**统计分析与建模分类模型研究报告**

1. 研究课题

泰坦尼克号的沉没是历史上最著名的沉船事故之⼀。1912年4⽉15⽇，在处⼥航中，被认为“永不沉没”的皇家邮轮泰坦尼克号与冰⼭相撞后沉没。不幸的是，船上没有⾜够的救⽣艇容纳所有⼈，导致2224名乘客和船员中的1502⼈死亡。虽然⽣存有⼀些运⽓因素，但似乎有些群体⽐其他群体更有可能⽣存下来。请使⽤乘客数据（包括姓名、年龄、性别、社会经济阶层等）建⽴⼀个预测模型，尝试思考和回答以下问题：“什么样的⼈更有可能存活下来？”。

1. 确定模型

对于这个问题，由于需要预测是否存活（Survived），是一个典型的二元分类问题。因此，可以使用各种二元分类模型来进行分析。可以使用广义线性回归中的逻辑回归模型。

1. 数据分析/结论（预处理前）

在对数据进行预处理前，可以先对原数据进行一系列分析，包括摘要分析