

0.1 2023-2024 数据分析课程作业

2023-2024 数据分析课程作业

课程作业任务

任务一：下载相应数据的财务报表数据，计算相应的基本面指标（参见附录3.1）

下载数据网址<https://data.eastmoney.com/>

各组选择一个行业，每个行业对应多只证券

下载各只证券（一个行业30只证券）的财务报表：主要指标、资产负债表、利润表、现金流量表

按照3.1.2的公式计算基本面指标（以2023年9.30数据为基础计算，股价按照2023.9.29的收盘价计算）

任务二：根据提供的基本面指标，分析行业内各个股票的表现，用十种不同的理由分别推荐十只股票，并逐个说明理由以及用数据和图表进行展示

任务三：预测自选行业内一只股票的收盘价

任务四：自选行业内一只股票，允许做空，进行证券买卖（买卖价格以当日收盘价为准），计算收益率

实验要求：

附录

基本面指标

基本面分析

基本面指标详细介绍

"做多" (Long) 和"做空" (Short)

均值回归策略

行情接口

使用baostock进行行情获取。(http://baostock.com/baostock/index.php/%E9%A6%96%E9%A1%B5)

使用API进行日K线行情数据获取。

(http://baostock.com/baostock/index.php/A%E8%82%A1K%E7%BA%BF%E6%95%B0%E6%8D%AE)

本实验只使用每日收盘价数据

行业清单

1 课程作业任务

1.1 任务一：下载相应数据的财务报表数据，计算相应的基本面指标（参见附录3.1）

1.1.1 下载数据网址<https://data.eastmoney.com/>

1.1.2 各组选择一个行业，每个行业对应多只证券

行业list见附录3.5 行业清单，行业对应证券见stock_industry.csv，按组号选择对应序号的行业

1.1.3 下载各只证券（一个行业30只证券）的财务报表：主要指标、资产负债表、利润表、现金流量表

例：https://emweb.securities.eastmoney.com/pc_hsf10/pages/index.html?type=web&code=SZ300418&color=b#/cwfx/zcfzb

网站首页 加收藏 移动客户端 东方财富 天天基金网 东方财富证券 东方财富期货 Choice数据 基金交易 基金交易									
300418 昆仑万维		最新价：38.82 涨跌：6.47 涨跌幅：20.00%							
操盘必读	股东研究	经营分析	核心题材	资讯公告	公司大事	公司概况	同行比较	盈利预测	研究报告
财务分析	分红融资	股本结构	公司高管	资本运作	关联个股	资金流向	龙虎榜单	机构评级	智能点评
主要指标 杜邦分析 资产负债表 利润表 现金流量表 百分比报表									
注：杜邦分析是一种以净资产收益率为核心，综合分析企业盈利水平的方法。为了聚焦重点，在收入总额、成本总额、流动资产和非流动资产细分科目中，仅展示重点科目，部分科目未展示。收入总额和成本总额划分口径在财务上尚有争议，因此两者差额不一定与净利润相等。									
○ 资产负债表									
按报告期		报告期同比		全部 年报 季报 中报 季报					
资产负债表		2023-09-30	2023-06-30	2023-03-31	2022-12-31	2022-09-30			
流动资产									
货币资金		16.58亿	13.42亿	13.05亿	10.00亿	13.52亿			
交易性金融资产		16.68亿	13.49亿	12.53亿	15.01亿	5.365亿			
应收票据及应收账款		6.933亿	6.615亿	7.191亿	7.852亿	6.945亿			
其中：应收账款		6.933亿	6.615亿	7.191亿	7.852亿	6.945亿			
预付款项		7602万	5178万	1.047亿	3343万	4863万			
其他应收款合计		9139万	9357万	8797万	8958万	5946万			
一年内到期的非流动资产		--	--	--	--	13.15亿			

1.1.4 按照3.1.2的公式计算基本面指标（以2023年9.30数据为基础计算，股价按照2023.9.29的收盘价计算）

计算行业内30只证券的基本面指标，如资产负债表、利润表、现金流量表中无法提供指标公式中数据，请在实验报告中说明缺什么数据。

1.2 任务二：根据提供的基本面指标，分析行业内各个股票的表现，用十种不同的理由分别推荐十只股票，并逐个说明理由以及用数据和图表进行展示

1.3 任务三：预测自选行业内一只股票的收盘价

- 行情数据获取，需要获取的数据为2023年迄今为止“收盘价”数据，获取方式参见附录3.4
- 自选本行业中一只股票
- 建模并预测接下来两天的收盘价（1月8日-9日）
- 2023年1月7日将预测结果发到指定邮箱
- 2023年1月9日晚计算预测误差，公式如下：

$$RMSE = \sum_{i=1}^2 W_i * (\frac{X_i - \hat{X}_i}{X_i})^2$$

其中：X1代表周一实际收盘价， \hat{X}_1 表示周一收盘价预测值，W1表示周一的权重
W1=1，W2=0.5

1.4 任务四：自选行业内一只股票，允许做空，进行证券买卖（买卖价格以当日收盘价为准），计算收益率

- 自选行业内一只股票，并在1月7日将选择的股票发到指定邮箱
- 给定日期12月29日，以当日收盘价为买入价（做多）或卖出价（做空），交易100股自选股票。"做多"（Long）和"做空"（Short）参见附录3.2
- 在1月8日或1月9日进行反向交易，以收盘价为卖出价（做多）或买入价（做空），交易上述100股自选股票，可以根据收益最高原则选择这两天对应自选股票的收盘价进行交易。
- 收益率计算公式 $收益率 = \frac{收益}{期初资金}$ ，其中收益为(卖出价 - 买入价) × 100，

1. 做多期初资金为 $\text{买入价} \times 100$
2. 做空期初资金为 $\text{卖出价} \times 100$
5. 介绍所使用的交易策略，并列出了算法流程以及代码
6. 附录3.3均值回归策略介绍了一种常用的交易策略

2 实验要求:

• 2.1 提交时间:

- 1月10日课堂提交纸质文档并演讲
- 2023年1月9日晚十点将实验报告电子档和PPT发到指定邮箱linjie891@qq.com
 - 邮件标题命名规范: 课程作业-第X组
 - 实验报告命名规范: 2023-2024 数据分析课程作业-第X组.doc
 - PPT命名规范: 2023-2024 数据分析课程作业-第X组.ppt

• 2.2 提交材料要求

1. 按照实验报告格式，需要按照数据分析步骤写出各个步骤处理过程。
2. 实验报告封面应按照提供的模板文件填写。
3. 实验报告需以纸质形式双面打印。
4. 演讲方式: 每组派一位同学上台演讲。
5. 演讲需要使用PPT进行展示。
6. 提供纸质PPT，将6页内容打印在一页上，并进行双面打印。

• 2.3 分组要求: 每组4-5人，预计20组

• 2.4 2023年1月7日将任务三选定的股票代码和预测结果以及任务四的股票代码发到指定邮箱linjie891@qq.com。

1. 邮件标题: 第X组自选股票信息及预测结果
2. 邮件内容:
 1. 任务三自选股票代码: XXXXXX, 预测收盘价: XXXX, XXXX
 2. 任务四自选股票代码: XXXXXX, 买入价: XXXXXX(或者 卖出价 XXXXXX)

• 2.5 评分要求:

1. 小组成绩为个人课程作业成绩
2. 实验报告占课程作业总分80%
 1. 任务1占课程作业总分10%
 2. 任务2占课程作业总分30%
 1. 一项有意义的知识占3%
 3. 任务3占课程作业总分20%
 1. $(1 - (\text{RMSE误差排名} - 1) / 30) * 5\%$
 2. 分析步骤占课程作业总分15%
 4. 任务4占课程作业总分20%

1. $(1 - (\text{收益率排名} - 1) / 30) * 5\%$ (不进行交易的排名最低)
2. 分析以及设计交易策略步骤占课程作业总分15%
3. 演讲占课程作业总分20%
 1. PPT占课程作业总分5%
 2. 演讲水平占课程作业总分15%

3 附录

3.1 基本面指标

基本面指标是用于评估和分析公司或资产财务和经营状况的关键指标。这些指标涵盖了财务报表数据、经济指标以及其他与公司价值和盈利能力相关的因素。基本面指标提供了对公司基本情况的定量和定性评估，用于判断公司的盈利能力、偿债能力、成长潜力和风险水平等方面。

3.1.1 基本面分析

基本面分析是一种投资分析方法，通过对公司的基本面指标进行研究和分析，以评估其内在价值和未来发展趋势。基本面分析的目标是揭示市场价格与公司基本价值之间的差异，以确定投资机会和风险。

在基本面分析中，投资者会关注公司的财务报表，如利润表、资产负债表和现金流量表，以了解公司的盈利能力、财务健康状况和现金流情况。此外，经济指标、行业趋势、市场竞争力和管理团队等因素也会被纳入考虑。通过综合分析这些因素，投资者可以评估一个公司的长期增长潜力和风险，并做出相应的投资决策。

基本面分析通常与技术分析相对应。技术分析更关注市场价格的走势和交易量等技术指标，通过图表和图形模式来预测未来价格走势。基本面分析和技术分析可以结合使用，以提供更全面和准确的投资决策依据。

3.1.2 基本面指标详细介绍

以下是常用的基本面指标、定义、用LaTeX表示的公式，以及它们在量化分析和量化选股中的意义：

1. 市盈率 (Price-Earnings Ratio, P/E Ratio) :

- 定义：市盈率是衡量公司股价相对于每股盈利的相对估值指标。
- 公式：
$$P/E \text{ Ratio} = \frac{\text{市价}}{\text{每股盈利}}$$
- 意义：市盈率可用于评估公司的估值水平，较低的市盈率可能表示投资机会，较高的市盈率可能表示高估。

2. 市净率 (Price-to-Book Ratio, P/B Ratio) :

- 定义：市净率是衡量公司股价相对于每股净资产的相对估值指标。
- 公式：
$$P/B \text{ Ratio} = \frac{\text{市价}}{\text{每股净资产}}$$
- 意义：市净率可以提供关于公司资产价值相对于股价的信息，低市净率可能暗示投资机会。

3. 企业价值与销售收入比 (Enterprise Value-to-Sales Ratio, EV/Sales Ratio) :

- 定义：企业价值与销售收入比是衡量公司相对估值的指标，将企业价值除以销售收入。
- 公式：
$$EV/Sales \text{ Ratio} = \frac{\text{企业价值}}{\text{销售收入}}$$
- 意义：EV/Sales比率可以提供关于公司销售收入相对于企业价值的信息，较低的比率可能表示投资机会。

4. 资产收益率 (Return on Assets, ROA) :

- 定义：资产收益率是衡量公司利用其资产创造利润能力的指标。
- 公式：
$$ROA = \frac{\text{净利润}}{\text{总资产}}$$
- 意义：ROA可以帮助评估公司资产的利用效率，较高的ROA可能表示较好的盈利能力。

5. 负债比率 (Debt Ratio) :

- 定义: 负债比率是衡量公司负债水平的指标, 将总负债除以总资产。
- 公式: $\text{Debt Ratio} = \frac{\text{总负债}}{\text{总资产}}$
- 意义: 负债比率可以提供关于公司财务风险和财务稳定性的信息, 较低的比率可能表示较好的财务状况。

6. 现金流量比率 (Cash Flow Ratio) :

- 定义: 现金流量比率是衡量公司现金流量状况的指标, 可以有多种形式, 如经营活动现金流量与净利润的比率、自由现金流量与净利润的比率等。
- 公式: 具体计算公式根据所选的现金流量比率指标而定。
- 意义: 现金流量比率可以提供关于公司现金流量状况和财务稳定性的信息, 有助于评估公司的偿债能力和经营活动的稳定性。

7. 股息收益率 (Dividend Yield) :

- 定义: 股息收益率是衡量公司派发的股息相对于股价的比率。
- 公式: $\text{Dividend Yield} = \frac{\text{每股股息}}{\text{股票价格}}$
- 意义 (续): 股息收益率可以提供关于投资者从股息收入中获取的回报的信息, 较高的股息收益率可能表示较好的投资机会。

8. 每股盈利 (EPS) :

- 公司的净利润除以发行的普通股总数。
- 公式: $\text{EPS} = \frac{\text{净利润}}{\text{普通股总数}}$
- 意义: EPS是衡量公司盈利能力的重要指标, 投资者可以通过比较不同公司的EPS来评估其盈利水平和增长潜力。

9. 营业收入 (Revenue) :

- 公司在一定期间的总销售收入。
- 公式: 无需公式表示。
- 意义: 营业收入是评估公司销售能力和市场规模的关键指标, 从中可以了解公司的业务增长情况。

10. 净利润 (Net Profit) :

- 公司在一定期间的总利润, 即扣除所有费用和税后的收入。
- 公式: $\text{净利润} = \text{营业收入} - \text{费用} - \text{税收}$
- 意义: 净利润反映了公司的盈利能力和效率, 是投资者评估公司综合财务状况的重要依据。

11. 毛利率 (Gross Margin) :

- 公司的毛利润与营业收入之间的比率。
- 公式: $\text{毛利率} = \frac{\text{营业收入} - \text{销售成本}}{\text{营业收入}}$
- 意义: 毛利率衡量公司产品或服务的生产和销售的盈利能力, 高毛利率通常表示公司在销售中有更多的利润空间。

12. 净利润率 (Net Profit Margin) :

- 净利润与营业收入之间的比率。
- 公式: $\text{净利润率} = \frac{\text{净利润}}{\text{营业收入}}$
- 意义: 净利润率是评估公司每一销售额的净利润水平, 可以反映公司盈利能力的高低。

13. 资产负债比率 (Debt-to-Equity Ratio) :

- 公司负债总额与股东权益之间的比率。
- 公式: $\text{资产负债比率} = \frac{\text{负债总额}}{\text{股东权益}}$
- 意义: 资产负债比率反映公司融资结构和负债水平, 高比率可能意味着公司承担较大的债务风险。

14. 流动比率 (Current Ratio) :

- 公司流动资产与流动负债之间的比率。
- 公式: $\text{流动比率} = \frac{\text{流动资产}}{\text{流动负债}}$

- 意义：流动比率衡量公司偿付短期债务的能力，较高的流动比率通常表示公司具备良好的流动性和偿债能力。

15. 速动比率 (Quick Ratio) :

- 公司流动资产中除去存货后与流动负债之间的比率。
- 公式:
$$\text{速动比率} = \frac{\text{流动资产} - \text{存货}}{\text{流动负债}}$$
- 意义：速动比率更加严格地衡量公司在没有存货支持的情况下偿付短期债务的能力。

16. 资本回报率 (Return on Investment, ROI) :

- 公司净利润与投资总额之间的比率。
- 公式:
$$ROI = \frac{\text{净利润}}{\text{投资总额}}$$
- 意义：资本回报率衡量投资者对投资所获得的回报，是评估投资项目或公司绩效的指标之一。

17. 股东权益回报率 (Return on Equity, ROE) :

- 公司净利润与股东权益之间的比率。
- 公式:
$$ROE = \frac{\text{净利润}}{\text{股东权益}}$$
- 意义：ROE衡量股东对其投资的回报率，反映公司利用股东权益创造价值的能力。

18. 现金流量 (Cash Flow) :

- 公司在一定期间的现金流入和流出情况。
- 公式：无需公式表示。
- 意义：现金流量是评估公司现金流动状况和经营能力的重要指标，可以用来分析公司的偿债能力和投资活动。

19. 销售增长率 (Sales Growth Rate) :

- 公司营业收入在一定期间的增长率。
- 公式:
$$\text{销售增长率} = \frac{\text{当前期营业收入} - \text{上期营业收入}}{\text{上期营业收入}} \times 100\%$$
- 意义：销售增长率反映了公司业务的增长速度，是评估公司发展潜力和市场竞争力的重要指标。

在量化分析中，这些基本面指标可以用于筛选和评估股票，帮助投资者进行量化选股。通过对这些指标进行定量分析，并与行业平均值、竞争对手或市场指数进行比较，投资者可以识别出具有潜在投资价值的股票。这些指标可以作为量化模型的输入变量，结合其他技术指标和市场数据，来制定投资策略和决策。

量化选股的意义在于提供系统性和规则化的方法来评估股票，减少主观判断和情绪影响，通过数据驱动的方法来选择投资组合。通过量化分析，投资者可以快速筛选出满足特定条件的股票，提高投资效率和决策的一致性。同时，量化选股也可以帮助投资者控制风险，通过建立多因子模型和风险模型来优化投资组合的风险收益特征。

然而，需要注意的是，基本面指标只是量化选股的一部分，投资者还应考虑其他因素，如行业趋势、公司竞争优势、管理团队等。综合考虑多个因素，并进行有效的组合和权衡，才能做出更准确和全面的投资决策。

3.2 "做多" (Long) 和"做空" (Short)

"做多" (Long) 和"做空" (Short) 是金融领域中常用的术语，用于描述投资者对某种资产或证券的交易方向和立场。

- 做多 (Long)：指投资者购买某种资产或证券，期望其价格上涨，从中获取利润。投资者预期资产的价值将在未来增加，因此愿意持有该资产。做多是投资者最常见的交易方向，通常表示看涨市场。

例如，如果一个投资者认为某只股票的价格将上涨，他们可以选择以做多的方式购买该股票，并在未来将其卖出以获得利润。

- 做空 (Short)：指投资者借入某种资产或证券并立即卖出，然后在未来的某个时候再以更低的价格买入，归还借入的资产或证券，从中获取利润。投资者预期资产的价值将在未来下降，因此选择

做空该资产。做空是相对较为复杂和风险较高的交易策略。

做空通常用于投资者预计某个市场或资产将下跌时，以赚取价格下跌的利润。这种策略允许投资者从市场的下降中获利，与传统的做多策略相反。

需要注意的是，做空交易可能面临更高的风险，因为资产的价格理论上可以无限上涨，从而导致投资者的损失增加。因此，在进行做空交易时，投资者需要有一定的市场了解和风险管理策略。

3.3 均值回归策略

均值回归策略是一种常见的量化交易策略，基于统计学原理和价格趋势的理论。该策略的核心思想是，当价格偏离其均值时，存在一种趋势，价格会回归到其长期平均水平。

在均值回归策略中，投资者会选择一个或多个指标或价格序列作为基准，通常使用统计学中的平均值或移动平均线作为基准。当价格偏离基准时，投资者会采取相应的交易策略来获利。

具体而言，均值回归策略通常分为以下几个步骤：

1. 选择基准：选择一个或多个适当的指标或价格序列作为基准，如股价、收益率、相对强弱指标等。
2. 计算均值：计算基准的均值或移动平均线，用于衡量价格的长期平均水平。
3. 确定偏离阈值：设定价格与基准之间的偏离阈值，当价格偏离超过该阈值时认为存在回归机会。
4. 生成交易信号：当价格偏离超过阈值时，产生相应的交易信号。一般而言，当价格偏低时产生买入信号，当价格偏高时产生卖出信号。
5. 执行交易策略：根据交易信号执行相应的买入或卖出操作。可以使用各种交易工具，如股票、期货或衍生品等。
6. 风险管理：实施适当的风险管理策略，包括设定止损点和止盈点，控制仓位大小等。

均值回归策略适用于市场存在一定的价格波动和趋势回归的情况。然而，需要注意的是，市场条件可能会发生变化，导致价格无法回归到均值，因此在实施均值回归策略时，投资者需要对市场进行充分的研究和分析，并灵活调整策略参数。

3.4 行情接口

3.4.1 使用baostock进行行情获取。 (<http://baostock.com/baostock/index.php/%E9%A6%96%E9%A1%B5>)

3.4.2 使用API进行日K线行情数据获取。 (<http://baostock.com/baostock/index.php/A%E8%82%A1K%E7%BA%BF%E6%95%B0%E6%8D%AE>)

3.4.3 本实验只使用每日收盘价数据

3.5 行业清单

1	钢铁
2	农林牧渔
3	汽车
4	综合
5	公用事业
6	采掘
7	化工
8	计算机
9	家用电器
10	电子
11	国防军工
12	通信
13	机械设备
14	交通运输
15	食品饮料
16	建筑装饰
17	休闲服务
18	建筑材料
19	非银金融
20	房地产
21	轻工制造
22	纺织服装
23	电气设备
24	商业贸易
25	医药生物
26	传媒
27	银行
28	有色金属

行业对应证券见stock_industry.csv

