



量化金融分析师（AQF®）全国统一考试

模拟题

适用场次：2021 年 3 月

使用本模拟题，您应该遵守：

1. 本模拟题仅提供给参加 2021 年 3 月份 AQF 全国统一考试的考生，考生仅可以出于准备个人考试的目的查阅和打印本模拟题；
2. 严禁出于任何目的的复制、网络发布和传播、抄袭本模考题内容，如有违反，可能导致违纪或违法行为；

© 版权所有，侵权必究。

量化金融标准委员会

Standard Committee of Quantitative Finance

量化金融分析师（AQF®）全国统一考试模拟题

说明：本场考试中的代码都应采用 Python 3.X 版本作答。

解答题（每题 20 分，本部分共 100 分）

1. 李明，AQF，某量化基金经理，在量化研究的过程中，获取了 10 支股票的行业和量价信息存储在 SQLite 数据库 stock_data 中。其中，表 stock_price 存放这 10 支股票 2020 年的行情数据，部分数据如下：

Id	Date	Open	Close	High	Low	Volume	Code
0	2020-01-02	16.65	16.87	16.95	16.55	1530231.0	000001
1	2020-01-03	16.94	17.18	17.31	16.92	1116194.0	000001
2	2020-01-06	17.01	17.07	17.34	16.91	862083.0	000001

表 industry 存放相关行业的基本信息，部分数据如下：

Id	Industry	Description
1	financial service	Financial services include banking, securities, insurance, trust and fund
2	real estate	The real estate industry is an industry of real estate investment, development and management
3	internet	Internet industry is an industry based on computer network technology and using network platform to provide services

李明现希望在 SQLite 数据库 stock_data 中创建第三张表 stock_info，该表中存储 10 支股票的基本信息，包括股票代码（Code）、公司名称（CompanyName）、上市日期（ListingDate）、行业编号（IndustryId）四个字段，其中 Code、CompanyName、ListingDate 数据类型为文本类型，IndustryId 数据类型为整数型，并将 Code 设置为主键且非空。目前该信息存放在 csv 文件 stock_info 中， csv 文件部分数据如下：

Code	CompanyName	ListingDate	IndustryId
000001	平安银行	2001-04-03	1

000002	万科 A	1991-01-29	2
601939	建设银行	2007-09-25	1

1.1. 请你在数据库 `stock_data` 中按照以上要求创建数据表 `stock_info`, 并将 csv 文件 `stock_info` 中的数据录入该表。请在 Python 环境下完成该任务。(8 分)

参考答案:

创建数据库连接 (2 分)

```
import sqlite3
```

```
import pandas as pd
```

```
conn = sqlite3.connect('stock_data.sqlite')
```

```
cur = conn.cursor()
```

创建数据表 (2 分)

```
sql = """
```

```
CREATE TABLE stock_info (
```

```
    Code TEXT PRIMARY KEY NOT NULL,
```

```
    CompanyName TEXT,
```

```
    ListingDate TEXT,
```

```
    IndustryId INTEGER
```

```
);
```

```
"""
```

```
cur.execute(sql)
```

```
conn.commit()
```

从 csv 文件中读取数据 (2 分)

```
stock_info = pd.read_csv('stock_info.csv', dtype={'Code':str},  
encoding='utf-8')
```

将数据录入数据库 (2 分)

```
pd.io.sql.to_sql(stock_info, 'stock_info', con=conn, if_exists='replace',  
index=False)
```

```
conn.commit()
```

```
conn.close()
```

1.2. 李明发现表 `stock_info` 中数据有误，平安银行的上市日期应为 1991 年 4 月 3 日，误写为 2001 年 4 月 3 日，请你修改此项错误，只需写出 SQL 命令即可。（4 分）

参考答案：

```
UPDATE stock_info
SET ListingDate = "1991-04-03"
WHERE CompanyName = "平安银行";
```

1.3. 请查找出中国银行 2020 年 1 月 1 日到 2020 年 1 月 31 日的股票行情信息，只需写出 SQL 命令即可（4 分）

参考答案：

```
SELECT Date, Open, Close, High, Low
FROM stock_info AS s
JOIN stock_price AS p
ON s.Code = p.Code
WHERE CompanyName = "中国银行"
AND Date BETWEEN "2020-01-01" AND "2020-01-31";
```

1.4. 请查询并计算数据库中 10 支股票收盘价的均值，只需写出 SQL 命令即可（4 分）

参考答案：

```
SELECT Code, AVG(Close)
FROM stock_price
GROUP BY Code;
```

2. 某研究员正在构建一个配对交易策略，他选取了两种期货指数作为配对交易的对象。为了验证此两种期货具有一定相关性，他获取此两种期货指数 2019 年的价格数据作为训练集，分别存储于 csv 文件 `training1` 和 `training2` 中，部分数据如下：

trade_date	close
20191231	1082.101
20191230	1084.934

trade_date	close
20191231	929.528
20191230	928.302

20191227	1083.801
20191226	1084.934

20191227	929.936
20191226	929.936

其中，`trade_date` 为交易日期，`close` 为期货指数当日收盘价。同时，他选取 2020 年作为回测区间，将 2020 年的期货指数价格数据存储于 csv 文件 `test` 中，部分数据如下：

trade_date	close_1	close_2
2020/1/2	1086.633	936.065
2020/1/3	1086.633	936.065
2020/1/6	1083.801	931.162
2020/1/7	1091.166	934.839

其中，`trade_date` 为交易日期，`close_1` 和 `close_2` 分别为两个期货指数当日收盘价。

2.1. 为了检验这两个期货指数是否适合进行配对交易，该研究员准备使用训练集数据进行相关性检查，请你使用 `training1` 和 `training2` 中的数据，绘制两个期货指数的价格走势图，并计算相关系数。（8 分）

参考答案：

读取数据并进行拼接（4 分）

```
training1 = pd.read_csv('training1.csv', index_col=0,
parse_dates=True).sort_index()
```

```
training2 = pd.read_csv('training2.csv', index_col=0,
parse_dates=True).sort_index()
```

```
training = training1.join(training2, how='inner', lsuffix='_1',
rsuffix='_2')
```

绘制价格走势图（2 分）

```
training.plot(figsize=(8,6))
```

计算相关系数（2 分）

```
training.corr()
```

2.2. 在检验了两个期货指数的相关性之后，他认为这两种期货指数适合进行配对交易，并准备使用 2020 年价格数据进行配对交易策略回测。该研究员认为，当两个期货指数价差偏离均值 1.5 倍标准差时存在错误定价，并进行开仓操作，回归正常价格区间时进行平仓。

请你使用题目所提供的数据，计算交易信号。（9 分）

参考答案：

计算训练集均值和标准差（2 分）

```
training['spread'] = training['close_1'] - training['close_2']  
mu = training['spread'].mean(); sigma = training['spread'].std()
```

读取测试集数据

```
test = pd.read_csv('test.csv', index_col=0, parse_dates=True)
```

计算测试集价差上下轨（4 分）

```
test['spread'] = test['close_1'] - test['close_2']
```

```
test['up']=mu + 1.5 * sigma
```

```
test['down']=mu - 1.5 * sigma
```

计算交易信号（3 分）

```
test['position_1'] = np.where(test['spread'] > test['up'], -1, 0)
```

```
test['position_1'] = np.where(test['spread'] < test['down'], 1,  
test['position_1'])
```

```
test['position_2'] = -test['position_1']
```

2.3. 请问构建配对交易策略的前提条件是什么？答出任意两点即可。（3 分）

参考答案：

“卖空机制”

作为市场中性策略的一种，配对交易需要同时持有多头和空头仓位，因此需要对其中一个投资标的进行卖空。

“构建配对组”

配对交易需要有两个风险收益特征非常相似的投资标的进行配对。比如，通过统计原理进行筛选，寻找相关系数较高的期货品种。

“价差收敛”

配对交易的获利关键为投资标的之间的价差收敛。如果因异常因素影响，导致某投资标的基本面发生变化，投资标的之间的价差可能持续扩大，此时需要对此配对组合仓位进行止损。

3. 在多因子策略中，对于有效因子的挖掘与测试是模型构建的重要前提。除了使用回归法

和因子 IC 值来衡量因子对收益的预测能力之外，我们通常也会关注因子对股票未来收益的分层效果，即对因子进行单调性检验。李明，某量化研究员，将某一期处理好的因子数据保存于文件 `month_data.csv` 中，包含期初股票池中所有股票的市值对数（`size`）、账面市值比（`btm`）、市盈率倒数（`ep`）因子值数据，以及该期该股票的收益率数据（`monthly_r`）和相对大盘的超额收益数据（`excess_r`），部分数据如下：

	size	btm	ep	monthly_r	excess_r
000002.SZ	14.07125	0.442145	0.04199	0.003515	-0.09711
000004.SZ	10.80556	0.196924	0.007834	0.10961	0.008983
000005.SZ	11.91049	0.603573	0.001456	0.073427	-0.0272
000007.SZ	10.89391	0.359467	0.027798	0.091667	-0.00896
000008.SZ	11.28962	0.03912	0.001234	0.118806	0.018178

3.1. 现李明想要对市值对数因子进行单因子有效性检验，使用因子分层测试法检验因子的单调性，即因子值对于股票未来收益的分层效果。请你将所有股票按照该因子值的大小分成五组，并计算该期每组的平均超额收益。（12 分）

参考答案：

读取数据（1 分）

```
data = pd.read_csv('month_data.csv', index_col=0)
```

按因子值排序（1 分）

```
data.sort_values(by='size', inplace=True)
```

计算因子值的五分位数（4 分）

```
a1, a2, a3, a4, a5, a6 = [data['size'].quantile(i) for i in
np.linspace(0,1,6)]
```

按因子值从小到大分组（4 分）

```
data['group'] = np.where(data['size']<=a2, 1,
                        np.where(data['size']<=a3, 2,
                        np.where(data['size']<=a4, 3,
                        np.where(data['size']<=a5, 4, 5))))
```

计算每组超额收益平均值（2 分）

```
grouped = data.groupby(by='group')['excess_r'].mean()
```

3.2. 李明按照以上方法，计算出测试时段中所有期的股票分组收益，并保存于文件 `size_factor.csv` 中，部分数据如下：

date	group1	group2	group3	group4	group5
20050131	0.0021	0.005	0.0034	0.0014	-0.0152
20050228	-0.0053	-0.0127	-0.0052	0.0024	0.0289
20050331	-0.0394	-0.0249	-0.0028	0.0195	0.056
20050429	0.0411	0.0267	0.0082	-0.0265	-0.0633
20050531	0.02	0.0084	0.003	-0.006	-0.0119

其中，每一行表示该期各组的股票收益率情况，例如，20050131 一行表示，2005 年 1 月期间，按照市值大小将股票分为 5 组后，每组的平均收益率。请你计算在整个测试期间，每组的月平均收益率，并用柱形图展示该因子的单调性情况。（8 分）

参考答案：

读取数据（2 分）

```
size_r=pd.read_csv('size_factor.csv',index_col=0,parse_dates=True)
```

数据表求每一列的均值（2 分）

```
avg_size_r = size_r.mean()
```

画图（4 分）

```
plt.bar(avg_size_r.index, avg_size_r)
```

```
plt.ylabel('Monthly Excess Return')
```

```
plt.title('Size Factor Group Test')
```

4. 李明，某量化基金经理，正在测试逻辑回归模型的预测能力。他认为，根据个股的市净率和市值大小可以判断出个股是否存在投资机会。他将个股市净率和对数市值作为逻辑回归模型的特征数据储存在文件 `data_x.csv` 中，将个股当年是否上涨作为标签数据储存在文件 `data_y.csv` 中。

`data_x` 部分数据如下：

code	pb	log_mv
002115.SZ	3.0881	13.55167

002408.SZ	1.4628	13.92102
603208.SH	3.7099	13.09767
002401.SZ	3.2581	12.62328
600644.SH	1.8101	12.47892

data_y 部分数据如下：

code	change
002115.SZ	-1
002408.SZ	1
603208.SH	1
002401.SZ	1
600644.SH	-1

现需要你根据以下步骤描述，编写相应代码。

4.1. 定义函数 `standardize` 用于数据标准化。假设函数参数为 `data_x` 中的一列特征数据 `x`，函数返回值为标准化之后的特征数据，标准化公式为：

$$x^* = \frac{x - \bar{x}}{\sigma}$$

（2 分）

参考答案：

```
def standardize(x):
    return (x - x.mean())/x.std()
```

4.2. 请使用第一小问中定义的函数，对特征数据进行标准化，并将标准化后的数据保存在变量 `data_2` 中。（4 分）

参考答案：

```
# 读取数据
data_x = pd.read_csv('data_x.csv', index_col=0)
data_y = pd.read_csv('data_y.csv', index_col=0)

# 标准化
data_2 = data_x.apply(standardize)
```

4.3. 将特征数据和标签数据进行分组，其中 80%作为训练集，20%为测试集，分别将训练集和测试集的特征数据和标签数据存储于变量 `train_x`, `train_y`, `test_x`, `test_y` 中。（4 分）

参考答案：

```
from sklearn.model_selection import train_test_split
features = data_2.values
labels = data_y.values.flatten()
train_x, test_x, train_y, test_y = train_test_split(features, labels,
test_size=0.2)
```

4.4. 使用逻辑回归模型进行模型训练。（4 分）

参考答案：

```
from sklearn.linear_model import LogisticRegression
lr = LogisticRegression()
lr.fit(train_x, train_y)
```

4.5. 使用测试集数据对模型进行测试，根据测试集特征预测标签，并将预测的结果保存在变量 `predict_y` 中。（2 分）

参考答案：

```
predict_y = lr.predict(test_x)
```

4.6. 根据以下公式计算模型准确率：

模型准确率=测试集中预测正确的次数/总的预测次数

（4 分）

参考答案：

```
accuracy = (test_y == predict_y).sum()/len(test_y)
```

5. 在投资过程中，风险管理对于稳定投资收益具有重要作用。在险价值 VaR 衡量了在一定置信水平下，某资产或资产组合在未来特定一段时间内的最大可能损失。李明，某量化基金经理，希望使用在险价值 VaR 来衡量某投资标的的投资风险，他已将该投资标的的近一年

的价格数据保存于文件 `data.csv` 中，部分数据如下：

date	open	close	high	low
2020-01-02	22.68	22.88	23.18	22.5
2020-01-03	22.89	23.04	23.2	22.71
2020-01-06	22.94	23.23	23.33	22.76
2020-01-07	23.23	23.15	23.28	22.91
2020-01-08	23.01	22.64	23.03	22.6

李明想观察该投资标的收盘价的收益率分布是否符合正态分布。请你：

5.1. 计算该投资标的收盘价的日收益率序列保存于变量 `returns` 中，数据类型 `pandas.Series`，使用算数收益率计算方法；（4分）

参考答案：

```
# 导入模块

import numpy as np

import pandas as pd

# 读取数据（2分）

data=pd.read_csv('data.csv',index_col=0,parse_dates=True)

# 计算算数收益率（2分）

returns=data['close'].pct_change().dropna()
```

5.2. 绘制 QQ 图，并判断该投资标的的日收益率是否符合正态分布；（4分）

参考答案：

```
# 导入模块（1分）

from scipy import stats

from matplotlib import pyplot as plt

# 绘制图形（2分）

plt.figure()

stats.probplot(returns, dist="norm", plot=plt)

plt.title('QQ Plot')

# 解释图形（1分）
```

该投资标的的日收益率分布相对于正态分布呈现尖峰肥尾的特征。

5.3. 历史模拟法是一种常用的非参数计算 VaR 的方法，采用历史数据直接估计资产价格变化的概率分布。请你使用 `data.csv` 文件中最近 100 天的数据，使用历史模拟法计算 95%和 99%置信度水平下的日在险价值（Daily VaR），分别存储于变量 `dVaR_95_hs` 和 `dVaR_99_hs` 中。（4 分）

参考答案：

```
# 最近 100 天收益率（1 分）
recent_rs = returns[-100:]

# 计算 95%的 Daliy VaR（2 分）
dVaR_95_hs = np.percentile(recent_rs, 5)

# 计算 99%的 Daliy VaR（1 分）
dVaR_99_hs = np.percentile(recent_rs, 1)
```

5.4. 当能够确定资产收益近似服从正态分布的情况下，可以使用参数法计算在险价值 VaR。请你使用 `data.csv` 文件中最近 100 天的数据，使用参数法计算 95%和 99%置信度水平下的日在险价值（Daily VaR），分别存储于变量 `dVaR_95_param` 和 `dVaR_99_param` 中。（4 分）

参考答案：

```
# 计算收益率均值和标准差（2 分）
mu = recent_rs.mean()
sigma = recent_rs.std()

# 计算在险价值（2 分）
dVaR_95_param = mu - stats.norm.ppf(0.95) * sigma
dVaR_99_param = mu - stats.norm.ppf(0.99) * sigma
```

5.5. 绘制该标的资产最近 100 天日收益率的频率分布直方图，并标注参数法下 95%VaR 和 99%VaR 所在位置。（4 分）

参考答案：

```
# 绘制频率分布直方图（2 分）
```

```
plt.figure()  
plt.hist(recent_rs, bins=50, density=True)  
# 标注 VaR 的位置 (2 分)  
plt.axvline(dVaR_99_param, color='r', label = '99% VaR')  
plt.axvline(dVaR_95_param, color='y', label = '95% VaR')  
plt.legend()  
plt.xlabel('returns')
```