

分析流程

数据源：
data_new__.csv

算法配置：
算法： 逻辑回归
变量： 变量 Y:{旅居养老意愿}； 变量 X:{我曾经在旅居地投资买房，我的年出游花费消费数额较大，我的子女都支持我选择旅居养老，我曾经有过旅居养老的经历，通过宣传，我对旅居养老有了充分的了解，旅居养老地所在的交通要方便，旅居养老机构的服务质量要完善，旅居养老服务机构所提供的旅游产品应有较高的性价比，旅居养老地的安全系数要高，旅居养老的自然风光要优美，年龄，性别，婚姻状况，教育程度，月收入状况（退休金、养老金等总和），每次旅居养老的预算范围，年出游次数(次)，次出游时长(天)，自理状况，我的月收入水平很高}收起

分析结果：
逻辑回归是用于解决因变量为二分类变量的回归：模型的似然比卡方检验的结果显示，显著性 P 值 0.000***，水平上呈现显著性，拒绝原假设，因而模型是有效的。

分析步骤

- 1. 对分类因变量分布状况进行描述。
- 2. 对模型进行似然比卡方检验，分析似然检比卡方显著性，若拒绝原假设(P<0.05)，说明模型有效，反之模型不成立，若设计多个模型，可以结合其他分类评价或者信息准则（AIC、BIC 值越小越好）进行综合分析。
- 3. 根据模型参数表，分析 X 是否呈显著性(P<0.05)，用于探究 X 对 Y 的影响关系。
- 4. 分析回归系数 B 与 OR 值（优势比），对比分析 X 对 Y 的影响程度。
- 5. 结合预测分类混淆矩阵与模型评价中的分类指标，分析模型预测。

详细结论

输出结果 1：二分类因变量基本汇总

因变量	选项	频数	百分比(%)
旅居养老意愿	1	235	85.766
	0	39	14.234
	总计	274	100

图表说明：

- 上表展示了因变量各分组的分布情况。
- 选项：当前字段数据下的去重类别
 - 频数：当前去重类别在数据中出现的次数
 - 百分比：当前去重类别的频数占比
 - 当因变量分类水平的数据量出现严重不平衡时，建议对数据进行过采样或者欠采样

输出结果 2：模型评价

似然比卡方值	P	AIC	BIC
152.609	0.000***	194.609	270.485

注：***、**、*分别代表 1%、5%、10%的显著性水平

图表说明：

上表展示了模型评价指标，可用于对模型的表现进行评估或有效性进行验证，其包括似然比检验、P 值、AIC 值、BIC 值。

- 对 P 值进行分析，如果该值小于 0.05，则说明模型有效；反之则说明模型无效。
- AIC 值和 BIC 值用于对比两个模型的优劣时使用，此两个值均为越小越好。

智能分析：

模型的似然比卡方检验的结果显示，显著性 P 值 0.000***，水平上呈现显著性，拒绝原假设，因而模型是有效的。

输出结果 3：二分类逻辑回归结果

项	回归系数	标准误差	Wald	P	OR	OR 值 95%置信区间	
						上限	下限
常数	-2.598	0.294	78.219	0.000***	0.074	0.042	0.132
我曾经在旅居地投资买房	0.25	0.4	0.39	0.532	1.284	0.586	2.814
我的年出游花费消费数额较大	0.533	0.364	2.146	0.143	1.704	0.835	3.474
我的子女都支持我选择旅居养老	-0.782	0.264	8.812	0.003***	0.457	0.273	0.767
我曾经有过旅居养老的经历	-0.252	0.305	0.682	0.409	0.778	0.428	1.413
通过宣传，我对旅居养老有了充分的了解	-0.501	0.291	2.967	0.085*	0.606	0.343	1.071
旅居养老地所在的交通要方便	0.011	0.252	0.002	0.965	1.011	0.617	1.657
旅居养老机构的服务质量要完善	-0.492	0.259	3.598	0.058*	0.611	0.368	1.017
旅居养老服务机构所提供的旅游产品应有较高的性价比	-0.139	0.307	0.206	0.650	0.87	0.477	1.588
旅居养老地的安全系数要高	-0.026	0.285	0.008	0.928	0.975	0.557	1.704
旅居养老的自然风光要优美	0.047	0.321	0.022	0.883	1.048	0.559	1.965
年龄	0.564	0.238	5.606	0.018**	1.757	1.102	2.801
性别	-0.061	0.235	0.068	0.795	0.941	0.594	1.49
婚姻状况	0.029	0.216	0.019	0.891	1.03	0.674	1.573
教育程度	-0.158	0.254	0.385	0.535	0.854	0.519	1.406
月收入状况（退休金、养老金等总和）	0.252	0.327	0.594	0.441	1.286	0.678	2.441
每次旅居养老的预算范围	-0.244	0.329	0.549	0.459	0.784	0.411	1.494
年出游次数(次)	0.122	0.274	0.197	0.657	1.129	0.66	1.931
次出游时长(天)	-0.268	0.301	0.797	0.372	0.765	0.424	1.378
自理状况	0.011	0.248	0.002	0.963	1.012	0.622	1.646
我的月收入水平很高	-0.496	0.347	2.049	0.152	0.609	0.308	1.201

因变量：旅居养老意愿

注：***、**、*分别代表 1%、5%、10%的显著性水平

图表说明：

上表展示了模型的参数结果。包括模型的系数、标准误差、OR 值、置信区间等用于分析模型的公式。

- OR 值（优势比）：为实验组的事件发生概率/对照组的事件发生概率。
- 对于连续自变量的 OR 值的意义为：该变量每升高一个单位，发生实验组事件的概率比发生对照组事件的概率变化了（OR 值-1）%。
- 对于哑变量化的 0-1 分类自变量的 OR 值意义为：该变量每升高一个单位（即分类水平从 0 变为 1），发生实验组事件的概率比发生对照组事件的概率变化了（OR 值-1）%。

智能分析：

二分类逻辑回归的结果显示：
 字段常数显著性 P 值为 0.000***，水平上呈现显著性，拒绝原假设，因此常数会对旅居养老意愿产生显著性影响，意味着常数每增加一个单位，旅居养老意愿为 0.0 的概率比 1.0 的概率低了

92.556%。

字段我曾经在旅居地投资买房显著性 P 值为 0.532，水平上不呈现显著性，不能拒绝原假设，因此我曾经在旅居地投资买房不会对旅居养老意愿产生显著性影响。

字段我的年出游花费消费数额较大显著性 P 值为 0.143，水平上不呈现显著性，不能拒绝原假设，因此我的年出游花费消费数额较大不会对旅居养老意愿产生显著性影响。

字段我的子女都支持我选择旅居养老显著性 P 值为 0.003***，水平上呈现显著性，拒绝原假设，因此我的子女都支持我选择旅居养老会对旅居养老意愿产生显著性影响，意味着我的子女都支持我选择旅居养老每增加一个单位，旅居养老意愿为 0.0 的概率比 1.0 的概率低了 54.264%。

字段我曾经有过旅居养老的经历显著性 P 值为 0.409，水平上不呈现显著性，不能拒绝原假设，因此我曾经有过旅居养老的经历不会对旅居养老意愿产生显著性影响。

字段通过宣传，我对旅居养老有了充分的了解显著性 P 值为 0.085*，水平上不呈现显著性，不能拒绝原假设，因此通过宣传，我对旅居养老有了充分的了解不会对旅居养老意愿产生显著性影响。

字段旅居养老地所在的交通要方便显著性 P 值为 0.965，水平上不呈现显著性，不能拒绝原假设，因此旅居养老地所在的交通要方便不会对旅居养老意愿产生显著性影响。

字段旅居养老机构的服务质量要完善显著性 P 值为 0.058*，水平上不呈现显著性，不能拒绝原假设，因此旅居养老机构的服务质量要完善不会对旅居养老意愿产生显著性影响。

字段旅居养老服务机构所提供的旅游产品应有较高的性价比显著性 P 值为 0.650，水平上不呈现显著性，不能拒绝原假设，因此旅居养老服务机构所提供的旅游产品应有较高的性价比不会对旅居养老意愿产生显著性影响。

字段旅居养老地的安全系数要高显著性 P 值为 0.928，水平上不呈现显著性，不能拒绝原假设，因此旅居养老地的安全系数要高不会对旅居养老意愿产生显著性影响。

字段旅居养老的自然风光要优美显著性 P 值为 0.883，水平上不呈现显著性，不能拒绝原假设，因此旅居养老的自然风光要优美不会对旅居养老意愿产生显著性影响。

字段年龄显著性 P 值为 0.018**，水平上呈现显著性，拒绝原假设，因此年龄会对旅居养老意愿产生显著性影响，意味着年龄每增加一个单位，旅居养老意愿为 0.0 的概率比 1.0 的概率高了 75.681%。

字段性别显著性 P 值为 0.795，水平上不呈现显著性，不能拒绝原假设，因此性别不会对旅居养老意愿产生显著性影响。

字段婚姻状况显著性 P 值为 0.891，水平上不呈现显著性，不能拒绝原假设，因此婚姻状况不会对旅居养老意愿产生显著性影响。

字段教育程度显著性 P 值为 0.535，水平上不呈现显著性，不能拒绝原假设，因此教育程度不会对旅居养老意愿产生显著性影响。

字段月收入状况（退休金、养老金等总和）显著性 P 值为 0.441，水平上不呈现显著性，不能拒绝原假设，因此月收入状况（退休金、养老金等总和）不会对旅居养老意愿产生显著性影响。

字段每次旅居养老的预算范围显著性 P 值为 0.459，水平上不呈现显著性，不能拒绝原假设，因此每次旅居养老的预算范围不会对旅居养老意愿产生显著性影响。

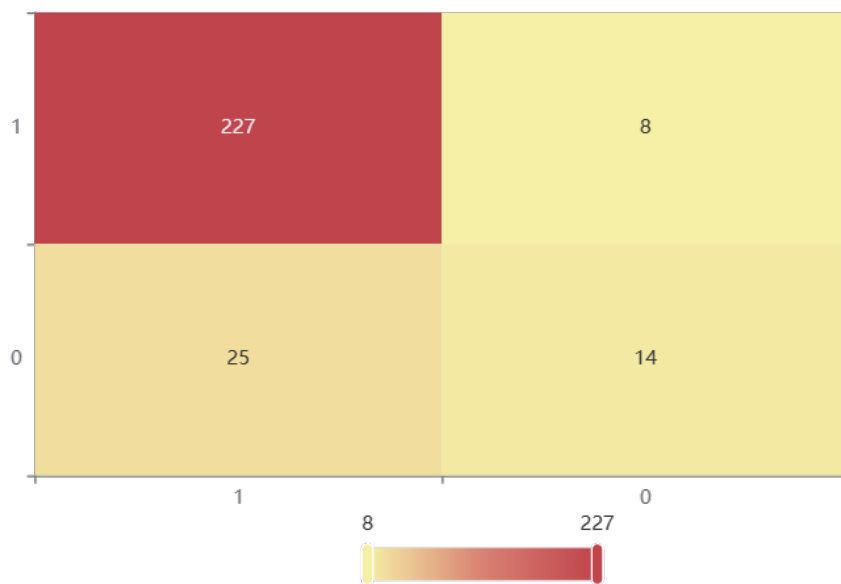
字段年出游次数(次)显著性 P 值为 0.657，水平上不呈现显著性，不能拒绝原假设，因此年出游次数(次)不会对旅居养老意愿产生显著性影响。

字段次出游时长(天)显著性 P 值为 0.372，水平上不呈现显著性，不能拒绝原假设，因此次出游时长(天)不会对旅居养老意愿产生显著性影响。

字段自理状况显著性 P 值为 0.963，水平上不呈现显著性，不能拒绝原假设，因此自理状况不会对旅居养老意愿产生显著性影响。

字段我的月收入水平很高显著性 P 值为 0.152，水平上不呈现显著性，不能拒绝原假设，因此我的月收入水平很高不会对旅居养老意愿产生显著性影响。

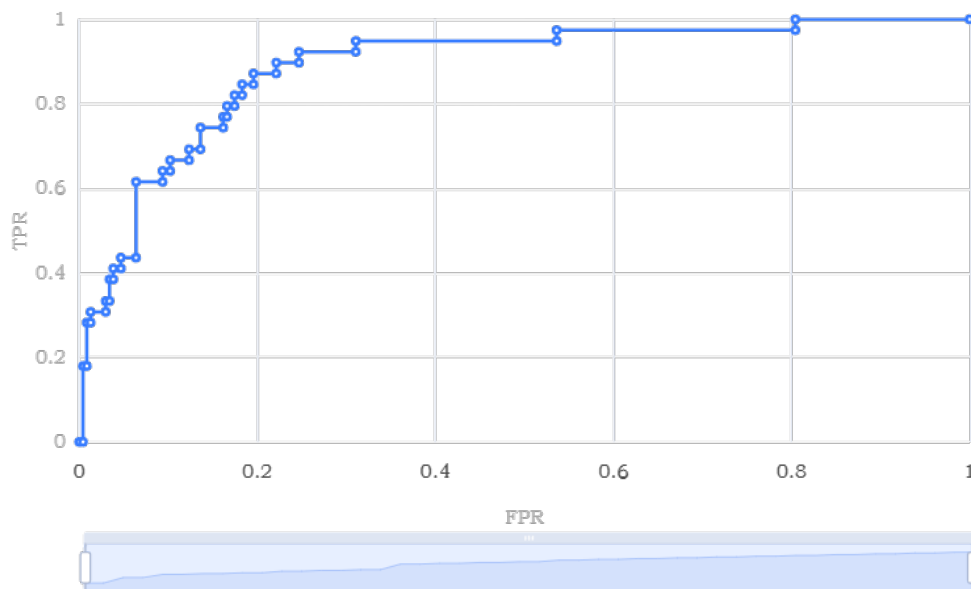
输出结果 4：混淆矩阵热力图



图表说明：

上表以热力图的形式展示了混淆矩阵。

输出结果 5：ROC 曲线



图表说明：

上图展示了 ROC 曲线图，用于衡量逻辑回归的分类效果。

- ROC 曲线图把灵敏度（TPR）和特异度（FPR）结合，可以同时衡量两者关系。理想情况下，TPR 应该接近 1，FPR 应该接近 0。
- 灵敏度：实际为正样本的结果中，预测为正样本的比例。

- 特异度：实际为负样本的结果中，预测为正样本的比例。

输出结果 6：分类评价指标

准确率	召回率	精确率	F1	AUC
0.88	0.88	0.863	0.865	0.89

图表说明：

上表中展示了分类评价指标，进一步通过量化指标来衡量逻辑回归的分类效果。

- 准确率：预测正确样本占总样本的比例，准确率越大越好。
- 召回率：实际为正样本的结果中，预测为正样本的比例，召回率越大越好。
- 精确率：预测出来为正样本的结果中，实际为正样本的比例，精确率越大越好。
- F1：精确率和召回率的调和平均，精确率和召回率是互相影响的，虽然两者都高是一种期望的理想情况，然而实际中常常是精确率高、召回率就低，或者召回率低、但精确率高。若需要兼顾两者，那么就可以用 F1 指标。
- AUC：AUC 值越接近 1 说明分类效果越好。

参考文献

- [1] Scientific Platform Serving for Statistics Professional 2021. SPSSPRO. (Version 1.0.11)[Online Application Software]. Retrieved from <https://www.spsspro.com>.
- [2] 袁翔宇,张蓬鹤,熊素琴,等. 基于逻辑回归算法的异常用电辨识方法研究[J]. 电测与仪表,2021,58(12):81-87.