

# 基于判别分析与 Logistic 回归的股票价格涨跌预测

周世龄<sup>1</sup>, 范英兵<sup>2</sup>

(1.黑河学院, 黑龙江省黑河市 164300; 2.黑河学院, 黑龙江省黑河市 164300)

**摘要:** 本文深入探讨了股票市场预测的科学方法, 聚焦于 2022 年 9 月至 2024 年 9 月期间五粮液股票的历史交易数据, 运用判别分析与 Logistic 回归两种统计模型, 对股票价格的涨跌趋势进行预测分析。研究过程中, 首先对数据进行了细致的预处理, 确保分析的有效性。通过对两种模型的性能对比, 发现 Logistic 回归模型在预测准确性 (85%)、AUC 值 (0.9291) 及区分股票涨跌的混淆矩阵分析中表现更为优异, 显示出更强的预测能力与区分力。同时, 判别分析模型在交叉验证分数上虽略低, 但其在不同数据集上的稳定表现, 体现了较好的泛化能力。本文的研究成果为投资者提供了科学的决策支持, 同时对金融市场的理论研究和实践应用具有重要参考价值, 展现了统计学方法在金融市场分析中的巨大潜力。

**关键字:** 股票价格预测; 判别分析; Logistic 回归; 金融市场

**中图分类号:** F224

## Predicting Stock Price Trends Using Discriminant Analysis and Logistic Regression

Zhou Shiling<sup>1</sup>, Fan Yingbing<sup>2</sup>

(1.Heihe University, Heihe City, Heilongjiang Province 164300;

2.Heihe University, Heihe City, Heilongjiang Province 164300)

**Abstract:** This paper delves into the scientific methods of stock market prediction, focusing on the historical trading data of Wuliangye from September 2022 to September 2024. Two statistical models, discriminant analysis and logistic regression, are employed to predict the trend of stock price fluctuations. During the research process, the data was meticulously preprocessed to ensure the effectiveness of the analysis. By comparing the performance of the two models, it was found that the logistic regression model demonstrated superior predictive accuracy (85%), AUC value (0.9291), and discrimination power in the confusion matrix analysis for distinguishing between rising and falling stock prices, indicating stronger predictive capabilities and differentiation ability. Meanwhile, although the discriminant analysis model scored slightly lower in cross-validation, its stable performance across different datasets highlighted better generalization capability. The research findings provide investors with scientific decision-making support and offer significant reference value for both theoretical research and practical applications in financial markets, showcasing the great potential of statistical methods in financial market analysis.

**Keywords:** Stock price prediction; Discriminant analysis; Logistic regression; Financial market

## 1 绪论

### 1.1 研究背景

随着全球经济一体化进程的加速以及金融市场的快速发展, 股票市场作为资本市场的重要组成部分, 在促进资本流动、资源配置以及经济增长等方面发挥着至关重要的作用。然而, 股票市场的高度不确定性和复杂性使得投资者面临着巨大的挑战。因此, 开发有效的方法来预测股票价格变化趋势, 对于提升投资决策的质量和实现更高的投资回报具有重要意义。

近年来, 随着大数据技术和机器学习算法的发展, 越来越多的研究者开始尝试利用这些先进的技术手段来分析和预测股票市场。判别分析和 Logistic 回归分析作为经典的统计方法, 在处理分类问题上有着广泛的应用。尤其是对于二分类问题, 如股票价格的涨跌预测, 这两种方法因其直观性和解释性强的特点而备受青睐。

### 1.1 研究意义

本研究的意义在于通过应用判别分析和 Logistic 回归分析两种统计方法, 探索其在预测特定股票价格涨跌方面的有效性。一方面, 这有助于我们更好地理解不同统计模型在股票市场预测中的适用性和局限性; 另一方面, 研究成果可以为投资者提供科学的投资决策依据, 帮助他

基金项目: 无

作者简介: 周世龄 (2003-), 男, 四川宜宾人, 本科在读, 主要从事编程语言研究; 范英兵 (1984-), 男, 黑龙江黑河人, 副教授, 硕士, 主要从事金融数据处理与分析

通讯作者: 范英兵 (1984-), 男, 黑龙江黑河人, 副教授, 硕士, 主要从事金融数据处理与分析

们在复杂的市场环境中做出更为明智的选择。

### 1.3 研究目的

本研究将采用判别分析和 Logistic 回归分析两种统计方法。判别分析是一种用于分类的统计技术，它可以用来确定哪些变量最能区分不同类别的对象。而 Logistic 回归则是一种广泛应用于社会科学、医学研究等领域的统计方法，主要用于处理因变量为二元或多项分类的情况。在本研究中，将利用这两种方法来识别影响股票价格变动的关键因素，并构建相应的预测模型。通过比较两种模型的预测准确率、敏感性以及其他评价指标，可以得出哪种方法更适合用于预测股票的涨跌情况。

## 2 数据与方法

### 2.1 数据来源及描述

本研究中所使用的数据集来自于 英为财经 (<https://cn.investing.com>) 该网站提供了全面的金融信息,包括实时股票报价、历史价格数据等。所收集的数据涵盖了一只特定股票(代码 000858——五粮液)在 2022 年 9 月至 2024 年 10 月期间的日交易数据,包括每天的开盘价、收盘价、最高价、最低价、成交量以及涨跌幅等关键指标。选择此时间段是因为期间股票市场经历了多个周期性的波动,这为研究提供了丰富且多变的数据样本,有利于进行深度分析。

### 2.2 变量选择与定义

表 1 指标符号说明

符号	指标	指标含义
Open	开盘价	当日交易开始时的价格
Close	收盘价	当日交易结束时的价格
High	最高价	当日股票达到的最高价位
Low	最低价	当日股票触及的最低价位
Volume	成交量	当日的成交股数
Change	涨跌幅	相对于前一交易日收盘价的变化百分比
Rise/Fall	涨跌情况	表示股票当日是上涨还是下跌,用 1 或 0 表示

### 2.3 判别分析方法介绍

将每天的股票价格数据转化为涨跌标签,例如,如果收盘价相对于前一天有所增加,则标记为上涨组  $\pi_1$ , 否则标记为下跌组  $\pi_2$ 。假设  $n_1$  为上涨组样本个数,  $n_2$  为下跌组样本个数,  $n = n_1 + n_2$  为研究样本总数。每个观测值  $x_i^{(\alpha)}$  由开盘价、收盘价、最高价、最低价、成交量、涨跌幅六个向量元素组成。

设来自组  $\pi_\alpha$  的观测值为  $x_i^{(\alpha)}$ ,  $i = 1, 2, 3, \dots, n_\alpha$ ,  $n_\alpha = 1, 2$  将它们投影到某一个共同方向,得到的投影点是线性组合  $y_i^{(\alpha)} = c'x_i^{(\alpha)}$ , 其中  $c = (c_1, \dots, c_6)$  为六维常数向量,表示投影方向。

在确定了需使用的判别函数  $Z(x) = c_0 + c_1x_1 + c_2x_2 + c_3x_3 + c_4x_4 + c_5x_5 + c_6x_6$ , 其中  $c_1, c_2, c_3, c_4, c_5, c_6$  是判别函数系数后, 当将新样本  $X = (x_1, x_2, x_3, x_4, x_5, x_6)^T$  代入判别函数  $Z(x)$  中, 其判别规则应使用如下准则:  $x \in \pi_i$ , 若  $|Z(x) - \bar{Z}(i)| = \min_{1 \leq j \leq 2} |Z(x) - \bar{Z}(j)|$ , 其中  $\bar{Z}(1)$ ,  $\bar{Z}(2)$  分别为  $\pi_1, \pi_2$  组均值处的典型判别函数值。

### 2.4 Logistic 回归分析方法介绍

Logistic 回归分析是一种广泛应用于分类问题的统计方法,尤其适用于二分类问题。在本研究中,利用 Logistic 回归来预测股票价格的涨跌情况。具体而言,将每日股票的开盘价、收盘价、最高价、最低价、成交量、涨跌幅作为自变量,以股票价格是否上涨作为因变量来进行预测。

Logistic 回归模型基于逻辑函数,用来估计某一事件发生的概率。模型的基本形式如下:

$$P(Y = 1|x) = \frac{1}{1 + e^{-z}}$$

其中，  
 $Y$  是二分类的响应变量（0 或 1），在这里表示股票价格是否上涨（1）或下跌（0）； $z$  是一个线性组合的形式，由输入变量和回归系数组成：

$$z = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p$$

这里的  $b_0$  是截距项， $b_1, b_2, \dots, b_p$  是回归系数，对应于输入变量  $x_1, x_2, \dots, x_p$ ，分别是开盘价、收盘价、最高价、最低价、成交量和涨跌幅。

使用最大似然估计法来估计模型中的参数  $b_0, b_1, b_2, \dots, b_p$ ，模型可以写作：

$$\ln\left(\frac{P(Y=1|x)}{1-P(Y=1|x)}\right) = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p$$

这里，左边的部分称为对数几率，它是事件发生的几率与不发生的几率的比值的自然对数。

使用训练好的模型对未来数据进行预测。对于一个新的观察  $x$ ，计算  $P(Y=1|x)$ ，如果  $P(Y=1|x) > 0.5$ ，则预测股票会上涨，反之则预测会下跌。

## 2.5 模型评估标准

### 2.5.1 混淆矩阵

混淆矩阵是评估分类模型性能的重要工具之一，它可以直观地展示模型的预测结果与实际情况之间的关系。混淆矩阵包括以下四个基本指标：

表 2 四个基本指标及含义

指标	含义
真正例（True Positive, TP）	模型正确地预测了上涨的情况
假正例（False Positive, FP）	模型错误地预测为上涨，实际上下跌
真负例（True Negative, TN）	模型正确地预测了下跌的情况
假负例（False Negative, FN）	模型错误地预测为下跌，实际上上涨

基于混淆矩阵，可以得到如下指标：

表 3 混淆矩阵指标及公式

指标	含义	公式
准确率（Accuracy）	模型正确分类的比例	$\frac{TP + TN}{TP + TN + FP + FN}$
精确率（Precision）	模型预测为上涨时正确预测的比例	$\frac{TP}{TP + FP}$
召回率（Recall）	模型识别出的所有上涨情况中的比例	$\frac{TP}{TP + FN}$
特异性（Specificity）	模型识别出的所有下跌情况中的比例	$\frac{TN}{TN + FP}$

### 2.5.2 ROC 曲线与 AUC 值

ROC 曲线下面积（Area Under the Curve, AUC）是一个介于 0 到 1 之间的数值，用于量化分类器的性能。AUC 值越高，模型的分类性能越好。通常认为：

表 4 ROC 曲线性能评判表

区间范围	分类性能
AUC = 1	完美分类器
$0.9 \leq \text{AUC} < 1$	优秀分类器
$0.7 \leq \text{AUC} < 0.9$	良好分类器
$0.5 \leq \text{AUC} < 0.7$	一般分类器
AUC < 0.5	分类器不如随机猜测

2.5.3 交叉验证分数

每个折叠都会轮流作为测试集，其余部分作为训练集。这样可以多次评估模型的性能，并最终得到一个更可靠的性能估计。交叉验证的推演过程如图 1 所示：

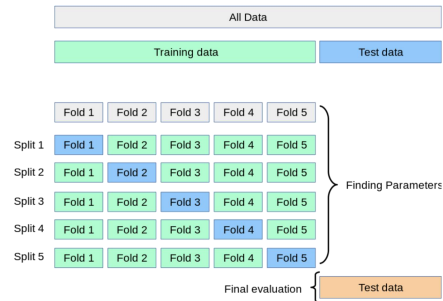


图 1 交叉验证推演图

3 实证分析

3.1 数据预处理

为了更好地理解数据的分布情况，进行描述性统计分析，以了解各变量的基本统计特征。以下是对股票代码为 000858 的历史日交易数据进行的描述性统计分析表。

表 5 数据描述统计表

	样本量	缺失	均值	标准差	均值 SE	均值的 95%置信 区间下限	均值的 95%置信 区间上限
收盘	485	0	156.93192	23.92064	1.08618	154.79771	159.06613
开盘	485	0	157.00951	24.01666	1.09054	154.86673	159.15228
高	485	0	158.97231	24.29401	1.10313	156.80479	161.13983
低	485	0	155.08619	23.55787	1.06971	152.98434	157.18803
交易量	485	0	17.83701	7.86157	0.35698	17.1356	18.53842

所有变量的样本量均为 485 条记录，且均无缺失值，表明数据完整性较好。收盘价、开盘价、最高价、最低价的均值和标准差非常接近，表明这些变量在统计意义上具有较高的相关性。交易量的均值为 17.83701 万手，标准差为 7.86157 万手，相较于价格变量，交易量的变异性较大。置信区间提供了估计均值的精度范围，对于收盘价和开盘价，置信区间的宽度较小，表明均值估计较为精确。

表 6 数据五数概括表

	最小值	Q1	中位数	Q3	最大值	Q3-Q1	极差	中位绝对 偏差	稳健变异 系数
收盘	107.7	139.1	155.1	171.4	217.6	32.3	109.9	16.4	0.16
开盘	107.7	138.9	155.0	172.1	218.6	33.2	110.9	16.4	0.16
高	109.6	140.3	156.6	174.0	219.9	33.7	110.3	16.7	0.16
低	106.3	136.7	153.6	169.4	214.3	32.7	108.0	16.0	0.15
交易量	7.0	12.6	15.9	20.8	54.6	8.2	47.6	3.9	0.36

通过表 6，可以解数据的分布情况。交易量的极差（47.6 万手）远大于其他变量，说明交易量的波动范围更大。中位数与均值相近，表明数据分布较为对称。中位绝对偏差（MAD）和稳健变异系数（RCV）提供了对数据离散程度的稳健估计。交易量的 RCV 值较高（0.36），说明交易量相对于其他变量而言更加不稳定。

为了更好地理解数据中的模式，绘制了各变量的箱线图和散点图。用于显示数据分散情况的数据分析图。它能够清晰地展示数据的中心趋势、分散性和异常值。

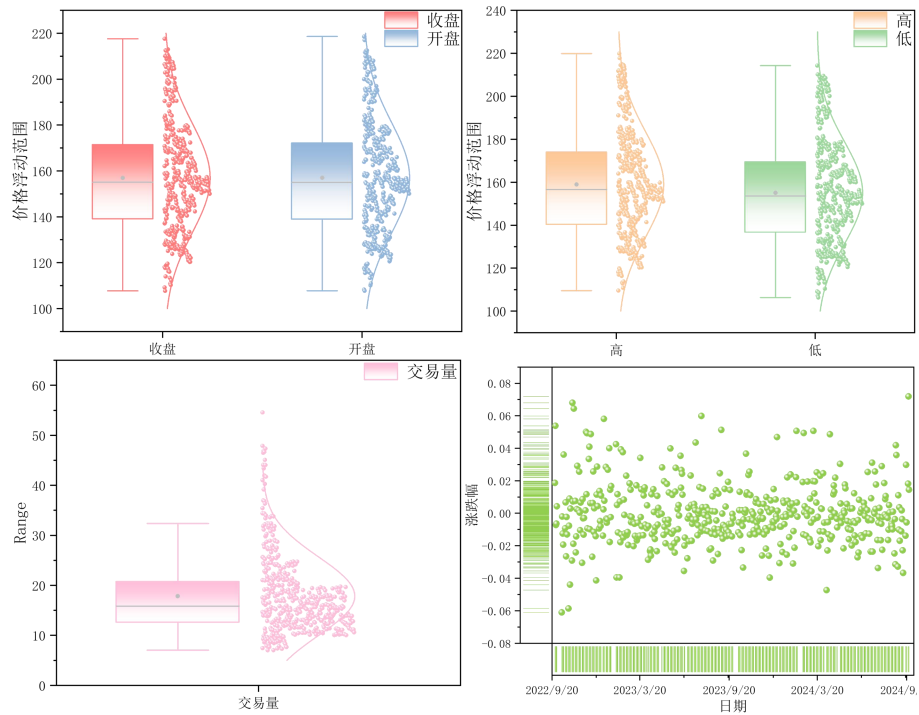


图 2 各变量的预处理图

通过对开盘价、收盘价、最高价、最低价及成交量的箱线图分析，可以看出大多数交易日的价格和成交量数据分布较为均匀，其中开盘价和收盘价的中位数位于箱体中部，显示出价格的稳定性，尽管开盘价存在少量市场突然波动导致的异常值，而收盘价则表现出较小的价格波动；最高价虽然其分布范围较广且有更多的异常值，反映出市场内的较大股价波动；最低价的中位数较低，且有少数极端低价情况，可能与市场特殊事件相关；成交量则体现了市场的活跃程度，中位数较高并且存在大量的异常值，特别是在高成交量的交易日；涨跌幅的散点图反映出重心位置接近于零，表明市场价格整体上波动不大。

为进一步揭示五粮液股票价格市场动态，绘制出图 3 的 2022 年-2024 年 9 月五粮液股票市场价格 K 线图。

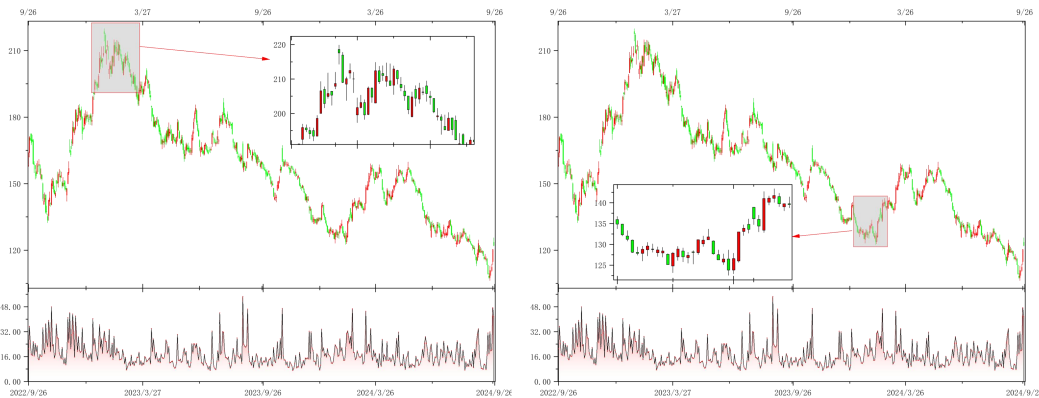


图 3 2022 年-2024 年 9 月五粮液股票市场价格 K 线图

阳线（红色）表示当日收盘价高于开盘价，表明股票上涨；阴线（绿色）则相反，表示股票下跌。从图中可以看出，阳线与阴线交替出现，反映了五粮液股票市场价格的波动。较长的上下影线表示在当日交易中股票价格经历了较大的波动，但最终收盘价与开盘价接近，说明买卖双方力量相当。

整体来看，价格呈现出一定的波动趋势，但总体保持在一个相对稳定的区间内。在某些交易日，价格出现了较为明显的上升或下降，这可能与市场中的重大新闻或经济数据发布有关。

大多数交易日的成交量集中在某个范围内，表明市场活跃度相对稳定。但在某些交易日，成交量出现显著增加。在价格大幅波动的日子里，往往伴随着成交量的激增，这表明市场情绪强烈，买卖双方的活动频繁。而在价格平稳的日子，成交量通常较低，表明市场参与者对该股票的关注度不高。

### 3.2 判别分析结果

为了分析股票代码为 000858 的历史日交易数据中的涨跌情况，使用了判别分析方法，并通过 *Python* 代码实现了模型的训练和检验。为了评估判别分析模型在区分股票涨跌情况时的有效性，如表 7 所示：

表 7 典型判别分析特征值表

	特征值	方差百分比	累积	典型相关
1	0.8407	100%	100%	0.67582

特征值为 0.8407，表示第一个典型判别函数解释了大量方差。典型相关系数为 0.67582，表示判别函数对组别的区分能力较强。

为了进一步评估模型的显著性，进行 Wilks' Lambda 检验。表 8 展示了 Wilks' Lambda 值以及显著性水平。

表 8 Wilks' Lambda 检验表

	Wilks' Lambda	卡方	df	Sig.
1 to 1	0.54327	241.31325	5	4.01395E-50

在 0.05 水平下，维度显著地大于 1。

Wilks'Lambda: Wilks'Lambda 值为 0.54327，表示模型在区分组别时具有显著性；卡方统计量为 241.31325，表明模型在区分组别时的统计显著性非常高；p 值小于 0.05，表明模型在 0.05 的显著性水平上拒绝原假设，也就是不同组的 Fisher 判别函数值存在显著差异，这就意味着判别函数是有效的。

由此得到非标准化 Fisher 判别系数。非标准化的 Fisher 判别系数是指将实测的样本观测值直接带入求出判别函数值，而不需要将观测值进行标准化。

表 9 非标准化判别函数系数

变量	函数	变量	函数
常量	0.64733	高	0.01287
收盘	-0.40095	低	-0.09895
开盘	0.48623	交易量	-0.0427

由表 9 可得，Fisher 判别函数为：

$$Z(x) = 0.64733 - 0.40095x_1 + 0.48623x_2 + 0.01287x_3 - 0.09895x_4 - 0.0427x_5$$

表 10 类中心处的 Fisher 判别函数值

组别	典型变量 1
1 (涨)	-1.04235
0 (跌)	0.80251

表 10 给出了类中心处的 Fisher 判别函数值，该函数值是根据为标准化 Fisher 判别函数计算的，股票价格涨的分布中心是 **-1.04235**，股票价格跌的分布中心是 **0.80251**。这样，只要根据 Fisher 判别函数计算出各观测数据的函数值之后，再比较他们分别距各类中心的距离，就可以把它们进行分类了。也就是说如果运用判别模型计算出来的判别值是大于 0 的，那么便可以把

观测样本归类为股票价格跌，判别值小于 0，则归类为股票价格涨。

### 3.3 Logistic 回归分析结果

为了评估 Logistic 回归模型的性能，计算模型的对数似然值以及两种不同的 R 方。表 11 展示了模型摘要的结果。

表 11 Logistic 模型摘要表

步骤	-2 对数似然	考克斯-斯奈尔 R 方	内戈尔科 R 方
1	303.187a	.523	.702

a. 由于参数估算值的变化不足 .001，因此估算在第 7 次迭代时终止。

该逻辑回归模型不仅具有良好的拟合效果（低-2 对数似然值），而且能够解释相当大比例的因变量变异（高 R 方值），这使得模型在预测股票市场涨跌情况时具备较高的可靠性和解释力。表 12 展示了模型中各变量的回归系数及其统计显著性。

表 12 Logistic 回归方程中的变量系数表

		B	标准误差	瓦尔德	自由度	显著性	Exp(B)
步骤 1a	Close	1.446	.192	56.523	1	.000	4.248
	Open	-1.534	.185	68.884	1	.000	.216
	High	-.203	.212	.925	1	.336	.816
	Low	.300	.235	1.623	1	.203	1.349
	Volume	.134	.035	14.674	1	.000	1.143
	常量	-2.937	1.138	6.661	1	.010	.053

a. 在步骤 1 输入的变量：Close, Open, High, Low, Volume。

根据表 3-8，可以得到 Logistic 回归方程为：

$$\text{Logit}(P) = -2.937 + 1.446x_1 - 1.534x_2 - 0.203x_3 + 0.300x_4 + 0.134x_5$$

通过方程和表中的内容，可以分析出收盘价和开盘价对股票涨跌情况具有显著的影响。

具体而言，收盘价的系数为 1.446，表明收盘价对股票涨跌情况有显著的正面影响；收盘价每增加一个单位，股票上涨的概率大约增加 4.248 倍。开盘价的系数为-1.534，表明开盘价对股票涨跌情况有显著的负面影响；开盘价每增加一个单位，股票上涨的概率大约降低至原来的 0.216 倍。然而，最高价和最低价对股票涨跌情况并没有显著的影响，最高价的系数为-0.203，交易量对股票涨跌情况也有显著的正面影响，系数为 0.134，交易量每增加一个单位，股票上涨的概率大约增加 1.143 倍。常量项为-2.937，这表明在所有自变量为 0 的情况下，股票上涨的概率较小。

综上所述，收盘价和开盘价以及交易量对股票涨跌情况具有显著的影响，而最高价和最低价则无显著影响，这些结果为理解股票市场涨跌提供了有价值的见解，并且模型中的显著变量可以作为预测未来股票走势的重要因素。

### 3.4 模型对比分析

在本研究中，分别使用了判别分析和逻辑回归两种不同的统计方法来预测股票的涨跌情况。通过对两种模型的性能进行对比分析，旨在探讨不同模型在预测准确性、稳定性以及解释能力方面的差异。

#### 3.4.1 性能对比

##### （1）准确率

准确率是指模型预测正确的次数占总预测次数的比例。准确率是衡量模型性能的一个基本指标，它直接反映了模型预测的准确性。

表 13 模型准确率对比表

模型	准确率
判别分析	83%
Logistic 回归	85%

根据表 13 的数据，逻辑回归模型在预测股票涨跌情况时的准确率为 85%，略高于判别分析模型的 83%。这表明逻辑回归模型在预测准确性上具有轻微的优势。然而，判别分析模型的准确率也达到了较高的水平，说明两种模型在预测股票涨跌情况上都有很好的表现。



## （2）AUC 值

AUC 值是评估分类模型性能的一个重要指标，特别是在二分类问题中。AUC 值反映了模型在不同阈值下区分正负样本的能力。较高的 AUC 值表示模型具有较强的区分能力，而较低的 AUC 值则表明模型的区分能力较弱。

表 14 模型 AUC 值对比表

模型	AUC 值
判别分析	0.8111
Logistic 回归	0.9291

表 14 展示了判别分析模型和逻辑回归模型在预测股票涨跌情况时的 AUC 值对比。从表中的数据可以看出，逻辑回归模型的 AUC 值显著高于判别分析模型。AUC 值越高，表示模型在区分正负样本上的能力越强。逻辑回归模型的 AUC 值为 0.9291，接近于 1，这表明模型在区分股票涨跌情况时具有很强的区分能力。相比之下，判别分析模型的 AUC 值为 0.8111，虽然也表明模型有一定的区分能力，但与逻辑回归模型相比仍有差距。

## （3）交叉验证分数

交叉验证是一种常用的技术，用于评估模型的泛化能力，它可以减少过拟合的风险，并提供对模型性能的更可靠估计。

表 15 交叉验证分数对比表

折叠次数（Fold）	Logistic 回归模型得分	判别分析模型得分
Fold 1	0.82352941	0.88235294
Fold 2	0.82352941	0.86764706
Fold 3	0.89705882	0.86764706
Fold 4	0.82352941	0.77941176
Fold 5	0.80597015	0.80597015
平均得分	0.83472344	0.84060579

从表 15 的数据可以看出，判别分析模型的平均交叉验证分数略高于逻辑回归模型，这表明在不同数据分割下，判别分析模型的泛化能力稍微好一些。然而，逻辑回归模型在不同轮次中的得分相对较为稳定，这表明逻辑回归模型更加稳健。

## （4）混淆矩阵

混淆矩阵是一种用于评估分类模型性能的表格形式，它提供了模型在测试数据集上的预测结果与真实结果之间的详细对比。混淆矩阵可以帮助了解模型在各个类别上的预测正确率和错误率，从而更好地评估模型的性能。

表 16 展示了两种模型的混淆矩阵、图 3-3 是两种模型的混淆矩阵热力图。

表 16 模型混淆矩阵对比表

模型	预测 实际	预测为涨	预测为跌
	实际为涨	78	5
判别分析	实际为跌	20	43
	实际为涨	80	10
逻辑回归模型	实际为跌	12	44



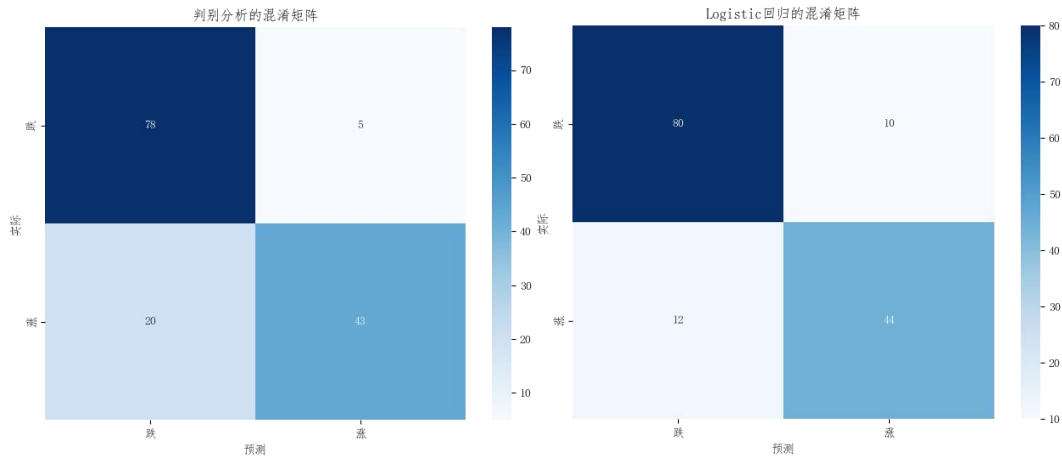


图 4 模型的混淆矩阵热力图

根据表 16 和图 4 的数据，可以得出以下结论：

针对判别分析模型在实际为涨的情况中，模型正确预测了 78 次，错误预测了 5 次。在实际为跌的情况中，模型正确预测了 43 次，错误预测了 20 次。

对于逻辑回归模型在实际为涨的情况中，模型正确预测了 80 次，错误预测了 10 次。在实际为跌的情况中，模型正确预测了 44 次，错误预测了 12 次。

通过对比两种模型的混淆矩阵及其图形表示，可以看到逻辑回归模型在预测股票上涨时的性能优于判别分析模型，并且在预测股票下跌时的错误也较少。

## 结论

随着经济全球化的推进以及金融市场的发展，股票市场作为资本市场的核心部分，在经济发展中扮演着越来越重要的角色。对于投资者而言，准确地预测股票价格的变化趋势至关重要，因为这直接影响到投资决策的有效性和投资收益的高低。然而，股票市场充满了不确定性，价格波动受多种因素影响，包括但不限于宏观经济状况、行业动态、公司基本面信息以及市场情绪等。因此，如何从海量的信息中提取有用的数据，并据此建立有效的预测模型，成为了研究者们关注的重点之一。

本研究通过分析特定股票（代码 000858——五粮液）的历史交易数据，运用统计学中的判别分析和 Logistic 回归分析两种方法，探索并比较这两种模型在预测股票价格涨跌方面的效果。得到的 Fisher 判别函数为：

$$Z(x) = 0.64733 - 0.40095x_1 + 0.48623x_2 + 0.01287x_3 - 0.09895x_4 - 0.0427x_5$$

Logistic 回归方程为：

$$\text{Logit}(P) = -2.937 + 1.446x_1 - 1.534x_2 - 0.203x_3 + 0.300x_4 + 0.134x_5$$

研究结果表明，逻辑回归模型在预测股票涨跌情况时的准确率为 85%，略高于判别分析模型的 83%。此外，逻辑回归模型的 AUC 值为 0.9291，显著高于判别分析模型的 0.8111，表明逻辑回归模型在区分股票涨跌情况方面具有更强的能力。然而，判别分析模型的平均交叉验证分数略高于逻辑回归模型，这表明在不同数据分割下，判别分析模型的泛化能力稍微好一些。但从总体上看，逻辑回归模型在预测股票涨跌情况上表现更好。

总之，本研究通过对比判别分析和逻辑回归两种方法，发现逻辑回归模型在预测股票涨跌情况方面具有较高的准确率和区分能力。这一发现对于投资者来说具有重要意义，同时也为未来相关领域的研究提供了参考依据。未来的工作可以考虑结合更多复杂的模型和技术，进一步提高预测的准确性，为投资者提供更加可靠的决策支持。

## 参考文献

- [1]苏为华.多指标综合评价理论与方法问题研究[D].厦门大学, 2000.
- [2]张玲.财务危机预警分析判别模型及其应用[J].预测, 2000, (06): 38-40.
- [3]张玲.财务危机预警分析判别模型[J].数量经济技术经济研究, 2000, (03): 49-51
- [4]吴世农, 卢贤义.我国上市公司财务困境的预测模型研究[J].经济研究, 2001, (06): 46-55+96.
- [5]刘立国, 杜莹.公司治理与会计信息质量关系的实证研究[J].会计研究, 2003, (02): 28-36+65.
- [6]于立勇, 詹捷辉.基于 Logistic 回归分析的违约概率预测研究[J].财经研究, 2004, (09): 15-23
- [7]施锡铨, 邹新月.典型判别分析在企业信用风险评估中的应用[J].财经研究, 2001, (10): 53-57
- [8]朱根海, 刘祖祺, 朱培仁.应用 Logistic 方程确定植物组织低温半致死温度的研究[J].南京农业大学学报, 1986, (03): 11-16.
- [9]姜广辉, 张凤荣, 陈军伟, 等.基于 Logistic 回归模型的北京山区农村居民点变化的驱动力分析[J].农业工程学报, 2007, (05): 81-87.