

基于四川人寿理赔的 SPSS 数据分析报告

聂丹杨 杉

四川大学锦城学院 四川 成都 611731

【摘要】本文对四川人寿理赔数据采用 spss 分析工具进行分析,采用数据可视化方法、探索分析、比较均值、频率分析的方法具体针对客户画像、理赔年龄与保额及赔款金额之间的关系、理赔的出险原因以及保费与赔偿金额之间的关系作出了分析,得出理赔人数主要在 41-50 岁且呈年轻化趋势,以女性和已婚人士居多,年平均工资水平较低的人更易产生理赔;赔款金额与总保费呈正向相关,群众可以针对自身情况可以投入更多的保费保障自身利益;平均保额总体来说较低;理赔出险原因主要集中在疾病和意外,保险公司应该主要针对这两类推出更多的险种。通过此次分析个人可以根据自身具体情况选择理赔的方向,企业也可以跟进流量走势,推出更好的健康保障措施。

【关键词】保险;理赔;SPSS;数据分析

1 引言

保险作为如今社会中不可或缺的一部分,正在慢慢影响和改变着人们的生活。疾病、意外、死亡、衰老、残疾的发生随时威胁着每个人,保险为公民的安全和健康提供了确切的保障,帮助人们在遇到风险时能够有效保护财产安全。对于群众而言,选择正确的、适宜的保险是有效保障的前提;而对于保险公司来说,充分了解群众情况,拟定符合时宜的、受群众所关注的保险方案是保险业长期运行的保障。

在现如今经济环境背景条件下,居民收入水平与生活条件逐渐提高,居民对于健康以及其他各方面的安全保险意识也有了更多的提升,但对待保险理赔的认识还不够深入;而意外和疾病的发生是难以避免的,当面对突如其来的灾害时,受害者及其家庭需要一笔金钱,需要经济上的救助,以避免由于经济上的拮据给当事人造成更严重的二次伤害^[1],选择合适的理赔可以充分保障自身权益,化解危难,帮助家庭渡过难关,起了非常大的作用。

2 分析思路

鉴于保险理赔在经济发展上的需求不断上升,以四川人寿保险的理赔数据为例,用 spss 数据分析工具采用可视化分析、探索分析、频率分析的方法针对年龄与保额、赔款金额之间的关系、出险原因以及保费与赔偿金额之间的关系作出分析,绘制客户画像,分别对个人和企业提出了一些建设性意见,之于个人,让他们注重个人情况,及时有效的选择理赔,之于企业,了解客户和自身,针对客户提供更多健康保障措施。

3 数据说明

3.1 数据来源

数据来源于四川人寿保险公司,获取了该公司新投保、理赔、续保以及退保数据;通过筛选主要针对理赔数据进行分析,理赔数据具体包括机构、险种、案件号、保单号、赔款金额、赔款时间、出险原因、费用类型、费用金额、总保费、保额、客户号、性别、年龄、婚姻状况、过去三年平均年收入、职业共 16 列 21 万条

数据, 17M。

3.2 数据清洗

1 删除空白数据:查找是否有空白数据,单击查找,查找空白,没有空白数据,即剩下的数据都为有效数据。

2 将空格中有两类数据的进行分列。

3 删除无用数据列:将与本次数据分析无关且无用的列进行整列删除,分别是:赔款时间、客户号、职业,对剩余有用数据进行分析。

4 将年龄重新编码定义为 5 个年龄层,针对不同年龄段进行后续统计分析,重定义的年龄层分别是 0-20 岁, 21-40 岁, 41-50 岁, 51-60 岁, 61 岁及以上;将过去三年平均年收入进行分组定义为年均收入,分别为:5000 以下、5001-20000、20001-40000、40001-60000、60000 以上。

5 选择自定义排序,对险种、费用类型、年龄、赔款金额进行降序排列便于直接观察。

4 数据分析

4.1 数据可视化分析

通过可视化的方法对理赔数据人群的年龄、性别、婚姻状况、过去三年平均收入进行分析。

在 spss 软件中选择分析,点击描述统计中的频率,将分组后的年龄层、收入以及性别、婚姻状况放入变量中,可以得到绘制出的分布图如下:

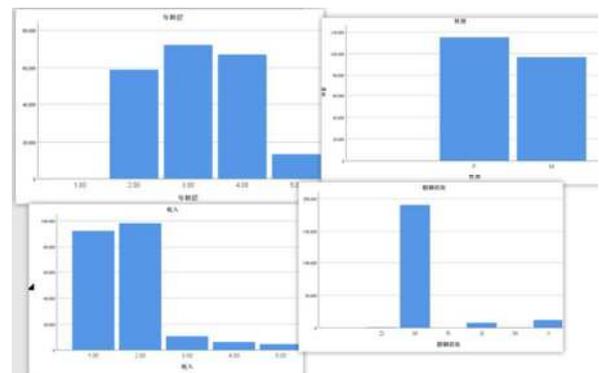


图 4.1 收入、性别、婚姻状况、不同年龄层人群分布图

由图可以得出,理赔人群年龄段在 41-50 岁中最多,

其次是51-60岁、21-40岁，这三个年龄段为主要构成；性别上女性理赔人数多于男性；产生理赔人群收入主要集中在5000以下及5001-20000之间，而后随着收入增高理赔人数逐渐减少；婚姻状况上已婚群众占比最大。

4.2 研究保额、赔款金额与年龄关系

通过比较均值来探究保额与年龄之间的关系。

在spss软件中选择分析，比较平均值，将保额和之前定义的年龄层放入变量，可以看出总体来说平均保额较低，年龄层在2，即20-40岁的人平均保额是最高的，年龄层在0-20岁的保额区别不大，年龄层在20-40岁之间的保额差异最大，且最为集中，呈右偏分布，右侧（低保额）极端值较多。

通过探索分析对赔付金额和年龄的关系进行探究。

在spss软件中点开分析，选择统计描述，再单击探索，将赔款金额放入因变量列表，年龄层放入因子列表，置信区间为95%，单击确定，生成描述统计表可以得出各年龄段数据偏度均大于0，为右偏（均值大于中位数大于众数），说明同组高赔付金额的人较少，大部分人的赔付金额低于平均值。年龄在61及以上数据峰度最大，为尖峰分布，说明样本之间差异较大，右边（高赔付金额区域）极端值较多，赔付金额更多的集中在众数周围。0-20岁的赔付金额平均值最高，可以说明21-40岁的人获得的赔付金额更多，其次是61岁及以上、21-40岁。

4.3 研究哪一类型出险原因最多

采用频率分析的方法分析统计出哪一类型的出险原因更多，点击分析里的描述统计，选择频率，将费用类型放入变量中。

统计出的出险原因一共有45种，其中以疾病、意外、呼吸疾病、消化疾病、泌尿生殖系病为前五类，分别占比为70.8%、15.3%、1.8%、1.4%、1.2%。

4.4 研究赔偿金额与总保费之间的关系

对赔偿金额与总保费的关系进行相关性分析，在spss软件中点击相关，选择双变量，将赔款金额与总保费放入变量当中，赔款金额为因变量，可以得到如下表格：

相关性		赔款金额	总保费
赔款金额	Pearson 相关性	1	.495**
	显著性 (双侧)	.000	.000
	N	212182	212182
总保费	Pearson 相关性	.495**	1
	显著性 (双侧)	.000	.000
	N	212182	212182

** 在 .01 水平 (双侧) 上显著相关。

模型	非标准化系数		t	Sig.	
	B	标准误差			
1	(常量)	2155.890	21.152	104.924	.000
	总保费	1.087	.004	262.249	.000

a. 因变量: 赔款金额

表 4.4 相关性、系数表

通过表格可以看出sig小于0.05是有相关且有效，呈正相关，从而可得出 $y=2155.890+1.087 \times$ 总保费。

5 结论及建议

5.1 结论

大数据工具可以快速方便针对相关主题进行分析得到结论，针对问题有更深刻认识及作出具体对策。通过大数据分析对四川人寿保险数据的分析能够了解理赔

数据的人物画像、趋势等相关信息，之于群众，能够帮助他们更深层次的认识理赔有关的知识，同时更好的针对自身情况选择是否购买保险；之于企业，能够充分了解自己的客户信息，对走势作出正确的判断，优化内部决策。保险业本身具有不稳定性且有大量不确定因素，对保险有更加全面的认识是保障自身权益更好的方式。

通过spss分析可视化方法得到理赔数据的人物画像：理赔人数年龄层前三的是41-50岁、51-60岁、21-40岁，61岁以上理赔人数较少是因为这个年龄段的人多没有购买相关保险；以已婚人士为主，多为女性，年收入主要在5000以下及5001-20000。

通过spss分析比较均值的方法得出总体来说保险需求高，但理赔的保额较低，20-40岁的人平均保额是最高最为集中但是组内保额的差异也是最大的；通过探索分析可以得出20岁以下的人获得的赔付金额最多，其次是21-40岁、61岁以上，但61岁以上的人群赔付金额差异较大，极端值较多，不稳定；

通过spss频率分析可以得出出险原因最多的为疾病和意外；

通过spss相关性分析得出赔偿金额与总保费呈正相关，总保费越高，所获得的赔偿金额越高， $y(\text{赔偿金额})=2155.890+1.087 \times$ 总保费。

5.2 建议

对于个人而言，针对自身情况投入保费从而获得与自己利益相关的理赔，选择性购买疾病和意外类这两种出险原因最多的保险；41-50岁的人群占理赔人数最多，处于上有老下有小的阶段，是家庭经济的主要来源，需要更加谨慎的选择机构、险种，可以适当投入更多的保费，从而在遇到疾病或意外时获得更多赔偿金额；理赔的年龄段有走低趋势，20-40岁理赔人群较大，这一年龄层大部分人群会面临学业、事业、家庭的问题，压力较大，建议年轻人即使保费预算有限，也可以从较短保障年限、较低保额的产品入手，再逐步为自己增加保障^[2]；女性，尤其是已婚女性的出现率显然高于男性，女性更应该注重自身权益，一定要为自身做好重大疾病保险类产品配置，有能力者应当配置终身型、高保额产品，即便预算不足，也要积极寻求重疾险的替代产品，比如价格低廉的防癌保险^[2]。

对于保险公司而言，保险公司应该适当增高保额，健康保障的需求是在不断升高的，大多数人都会考虑自己获得的保障是否够用的问题，加大保障，吸引和留住更多的客户；理赔数据趋近年年轻化，而21-40岁是主要客户，所以针对0-20、21-40这一年龄段的人群推出更多保障措施，多关注这一年龄层的用户；20岁以下理赔人数虽然占比不大，但是青少年存在一些特定疾病，更需要注重青少年疾病，并且这一年龄段赔付金额是最高的，向青少年家长普及更多的理赔知识，提醒家长为孩子购买特定的保障产品，吸引这部分客户；出险原因以疾病和意外居多，所以具体针对这两类领域多推出不同的险种供不同年龄段和收入段的人选择；考虑企业自身利益，可以在赔付金额高的21-40年龄层多收取一定的保费。总体来说，保险公司应该紧跟趋势，及时了解

群众动态,调整企业内部相关措施,结合大数据分析实施符合自身利益以及群众要求的保障措施,进一步加深大数据在理赔领域的应用。

【参考文献】

[1] 张媛. 针对人寿保险特性分析如何有效进行税收筹划 [J]. 中国集体经济. 2020, (12):22-31.

[2] 陈婷. 2020年保险理赔报告出炉 “新”重疾病险发展路在何方 [J]. 上海保险. 2021, (02):4-8.