(randn(1).*He+En);

```

8.    x(q)=randn(1).*Enn+Ex;
9.    y(q)=exp(-(x(q)-Ex).^2/(2.*Enn.^2));
10.   end
11.   x=real(x);
12.   y=real(y);
13.   %云模型的主函数
14.   clc;
15.   clear all;
16.   close all;
17.   %生成云滴
18.   N=800;
19.   %数据如下
20.   Y1=importdata('sjzdata.txt') ;
21.   Y=Y1';
22.   EX=[];
23.   EN=[];
24.   HE=[];
25.   for i=1:size(Y,1)
26.       subplot(size(Y,1)/2,2,i)
27.       [x,y,Ex,En,He]=cloud_compute(Y(i,:),N);
28.       EX=[EX Ex];
29.       EN=[EN En];
30.       HE=[HE He];
31.       plot(x,y,'r');
32.       xlabel('满意度分数分布');
33.       ylabel('确定度');
34.       title(strcat('第',num2str(i),'个指标满意度云模型还原图谱'));
35.       axis([1,8,0,1]);
36.   end
37.   A=[EX' EN' HE'];
38.   %相声满意度综合评价云图
39.
A_qwz=[4.03211678832117,0.708357204360529,0.374795849384843;4.12262773722628,0.77695
9943564148,0.197910172204427;3.92116788321168,0.792708337519302,0.295993455471740;4.03357

```



```
664233577,0.813382111130716,0.258481461409446;4.07737226277372,0.763017166012247,0.287857
980188250;4.05401459854015,0.816950607454343,0.237229951304212;3.98102189781022,0.7056968
58298910,0.408455539349088;4.13284671532847,0.815508251155884,0.171057139815700;3.9795620
4379562,0.747824346335006,0.352917770313949;4,0.735521581315082,0.382285896082681];
```

```
40. k=sum(A_qwz(:,1));
```

```
41. A_qwz=A_qwz/k;
```

```
42.
```

```
P_sjz=[3.91970802919708,0.802484307986729,0.236014958402649;4.00729927007299,0.733785
411696543,0.364085489854941;3.83795620437956,0.831352802197611,0.154220645781892;3.881751
82481752,0.807393661461892,0.252278746225238;3.82335766423358,0.810935447483688,0.1520653
53093058;3.87153284671533,0.837960930868747,0.172718129285871;3.84817518248175,0.84137984
9502187,0.0936031812751204;4.03503649635037,0.755655806829671,0.347472886706031;3.7927007
2992701,0.843933354356513,0;3.84671532846715,0.802965093419543,0.161743505314815]';
```

```
43. X=P_sjz*A_qwz;
```

```
44. Y=A_qwz.^2*P_sjz.^2;
```

```
45. Ex=X(1,1);
```

```
46. En=(Y(1,2)+Y(2,1))^0.5;
```

```
47. He=(Y(1,3)+Y(3,1))^0.5;
```

```
48. W=[Ex En He];
```

```
49. % Ex=8.5744;
```

```
50. % En=0.0711;
```

```
51. % He=0.0268;
```

```
52. for q=1:10000
```

```
53. Enn=real(randn(1).*He+En);
```

```
54. x(q)=randn(1).*Enn+Ex;
```

```
55. y(q)=exp(-(x(q)-Ex).^2/(2.*Enn.^2));
```

```
56. end
```

```
57. x=real(x);
```

```
58. y=real(y);
```

```
59. plot(x,y,'r');
```

```
60. xlabel('满意度分数分布');
```

```
61. ylabel('确定度');
```

```
62. title(strcat('相声满意度综合评价云图'));
```

```
63. axis([0,8,0,1]);
```

附录 4 委托书

衡阳市雁峰区三缘文化艺术中心

市场调查项目委托书

甲方（委托方）：衡阳市雁峰区三缘文化艺术中心

乙方（受托方）：“研途璀璨”市场调研团队

为对目前国内相声市场进行全方位、多层次的了解，从而优化甲方在相声表演方面的内容设置和题材选择等，丰富表演形式，现委托乙方为甲方针对“国内相声观众的偏好选择”以及“观众付费观看相声的影响因素”等相关课题进行市场调查与分析，形成调研报告。

甲乙双方就调查的相关事宜，达成如下协议：

一、调查期限

乙方在本协议签署之日起 45 日内完成所有调查工作，并在调查完毕后 30 日内向甲方提供市场调研报告的纸质资料和电子版资料。

二、甲方的权利与义务

1. 甲方有权要求乙方提供调查的全部原始资料，以核对调查数据是否真实。
2. 甲方有权对乙方的不合格调查问卷提出补做要求。
3. 乙方独立调查有困难的，经甲方同意，可请甲方提供协助。

三、乙方的权利与义务

1. 乙方承诺对甲方的商业秘密负有保密义务。
2. 乙方应收集不少于 1000 份有效问卷。
3. 乙方应保证调查结果的真实性和准确性。

甲方负责人（签名或盖章）

衡阳市雁峰区三缘文化艺术中心



二〇二〇年一月十九日