



量化金融分析师（AQF®）全国统一考试

模拟题

适用场次：2020 年 9 月

使用本模拟题，您应该遵守：

1. 本模拟题仅提供给参加 2020 年 9 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。

1. 单选题（每题 1 分，本部分共 20 分）：只有一个正确答案，选对得 1 分，选错或不选得 0 分。

1.1. 李明，AQF，某量化交易员，准备使用限价定单买入美元外汇，已知目前人民币兑美元报价为 0.1432，且在当前价格下该定单未被触发，则李明设置的限价单的价格可能为（ ）？

- A. 6.9500 人民币/美元
- B. 6.9900 人民币/美元
- C. 0.1420 美元/人民币
- D. 0.1400 美元/人民币

参考答案：A

解析：限价定单指客户对于证券经纪人的买卖委托中设有低于市场价格的买进价格或高于市场价格的卖出价格的指示。限价定单规定了买者愿意买的最高价格或卖者愿意卖的最低价格。题中人民币兑美元当前报价为 0.1432，则美元限价买单的价格应低于 6.9832 人民币/美元，只有 A 选项符合条件。

1.2. 某量化交易员在进行量化回测过程中发现，存储回测日期的列表 `date_list` 中存在非日期格式的字符串，`date_list` 部分数据如下：

```
date_list = ['2019-12-02', '日期', '2019-12-03', ...]
```

他编写了如下代码删除该列表中非日期格式的字符串：

```
import time

for i in date_list:

    try:

        time.strptime(i, "%Y-%m-%d")

    except:

        print(f'"{i}"不是有效日期')

        date_list.____(i)
```

则以上横线处的代码为（ ）？

- A. append
- B. delete
- C. remove
- D. pop

参考答案：C

解析：A 选项，append 方法可以在列表末尾添加新的对象。B 选项，列表没有 delete 方法。C 选项，remove 方法可以删除列表某个元素。D 选项，pop 方法可以按索引移除列表中的一个元素。

1.3. 随着计算机的普及和互联网技术的发展，量化投资从上世纪 70 年代以来发展迅猛，目前已成为海外主流投资方法之一。以下关于量化投资的说法中不正确的是（ ）？

- A. 量化投资通常以获取稳定收益为目的，并通过计算机程序化发出买卖指令
- B. 由于量化投资使用数量化方法，因此舆情数据等文本类数据不属于量化投资的研究范畴
- C. 基本面量化投资策略可以使用公司财务数据等基本面数据作为策略开发基础
- D. 量化投资依靠模型挑选股票，具有高度纪律性，因而可以避免贪婪、恐惧等人性弱点

参考答案：B

解析：舆情数据等文本类数据也是量化投资策略的数据来源之一，可使用自然语言处理对其进行处理和分析。

1.4. 在多因子策略研究过程中，信息系数（Information Coefficient，简称 IC）常用于衡量选股因子的有效性。则以下选项中，关于信息系数描述错误的是（ ）？

- A. 信息系数最大值为 1，表示因子的预测能力 100%准确
- B. 常用的信息系数有 Normal IC 和 Rank IC
- C. 信息系数计算的是某时点所选股票的因子值与股票下期收益率的协方差
- D. 研究时可以通过因子 IC 衰退，衡量该因子有效性的稳定情况

参考答案：C

解析：信息系数 IC（Information Coefficient），表示所选股票的因子值与股票下期收益率的截面相关系数，通过 IC 值可以判断因子值对下期收益率的预测能力，常用有 Normal IC 和 Rank IC。

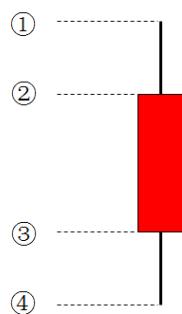
1.5. 以下哪种运算符不支持字符串数据类型（ ）？

- A. 算术运算符“+”
- B. 成员运算符“in”
- C. 比较运算符“>”
- D. 算术运算符“/”

参考答案：D

解析：算术运算符“/”不支持字符串数据类型的运算。

1.6. K线图，又称蜡烛图或阴阳线，广泛应用于股市及期货市场，用于记录股票或期货的开盘价、最高价、最低价、收盘价。根据频率不同，K线可分为日K线、周K线、月K线等。下图展示了某日某股票的日K线图，则哪个数字对应的位置表示该日该股票的收盘价（ ）？



- A. ①
- B. ②
- C. ③
- D. ④

参考答案：B

解析：图中所示K线图为阳线图，表示当日收盘价高于开盘价，则该K线实体上边线表示当日该股票的收盘价。

1.7. 对pandas模块中的数据结构DataFrame进行DataFrame.sort_index(axis=1)操作，对于其操作逻辑，下列描述正确的是（ ）？

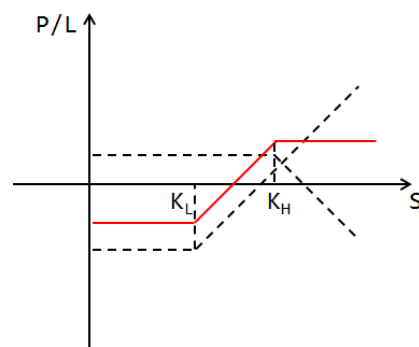
- A. 对DataFrame以行索引为参照以升序对行进行排序

- B. 对 DataFrame 以列索引为参照以升序对列进行排序
- C. 对 DataFrame 以行索引为参照以降序对行进行排序
- D. 对 DataFrame 以列索引为参照以降序对列进行排序

参考答案：B

解析：`DataFrame.sort_index(axis=1)`表示对 DataFrame 以列索引为参照以升序对列进行排序。

1.8. 近年来，随着期权品种的日趋丰富以及期权交易量的逐步扩大，中国衍生品市场逐步完善。下图展示了某种期权组合的收益曲线，其中纵轴代表损益，横轴代表标的资产的价格，则该图中的红色实线表示哪种期权组合的损益（ ）？



- A. 跨式期权组合
- B. 牛市价差期权
- C. 蝶式期权组合
- D. 熊市价差期权

参考答案：B

解析：牛市价差期权又称多头价差期权，可以通过买进一个低行权价的看涨期权，同时卖出一个高行权价的看涨期权进行构建。

1.9. 在进行量化策略回测的过程中，通常需要绘制策略的净值曲线，从而评判策略的可行性。则以下代码中，可以将绘制的净值曲线图嵌入在当前 Jupyter Notebook 文件中显示的魔法命令为（ ）？

- A. `import matplotlib`
- B. `import matplotlib.pyplot as plt`
- C. `%show_figure`

D. `%matplotlib inline`

参考答案：D

解析：Jupyter Notebook 魔法命令“`%matplotlib inline`”表示当使用 matplotlib 画图时，将图片嵌入在 Jupyter Notebook 里，不以单独窗口显示。

1.10. Python 是一个高层次的结合解释性、编译性、互动性和面向对象的计算机程序设计语言。自 2020 年 1 月 1 日起，Python2 不再得到支持，Python3 将正式成为主流 Python 语言。则以下关于 Python3 基础语法的说法中正确的是（ ）？

- A. 变量在使用前都必须赋值
- B. 变量名可以是字母或数字开头
- C. 使用缩进和大括号来表示代码块
- D. 可以在同一行中使用冒号分割多条赋值语句

参考答案：A

解析：A 选项正确，变量在赋值之后才被创建。B 选项错误，变量名不可以是数字开头。C 选项错误，Python 不使用大括号来表示代码块。D 选项错误，Python 可以在同一行中使用分号分割多条赋值语句，而不是冒号。

1.11. 在对股票进行研究过程中，研究员经常将行业划分为周期性行业与非周期性行业，以此进一步深入研究股票价格的走势特征。则以下选项中，属于非周期性行业的是（ ）？

- A. 医药
- B. 有色金属
- C. 水泥
- D. 房地产

参考答案：A

解析：周期性行业是指和国内或国际经济波动相关性较强的行业，其中典型的周期性行业包括大宗原材料（如钢铁，煤炭等），工程机械，船舶等。非周期性行业是指那些不受宏观经济影响的行业，这些行业往往集中在涉及人类日常消费的行业，例如：食品、医药、酒类、服装等。医药属于非周期性行业，而有色金属、水泥、房地产属于周期性行业。

1.12. 某量化研究员收集某股票 100 天的日收益率数据存储于变量 `a` 中，数据类型为

numpy.ndarray, 部分数据如下:

```
array([-0.58, 0.66, 1.29, ..., -2.00, -1.31])
```

现希望将该一维数组修改为 5 行一组的数组以方便查看, 修改后变量 **a** 部分数据如下:

```
array([[ -0.58,  0.66,  1.29, -0.63, -1.43], ...,  
       [ 0.27,  1.38, -0.49, -2.00, -1.31]])
```

则以下代码中能够满足要求的是 () ?

- A. `a.reshape((20,5))`
- B. `a.resize((20,5))`
- C. `a = a.shape((5,20))`
- D. `a = np.resize(a, (5,20))`

参考答案: B

解析: 题目要求将形状为(100,)的一维数组修改成形状为(20,5)的二维数组, B 选项符合条件。A 选项错误, `reshape` 方法将修改后数组返回, 不对原数组进行修改, 变量 **a** 不发生变化。C 选项错误, `shape` 为 ndarray 的属性, 不可以使用函数调用。D 选项错误, 形状修改为(5,20)不符合题目要求。

1.13. 以下关于多空 **alpha** 策略和多因子策略的说法错误的是 () ?

- A. 当一个投资策略在尽量避免市场风险时, 我们说这是一个市场中性策略
- B. 多因子策略的交易思想是: 创建具有尽可能高且稳定的 **alpha** 的投资组合
- C. 流通市值是一个价值因子, 我们可以通过市值因子分析公司估值对股价的影响
- D. 创建多因子模型的实际过程中, 应该注意多个因子数据的标准化问题

参考答案: C

解析: 流通市值属于规模因子, C 选项错误。

1.14. 某量化研究员在进行多因子策略的过程中发现因子之间相关性很高, 因此想采取模型缩减方法对模型进行简化, 使得缩减后的模型中新因子之间均线性不相关, 则他采取的机器学习算法为 () ?

- A. 支持向量机
- B. 惩罚回归
- C. 主成分分析

D. K 最近邻

参考答案: C

解析: 主成分分析可以通过正交变换将一组可能存在相关性的变量转换为一组线性不相关的变量, 转换后的这组变量叫主成分。

1.15. 某量化研究员在储存研究数据时需要储存三个因子的时间序列数据, 他希望将数据保存为 NumPy 中的一个 n 行 3 列 Nddarray, 并将其命名为 data, 数据形式如下:

```
array([[3.887000e-01, 7.586600e+00, 7.945000e-01],
       [6.962000e-01, 3.829858e+02, 6.186500e+00],
       [3.493000e-01, 2.601460e+01, 1.133700e+00],
       ...,
       [1.352100e+00, 1.019970e+01, 1.860600e+00]])
```

现在, 他希望获得一个新数组, 数据与 data 相同, 但新数组将 data 中的第一列数据和第三列数据调换位置, 调换后效果如下:

```
array([[7.945000e-01, 7.586600e+00, 3.887000e-01],
       [6.186500e+00, 3.829858e+02, 6.962000e-01],
       [1.133700e+00, 2.601460e+01, 3.493000e-01],
       ...,
       [1.860600e+00, 1.019970e+01, 1.352100e+00]])
```

则以下可以满足他需求的代码为 () ?

- A. np.dot(data, np.eye(3))
- B. np.dot(np.eye(3), data)
- C. np.dot(np.array([[0,0,1], [0,1,0],[1,0,0]]), data)
- D. np.dot(data, np.array([[0,0,1], [0,1,0],[1,0,0]]))

参考答案: D

解析: np.dot() 函数可以进行矩阵乘法运算, 右乘特定初等矩阵可以实现矩阵列位置对换。

1.16. 某研究员在编写代码过程中, 希望筛选出某对象 obj 中变量名仅包含小写字母的属性, 并将其保存在列表中。则以下代码中可以实现该需求的是 () ?

- A. [i for i in dir(obj) if i.isalpha() or i.islower()]

- B. `[i for i in help(obj) if i.isalpha() or i.islower()]`
- C. `[i for i in dir(obj) if i.isalpha() and i.islower()]`
- D. `[i for i in dir(obj) if i.isalpha() | i.islower()]`

参考答案: C

解析: `dir()` 函数可以查看某对象所有属性和方法, `isalpha()` 方法用于检测字符串是否只由字母组成, `islower()` 方法检测字符串是否由小写字母组成。

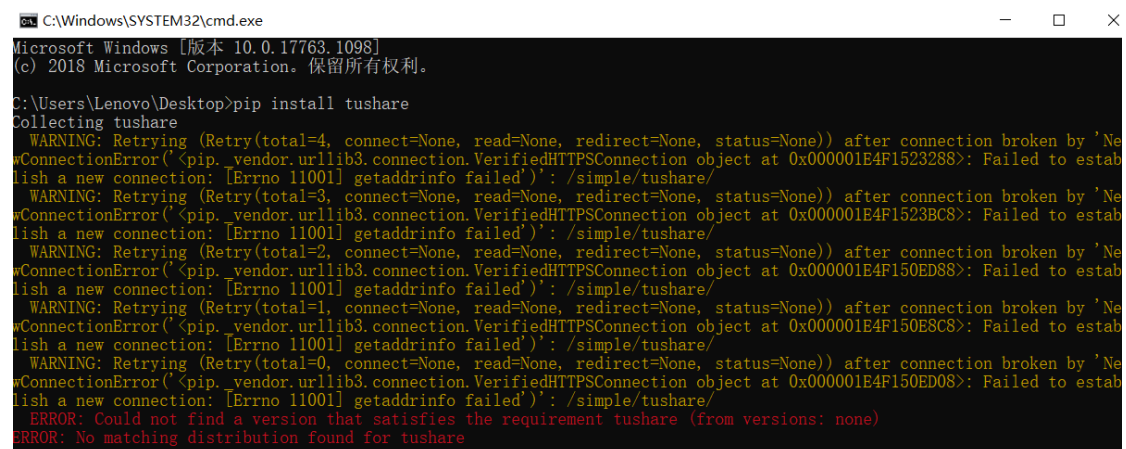
1.17. 通常在股票行情交易软件上, 通过十档行情界面可以看到股票的 20 个特殊价格及挂单情况, 即买一至买十和卖一至卖十的委托价格及委托数量。在股票高频交易研究过程经常需要对股票的买卖十档行情数据进行建模分析, 则以下对于十档行情的买卖价格比较中, 关系描述正确的是 () ?

- A. 买一价>买十价>卖一价>卖十价
- B. 卖十价>卖一价>买十价>买一价
- C. 卖十价>卖一价>买一价>买十价
- D. 买十价>买一价>卖一价>卖十价

参考答案: C

解析: 在十档行情中, 买一至买十价格依次降低, 卖一至卖十价格依次升高。

1.18. `pip` 是一种非常重要的工具, 可以用于 `Python` 工具包安装与管理。某同学在进行工具包安装时未能成功安装工具包, 并反馈以下警告信息:



```

C:\Windows\SYSTEM32\cmd.exe
Microsoft Windows [版本 10.0.17763.1098]
(c) 2018 Microsoft Corporation. 保留所有权利。

C:\Users\Lenovo\Desktop>pip install tushare
Collecting tushare
  WARNING: Retrying (Retry(total=4, connect=None, read=None, redirect=None, status=None)) after connection broken by 'NewConnectionError(<pip._vendor.urllib3.connection.VerifiedHTTPSConnection object at 0x000001E4F1523288>: Failed to establish a new connection: [Errno 11001] getaddrinfo failed)': /simple/tushare/
  WARNING: Retrying (Retry(total=3, connect=None, read=None, redirect=None, status=None)) after connection broken by 'NewConnectionError(<pip._vendor.urllib3.connection.VerifiedHTTPSConnection object at 0x000001E4F1523BC8>: Failed to establish a new connection: [Errno 11001] getaddrinfo failed)': /simple/tushare/
  WARNING: Retrying (Retry(total=2, connect=None, read=None, redirect=None, status=None)) after connection broken by 'NewConnectionError(<pip._vendor.urllib3.connection.VerifiedHTTPSConnection object at 0x000001E4F150ED88>: Failed to establish a new connection: [Errno 11001] getaddrinfo failed)': /simple/tushare/
  WARNING: Retrying (Retry(total=1, connect=None, read=None, redirect=None, status=None)) after connection broken by 'NewConnectionError(<pip._vendor.urllib3.connection.VerifiedHTTPSConnection object at 0x000001E4F150E8C8>: Failed to establish a new connection: [Errno 11001] getaddrinfo failed)': /simple/tushare/
  WARNING: Retrying (Retry(total=0, connect=None, read=None, redirect=None, status=None)) after connection broken by 'NewConnectionError(<pip._vendor.urllib3.connection.VerifiedHTTPSConnection object at 0x000001E4F150ED08>: Failed to establish a new connection: [Errno 11001] getaddrinfo failed)': /simple/tushare/
ERROR: Could not find a version that satisfies the requirement tushare (from versions: none)
ERROR: No matching distribution found for tushare
  
```

则以下说明中, 最可能是导致本次安装工具包失败的原因为 () ?

- A. 安装命令错误

- B. 网络传输问题
- C. 当前用户没有安装权限
- D. pip 站点无 tushare 工具包

参考答案：B

解析：通过报错信息和警告信息“建立新连接失败”可以推测安装失败的原因可能为网络问题。

1.19. 下列赋值代码中，错误的是（ ）？

- A. `import pandas as pd`
- B. `meter * 100 = 1000`
- C. `minute = 24 * 60`
- D. `a = lambda x:x+1`

参考答案：B

解析：赋值时被赋值部分不可以是表达式。B 选项会发生语法错误。

1.20. 很多量化对冲基金都会动用资金杠杆进行投资。某量化对冲基金，以其自有募集的 1 亿资金，通过杠杆放大使其规模达到了 20 亿并投资某标的资产，若该投资标的价格上涨 2%，则该对冲基金的收益率为（ ）？

- A. 2%
- B. 40%
- C. -2%
- D. -40%

参考答案：B

解析：杠杆投资时投资收益为标的资产价格波动乘以杠杆倍数。

2. 多选题（每题 2 分，本部分共 20 分）：有 2-5 个正确答案，全部选对得 2 分，少选得 1 分，选错或不选得 0 分。

2.1. 在量化交易策略的开发过程中，高质量的数据对于策略回测的准确度至关重要。在进行数据分析之前，量化研究员往往会耗费大量时间进行数据清洗工作。以下数据清洗的做法中，不正确的是（ ）？

- A. 在适当的情况下，可以使用样本的平均数、中位数等统计量填补缺失值
- B. 将样本量为 20 的数据去除最大的 5 个数据和最小的 5 个数据以避免极值的影响
- C. 在处理不同时区的数据时，未进行时区统一操作
- D. 对于上证 50 每日收盘价等结构化数据使用自然语言处理进行数据清洗
- E. 将不同量纲的数据进行 z-score 标准化处理

参考答案：BCD

解析：B 选项，将样本量为 20 的数据去除 10 个数据占比过大，会牺牲大量有效信息，删除异常值应考虑样本量的大小；C 选项，处理不同时区的数据时，应先对时间进行对齐操作；D 选项，自然语言处理一般用来处理文本等非结构化数据。

2.2. 以下代码中，哪些不能删除某列表中索引为偶数的元素？（假设待处理的列表名称为 a）

- A.

```
for i, item in enumerate(a):  
    if i % 2 == 0:  
        del a[i]
```
- B.

```
for i in range(len(a)):  
    if i % 2 == 0:  
        del a[i]
```
- C.

```
for i in reversed(range(len(a))):  
    if i % 2 == 0:  
        del a[i]
```
- D.

```
for i in range(len(a)):  
    if i % 2 != 0:  
        a.append(a[0])  
del a[0]
```

```
E. for i in range(len(a)):
    if i % 2 == 0:
        del a[0]
    else:
        a.append(a[0])
        del a[0]
```

参考答案：AB

2.3. 套利交易由于其较低的风险和相对稳定的收益率,长期以来受到国内外量化基金的青睐。常见的套利模式有跨期套利、跨市场套利、跨产品套利等,在期货、外汇等市场均有广泛应用。以下关于套利交易的说法中,正确的是()?

- A. 统计套利的优势是不依赖市场趋势,因此可以完全规避系统性风险
- B. 统计套利关注资产在统计意义上的错误定价,是一种无风险套利
- C. 跨市场套利策略利用同一资产在不同市场的价格差异进行套利
- D. 跨品种套利通常需要研究不同品种价格之间的相关性
- E. 买入近期的期货品种并卖出远期的期货品种是一种常用的跨期套利策略

参考答案：CDE

解析：A 选项错误,统计套利可以一定程度上规避系统性风险;B 选项错误,统计套利属于风险套利。

2.4. 李明,AQF,某量化分析师,获取某日市场所有股票的价格数据剔除空值后存储于变量 prices 中,数据类型为 pandas.DataFrame,部分数据如下:

	open	high	low	close
300299.SZ	4.96	5.17	4.85	5.16
300300.SZ	9.47	9.63	9.36	9.59
300306.SZ	8.67	9.01	8.51	8.94
300309.SZ	5.23	5.43	5.13	5.39
300315.SZ	3.29	3.51	3.27	3.49

其中行索引为股票代码，“open”、“high”、“low”、“close”分别为该股票该交易日的开盘价、最高价、最低价、收盘价。同时，他还获取了该日所有股票的市盈率、市净率数据剔除空值后存储于变量 `ratios` 中，数据类型为 `pandas.DataFrame`，部分数据如下：

	pe	pb
000411.SZ	31.4055	2.1083
603238.SH	42.2197	2.8465
603429.SH	81.4202	8.8701
603897.SH	17.8894	1.7972
601158.SH	18.6054	1.8742

其中行索引为股票代码，“pe”、“pb”分别为该股票该交易日的市盈率、市净率。现李明希望将这两个表格按照股票代码进行合并，并且只保留两个表格中均存在的股票数据，合并后的表格部分数据如下：

	open	high	low	close	pe	pb
300299.SZ	4.96	5.17	4.85	5.16	64.6262	2.968
300300.SZ	9.47	9.63	9.36	9.59	52.3122	2.64
300309.SZ	5.23	5.43	5.13	5.39	24.589	2.3738
300320.SZ	4.19	4.29	4.17	4.27	15.3743	1.6862
300336.SZ	3.75	3.89	3.73	3.88	98.9362	1.018

则以下代码中，可以实现该目的的代码为（ ）？

- A. `pd.concat([prices, ratios], axis=1, how='inner')`
- B. `prices.join(ratios, how='inner')`
- C. `pd.merge(prices, ratios, how='inner')`
- D. `pd.merge(prices, ratios, left_index=True, right_index=True)`
- E. `prices.append(ratios)`

参考答案：BD

解析：A 选项错误，`concat` 函数没有 `how` 参数，应该使用 `join` 参数。B 选项正确。C 选项错误，`merge` 函数默认按照相同列合并。D 选项正确。E 选项错误，`append` 方法可以在表格中添加行。

2.5. 在量化投资过程中，量化交易员往往会在实盘交易前对其交易策略进行回溯测试，即使用历史数据模拟策略的风险和盈利能力，从而评估策略的可行性。某量化交易员在进行量化交易的过程中，发现实盘收益大幅低于回测收益，则可能造成该情况的原因有（ ）？

- A. 在回测过程中，使用每日最低价作为买入的成本价
- B. 使用 2016 至 2018 年股价数据进行参数寻优，并使用 2019 年作为回测区间进行测试
- C. 使用机器学习算法进行参数优化时过于追求对训练集数据的拟合
- D. 回测过程中，执行交易前剔除股票池中当时已停牌或退市的上市公司股票
- E. 在对沪深 300 股票池进行回测的过程中考虑冲击成本

参考答案：AC

解析：A 选项，某日最低价在当日进行交易时是无法确定的，因此使用每日最低价作为买入成本价使用了未来数据，可能会高估回测收益。C 选项，过于追求对训练集数据的拟合可能会造成过拟合问题，从而降低策略对训练集外数据的适应性。

2.6. 在基金投资管理中，优秀的基金产品在于能够通过主动的投资管理，追求超越大盘的业绩表现。这说明基金投资不仅要有正收益，更要获得超越市场平均水准的超额收益。则以下哪些指标可以衡量超越市场的超额收益（ ）？

- A. 夏普比率
- B. 信息比率
- C. 最大回撤
- D. 詹森指数
- E. 特雷诺指数

参考答案：BD

解析：最大回撤不属于评价超额收益的指标，夏普比率和特雷诺指数将无风险收益作为基准。

2.7. Python3 中有六个标准的数据类型：数字、字符串、列表、元组、集合、字典，灵活使用各种数据类型是 Python 编程的重要基础。以下关于 Python 基本数据类型的说法中，正确的是（ ）？

- A. 列表与元组的区别在于，列表是有序的，而元组是无序的
- B. 由于元组属于不可变数据类型，因此无法进行加法运算
- C. 集合中的元素都是无序并且不重复的

- D. 字典的值可以取任何数据类型，但键只能是数字、字符串或元组等不可变数据类型
- E. 由于整数和浮点数属于不同的数据类型，因此需要先编写 Python 代码将其转换成同一数据类型后才可以进行算术运算

参考答案：CD

解析：A 选项错误，列表和元组都是有序数据结构，都可以通过下标索引访问序列中的值；B 选项错误，元组等序列类型都可以进行加法运算，效果为序列的组合；E 选项错误，整数和浮点数不需转换成同一数据类型就可以进行算术运算。

2.8. 以下 Python 打开文件的方法中，能够实现在文件中写入内容的操作是（ ）？

- A. `f=open('filename.txt', 'w')`
- B. `f=open('filename.txt', 'a')`
- C. `f=open('filename.txt', 'rb')`
- D. `f=open('filename.txt', 'r+')`
- E. `f=open('filename.txt', 'w+')`

参考答案：ABDE

解析：C 选项只可以实现文件读取操作。

2.9. 在量化投资的过程中，技术与多因子模型都是常用的分析方法，在量化交易策略中均有广泛应用。以下关于常见量化交易策略的说法中不正确的是（ ）？

- A. 技术分析是重要的投资分析方法之一，常见的技术指标有 MACD、KDJ、CAPM 模型等
- B. 技术分析指标 MACD 的顶背离现象通常是比较强烈的卖出信号
- C. 在多因子模型中，各因子之间的相关性越低越好
- D. 在多因子模型中，为了提高模型的准确度，有效因子的数量越多越好
- E. Fama-French 三因子模型运用上市公司的市值、账面市值比、销售额三个因子对股票回报率进行分析

参考答案：ADE

解析：A 选项，CAPM 资本资产定价模型不是技术指标；D 选项，多因子策略中有效因子数量过多可能会导致多重共线性等问题；E 选项，Fama-French 三因子模型中的三个因子分别为市场风险溢价因子、市值因子、账面市值比因子。

2.10. 以下说法正确的是（ ）？

- A. 函数调用时需要在函数名称后添加括号
- B. 函数定义中如果没有 `return` 语句，则返回结果为 `None`
- C. 函数调用时参数传入的顺序可以和定义时不同
- D. 不同的函数内部可以使用相同的变量名而不会相互影响
- E. 函数定义必须在程序开头

参考答案：ABCD

解析：函数定义不一定要在程序开头。

3. 解答题（每题 4 分，本部分总计 60 分）：按步骤、得分点给分。

3.1. 假设股票 A 和股票 B 长期以来价格走势高度一致，相关性达到 90%以上，并且在过去 3 年中股票 A 和股票 B 的平均价格比为 1.5。假设当前发现股票 A 和股票 B 的价格比为 3，请问可以采取怎样的策略获得收益？（无需代码，仅需文字描述策略逻辑）

参考答案：

买入股票 B，卖空股票 A，当价格比回归均值时对两只股票进行平仓获取收益。

3.2. 某交易员在分析某股票的过程中，通过爬虫技术获取某论坛关于某只股票的评论信息，并将所有评论以字符串的形式存储于列表 `comments` 中，部分数据如下：

`comments = ['这股票不行，买了三个月一点动静都没有。', '我看好这个股票，年后一定大涨！', '听说公司正在进行内部整合，再等等看，静观其变。', ...]`

他想筛选出不包含“好”字的评论存储到新的列表 `new_list` 中，从而研究该股票可能存在的问题，他编写了如下代码：

```
new_list=[]  
  
for comment in comments:  
    if your code here == -1:  
        your code here
```

请填补横线处缺失的代码。

参考答案：

- 1) `comment.find('好')`
- 2) `new_list.append(comment)`

3.3. 某量化交易员在进行美股分析的过程中，读取某 `json` 文件中的数据并存储于嵌套字典 `DICTION` 中，该字典包含美股代码、名称、市值、市盈率等信息，部分数据如下：

```
DICTION = {'AAPL':{'名称':'苹果','市值':'13973.21 亿美元','市盈率':25.29},  
           'MSft':{'名称':'微软','市值':'12747.73 亿美元','市盈率':31.02},  
           'AMZN':{'名称':'亚马逊','市值':'9245.23 亿美元','市盈率':81.48},  
           ...}
```

为了阅读方便，他希望将该字典中的数据打印出来查看，并且打印结果统一将美股代码修改为全部字母大写的形式，同时市盈率四舍五入至整数位，打印效果如下：

苹果公司的股票代码为 AAPL，市盈率为 25。

微软公司的股票代码为 MSFT，市盈率为 31。

亚马逊公司的股票代码为 AMZN，市盈率为 81。

.....

请写出相应代码。

参考答案：

```
for code in DICTION:
```

```
    print('{}公司的股票代码为{}, 市盈率为{}'.format(DICTION[code]['名称'],  
code.upper(), round(DICTION[code]['市盈率'])))
```

3.4. 某量化分析师在进行策略研究前需要对获取的数据进行整理。他将自己获取的某日频因子数据按时间升序保存并命名为 `data`，数据类型为 `pandas.DataFrame`，但整理数据时发现存在少量缺失值，缺失值在时间上不连续。部分数据如下

	factor
2020-01-02	1530231
2020-01-03	NaN
2020-01-06	862083
2020-01-07	728607
...	...

现该研究员决定使用缺失值前后各一个数据的平均值进行填充，请编写相应代码。

参考答案：

方法一：`data = data.interpolate()`

方法二：`data['factor'] = np.where(data['factor'].isna(),
(data['factor'].shift() + data['factor'].shift(-1))/2, data['factor'])`

3.5. 某量化分析师在研究中发现，上市公司的业绩预告通常会对股价产生影响，现获取实际数据进行验证。他整理出 2016 年第一季度至 2020 年第一季度所有季度的字符串，存储于列表 `season` 中，部分数据如下：

```
season = ['2016Q1', '2016Q2', ..., '2020Q1']
```

其中每个元素前四个字符表示年份，后两个字符表示第几季度，例如，'2016Q1'表示 2016 年第一季度。由于获取业绩预告的数据接口输入参数为年份和季度的数值，因此他希望将上述列表修改为如下形式：

```
[(2016, 1), (2016, 2), ..., (2020, 1)]
```

即将上述列表中的字符串拆分成由年份和季度组成的元组。请写出相应代码。

参考答案：

```
[(int(i[:4]), int(i[-1])) for i in season]
```

3.6. 李明，AQF，某量化研究员，在获取股票行情数据的过程中发现，从某数据接口获得的日线数据为 `pandas.DataFrame` 数据类型，部分数据如下：

	date	open	high	low	close
0	20191231	16.57	16.63	16.31	16.45
1	20191230	16.46	16.63	16.1	16.57
2	20191227	16.53	16.93	16.43	16.63
3	20191226	16.34	16.48	16.32	16.47
4	20191225	16.45	16.56	16.24	16.3

其中“date”一列为交易日，数据类型为字符串；“open”、“high”、“low”、“close”四列分别为该交易日该股票的开盘价、最高价、最低价、收盘价，数据类型为浮点型。李明现希望定义一个名为 `process_data` 的函数，可以将上述 `DataFrame` 作为原始数据进行处理后返回一个新的 `DataFrame`，处理过程中不对原始数据的 `DataFrame` 进行修改，处理操作如下：

- 1) 将交易日一列设为 `DatetimeIndex`；
- 2) 按交易日升序对 `DataFrame` 进行排序。

请写出定义该函数的代码。

参考答案：

```
def process_data(data):
    new_data = data.copy()
    new_data['date'] = pd.to_datetime(new_data['date'])
    new_data.set_index('date', inplace=True)
```

```
new_data.sort_index(inplace=True)

return new_data
```

3.7. 李明，AQF，某量化分析师，获取多只股票价格数据存储在变量 `df` 中，数据类型为 `pandas.DataFrame`，部分数据如下：

	code	date	open	high	low	close
0	000001.SZ	20191231	16.57	16.63	16.31	16.45
1	000776.SZ	20191231	15.2	15.32	15.04	15.17
2	600000.SH	20191231	12.32	12.38	12.21	12.37
3	000001.SZ	20191230	16.46	16.63	16.1	16.57
4	000776.SZ	20191230	14.45	15.44	14.4	15.3

其中“code”一列为股票代码，“date”为交易日，数据类型为字符串；“open”、“high”、“low”、“close”四列分别为该交易日该股票的开盘价、最高价、最低价、收盘价，数据类型为浮点型。李明现希望单独取出收盘价数据，并将日期作为行索引，股票代码作为列索引，修改后的部分数据如下：

code	000001.SZ	000776.SZ	600000.SH
date			
20190102	9.19	12.24	9.7
20190103	9.28	12.33	9.81
20190104	9.75	13.22	9.96
20190107	9.74	13.34	9.98
20190108	9.66	13.28	9.96

李明将修改后的收盘价数据存储于变量 `close_data` 中，请写出相应代码。

参考答案：

```
df.set_index(['date','code'], inplace=True)

close_data = df['close'].unstack()
```

3.8. 某量化分析师希望测试不同周期的技术指标结合后的策略效果，该研究员将某期货 5 分钟 K 线数据保存在 `data` 变量中，数据类型为 `pandas.DataFrame`，部分数据如下

	date	open	close	high	low
0	2020-08-04 09:35	13.66	13.61	13.67	13.56
1	2020-08-04 09:40	13.62	13.59	13.64	13.59
2	2020-08-04 09:45	13.59	13.55	13.60	13.54
...

该研究员希望分别计算出 5 分钟 K 线的 5 周期均线和日 K 线的 5 周期均线的数据，并分别保存在变量 `data_ma5_5min` 和 `data_ma5_1d`，请编写相应代码。（提示：计算均线时一般使用不同周期 K 线的收盘价进行计算）

参考答案：

```
data['date'] = pd.to_datetime(data['date'])
data.set_index('date', inplace=True)
data_ma5_5min = data['close'].rolling(5).mean()
data_ma5_1d =
data['close'].resample('1d').last().dropna().rolling(5).mean()
```

3.9. 李明, AQF, 量化分析师, 使用两只高度相关的证券进行配对交易策略的回溯测试, 他计算得到这两只证券的价差在 2016-2018 三年间的均值和标准差存储于变量 `mean` 和 `std` 中, 并使用 2019 年 1 月 1 日至 2019 年 12 月 31 日作为回测区间。他获取得到 2019 年两只证券的价格数据存储于变量 `data` 中, 数据类型为 `pandas.DataFrame`, 部分数据如下:

	price1	price2
2019/1/2	36.44	35.8
2019/1/3	36.2	35.92
2019/1/4	36.99	36.65
2019/1/7	36.99	36.48
2019/1/8	36.86	36.34

其中, 行索引为交易日, 数据类型为 `DatetimeIndex`; “`price1`”、“`price2`”分别为两只证券每日价格数据。李明认为, 当价差超过均值两倍标准差时存在错误定价, 可以进行套利交易。请分别计算两只证券的交易信号 (1 表示做多, -1 表示做空, 0 表示不持仓), 并保存于 `data` 新建列 “`position1`” 和 “`position2`” 中。请写出相应代码。

参考答案:

```
data['spread'] = data['price1'] - data['price2']
data['z-spread'] = (data['spread'] - mean)/std
data['position1'] = np.where(data['z-spread']>2,1,0)
data['position1'] = np.where(data['z-spread']<-2,-1,data['position1'])
data['position2'] = -data['position1']
```

3.10. 某量化研究员在研究过程中发现,金融资产的收益率近似服从正态分布,因此想使用实际数据进行验证。他获取某金融资产某段时间的价格数据,存储于变量 **data** 中,数据类型为 **pandas.Series**,部分数据如下:

```
date
20191231    8.48
20191230    8.36
20191227    8.28
20191226    8.19
20191225    8.19
Name: price, dtype: float64
```

其中 **data** 的索引为交易日,数据类型为字符串; **data** 的值为该金融资产每日价格,数据类型为浮点型。请写出相应代码,绘制该金融资产日收益率的频率分布直方图,并设置直方个数为 30。

参考答案:

```
data.sort_index(inplace=True)
returns = data.pct_change().dropna()
plt.hist(returns, bins=30, density=1)
```

3.11. 某量化分析师在进行量化策略实盘过程中,通过程序记录了某量化交易策略最近一年的日收益率数据,该数据储存在一个名为 **return_rate**、格式为 **csv** 格式的文件中,其部分数据如下:

date	return
------	--------

2019-01-02	0.015
2019-01-03	0.003
2019-01-04	-0.001
...	...
2019-12-31	0.037

现该研究员希望利用 Python 程序从 csv 文件中读取日收益率数据后, 计算出该段时间该策略的最大回撤率。请编写相应 Python 代码。

参考答案:

```
import pandas as pd

data = pd.read_csv('return_rate.csv')

data['strategy_index'] = (data['return'] + 1).cumprod()

data['cummax'] = data['strategy_index'].cummax()

draw_down = (1-data['strategy_index']/data['cummax']).max()
```

3.12. 某量化分析师在设计量化实盘系统时, 希望使用面向对象的技术进行实时数据的保存和记录。他希望创建一个名为 **Ticker** 的类, 该类在创建实例时需要传入两个参数, 其中 **time** 代表获取实时数据的时间, 格式为字符串; **price** 代表获取的实时成交价, 格式为浮点数。该类在创建实例时, 分别将 **time**、**price** 赋值给实例的 **time**、**price** 两个属性。请编写相应代码。

参考答案:

```
class Ticker:

    def __init__(self, time, price):

        self.time = time

        self.price = price
```

3.13. 李明, AQF, 某量化基金经理, 在进行策略研究时需要了解各行业情况, 当前已知股票数据存储于变量 **stock_data** 中, 数据类型为 **pandas.DataFrame**, 部分数据如下:

	industry	PE	ROE
stock1	industry1	30	0.05

stock2	industry1	15	0.07
stock3	industry2	5	-0.01
stock4	industry1	20	0.10
stock5	industry2	8	0.03

现要使用 `groupby` 技术从变量 `stock_data` 中取出各行业 PE 值的均值和各行业 ROE 的最大值分别储存在变量 `pe_mean` 和 `roe_max` 中，请编写相应代码。

参考答案：

```
pe_mean = stock_data.groupby('industry')['PE'].mean()
roe_max = stock_data.groupby('industry')['ROE'].max()
```

3.14. 某研究员在进行股票筛选时，希望根据季报中某些财务因子数据利用支持向量机模型筛选出问题股票。现将经过预处理的历史财务因子数据保存在变量名 `train_x` 中，数据类型为 `numpy.ndarray`，部分数据如下：

```
array([[ -0.94605487,  0.10085184,  1.25537349, -1.25166215],
       [ -1.39508908, -0.44441581,  1.24479955,  1.02083281],
       [  1.41366216,  0.45419956, -0.76307388,  1.55600654],
       [ -0.10950642,  0.27522134, -0.18718804, -1.51703105],
       [  0.96011602,  0.53424204,  0.45305838,  0.10289674]])
```

根据历史行情数据，将是否为问题股票的信息储存在变量名 `train_y` 中，数据类型为 `numpy.ndarray`，部分数据如下：

```
array([0, 1, 0, 0, 0, 1, 0])
```

`train_x` 每行数据和 `train_y` 数据一一对应。已知经过预处理的最新财务因子数据保存在变量名 `predict_x` 中，数据类型为 `numpy.ndarray`，部分数据如下：

```
array([[ 1.96135388,  0.82111093,  0.08576179,  0.1049082 ],
       [  0.44397289, -0.27871142,  0.42905546,  0.20329754],
       [  0.7493409 , -0.19221954,  2.94225732,  0.35231131],
       [  0.84939249, -0.7935356 , -1.42045482, -0.01109155],
       [  1.76566663,  0.19050908,  0.81923936, -0.69161678]])
```

请为该研究员编写代码，利用 `train_x`, `train_y` 训练支持向量机模型（核函数使用默认

的 rbf 核函数) 后使用 predict_x 数据完成预测。

参考答案:

```
from sklearn.svm import SVC  
clf_SVC = SVC()  
clf_SVC.fit(train_x,train_y)  
pred_SVC = clf_SVC.predict(predict_x)
```

3.15. 某研究员在研究某日各上市公司基本面数据时, 获取得到以下信息存储于变量 data 中, 数据类型为 pandas.DataFrame, 部分数据如下:

	code	pe	pb	ps
0	000681.SZ	56.1289	5.8711	20.0393
1	603603.SH	30.9796	4.3067	2.0553
2	600909.SH	26.0822	1.3697	8.8368
3	300194.SZ	13.3669	0.8885	1.8209
4	300152.SZ	50.0445	1.6580	2.8327

他想筛选出 pe 值小于 5 且 pb 值小于 1 的股票, 并将符合条件的股票代码以列表的形式存储在变量 chosed_stock 中, 请写出实现该目标的代码。

参考答案:

```
chosed_stock = list(data[(data['pe']<5) & (data['pb']<1)]['code'])
```