



# 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。

1. 单选题（每题 2 分，本部分共 40 分）：只有一个正确答案，选对得 2 分，选错或不选得 0 分。

1.1. 与仅投资一种资产类型相比，投资无风险资产和风险资产的投资组合具有更好的风险收益权衡，因为无风险资产和风险资产之间的相关系数等于（ ）？

- A. 1
- B. 0~1 之间
- C. 0
- D. -1

参考答案：C

解析：无风险资产的收益率是固定不变的，不受其他风险资产收益率变动的影响，所以无风险资产与风险资产之间的相关系数等于 0。

1.2. 在投资过程中，投资者通过财务报表分析可以对公司的整体财务状况有更清楚的认知，有助于进行投资决策。投资者如果担心公司的长期偿付能力，他很可能会检查公司的哪一个财务报表指标（ ）？

- A. 股本回报率
- B. 流动比率
- C. 速动比率
- D. 负债权益比

参考答案：D

解析：负债权益比率衡量公司的长期偿付能力。A 选项股本回报率为盈利指标。B 选项流动比率和 C 选项速动比率衡量公司短期偿付能力。

1.3. 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。现有一看涨期权，还有三个月到期，执行价格为 50 元。期权的标的资产为不支付股息的股票，标的资产的当前市场价格为 52 元。该看涨期权的期权费为 3 元，且不存在套

利机会，请问该期权的时间价值为（ ）？

- A. 3 元
- B. 2 元
- C. 1 元
- D. 0 元

**参考答案：C**

**解析：**期权价值由内在价值和时间价值两部分组成，其中，内在价值为期权持有者立即执行该期权合约所赋予的权利时所能获得的收益。题中看涨期权的期权价值等于期权费 3 元，内在价值为  $(52-50=2)$  元，则时间价值为  $(3-2=1)$  元。

1.4. 某交易员现有两个未完成的限价单。在正常交易时间内，第一个限价单成交的概率为 0.40，第二个限价单成交的概率为 0.25，两个订单都成交的概率为 0.15。则两个订单中至少有一个成交的概率是（ ）？

- A. 0.35
- B. 0.40
- C. 0.45
- D. 0.50

**参考答案：D**

**解析：** $0.40+0.25-0.15=0.50$

1.5. 线性回归模型通常利用判定系数（R-Squared）来判断模型拟合度。此时，如果在模型中增加一个特征，则下面哪个说法是正确的？

- A. 如果 R-Squared 增加，则这个特征有意义
- B. 如果 R-Squared 减小，则这个特征无意义
- C. 仅看 R-Squared 单一指标，无法确定这个特征是否有意义
- D. 以上说法都不对

**参考答案：C**

**解析：**无论特征是否有效，判定系数往往会随着模型特征数量的增加而增加。在样本容量一定的情况下，增加解释变量必定使得自由度减少，调整后的判定系数（Adjusted R-Squared）将残差平方和与总离差平方和分别除以各自的自由度，以剔除变量个数对拟合优度的影响。

1.6. 某量化交易员在 2 月某日发现股指期货 IF0803 合约价格为 4500 点，IF0804 合约价格为 5000 点，两合约的价差为 500 点。该交易员推测，在 IF0803 合约到期前，国家将出台利多政策，大盘将有望扭转疲弱局面，两合约价差有望进一步增大。在此预期下，他卖出一手 IF0803 合约的同时，买入一手 IF0804 合约。通过持续跟踪观察，在 IF0803 合约到期前，该交易员发现两合约的价差达到 800 点的水平，将两合约同时了结，从而获得 300 点收益。则该投资策略最有可能属于以下哪种类型？

- A. 期现套利
- B. 跨期套利
- C. 跨市场套利
- D. 跨品种套利

**参考答案：B**

**解析：**跨期套利策略是在同一期货品种的不同月份合约上建立数量相等、方向相反的交易头寸，最后以对冲或交割方式结束交易、获得收益的方式。

1.7. 下列关于量化交易策略的说法，不正确的是（ ）？

- A. 行业轮动策略是利用市场趋势获利的主动交易策略
- B. 多因子策略中，将越多的因子加入候选因子，越有可能筛选出强有效的因子
- C. Dual Thrust 系统捕捉长期信号，利用均值回归思想构建模型
- D. CCI 指标可以用来分析市场的超买超卖程度

**参考答案：C**

**解析：**Dual Thrust 是一个趋势跟踪系统。

1.8. 李明，某量化基金经理，正在研究多因子策略，信息系数（Information Coefficient, 简称 IC）常用于衡量选股因子的有效性，关于信息系数的描述错误的是（ ）？

- A. 信息系数的值越大，该因子越有效
- B. 信息系数最大值为 1，表示因子的预测能力 100% 正确
- C. 当样本股票过少时，IC 是没有统计意义的
- D. Rank IC 值是对因子值排名和下期股票收益排名求相关系数

**参考答案：A**

**解析：**信息系数的绝对值越大，该因子越有效。当信息系数为负时，值越小越有效。

1.9. 下列属于无监督学习算法的是（ ）？

- A. K 均值聚类算法
- B. 逻辑回归算法
- C. K 近邻算法
- D. 决策树算法

**参考答案：A**

**解析：**K 均值聚类算法属于无监督学习算法，逻辑回归算法、K 近邻算法、决策树算法属于监督学习算法。

1.10. 以下哪一句代码会导致 SyntaxError？

- A. `message="Once upon a time...", she said.'`
- B. `message="He said, 'Yes!'"`
- C. `message='3\'`
- D. `message="'That's ok'"`

**参考答案：C**

**解析：**Python 中字符串的最后一个字符是斜杠时，斜杠与后面的单引号结合为转义字符，从而无法构成字符串的两个单引号，会导致出错。

1.11. 李明，某量化研究员，在使用 Jupyter Notebook 编写策略代码的过程中，不小心把变量定义的单元格删掉了，他可以使用以下哪个魔法命令的功能来查看当前所有变量的类型和信息？

- A. `%matplotlib inline`
- B. `%time`
- C. `%whos`
- D. `%lsmagic`

**参考答案：C**

**解析：**`%whos` 魔法命令可以查看当前所有变量的类型和信息。

1.12. 李明，AQF，某量化基金经理，在策略研究过程中，把腾讯控股的股票信息存储在字典 `stock_info` 中，如下：

```
stock_info = {'code': '00700', 'name': '腾讯控股', 'PE': '10',
              'industry': 'Information Technology'}
```

但发现 PE 值和行业名称存在错误，所以他新建了一个字典 `modified_stock_info` 来存储正确的值，如下：

```
modified_stock_info = {'PE': '66.55', 'industry': 'Internet Content &
Information'}
```

请问以下代码哪个可以实现股票信息的修改？

- A. `stock_info.update(modified_stock_info)`
- B. `stock_info.modify(modified_stock_info)`
- C. `stock_info.append(modified_stock_info)`
- D. `stock_info.get(modified_stock_info)`

**参考答案：A**

**解析：**字典的 `update()` 方法可以将新字典的键值对更新到原字典中。

1.13. 李明，某量化基金经理，获取了某股票 2020 年的价格数据存储于变量 `stock_prices` 中，数据类型为 `pandas.DataFrame`，前五行数据如下：

	open	close	high	low
2020-01-02	25.74	25.49	25.89	25.26
2020-01-03	25.34	25.38	25.63	25.10
2020-01-06	25.06	25.32	25.85	25.00
2020-01-07	25.40	25.70	25.75	25.10
2020-01-08	25.43	24.94	25.48	24.85

在策略研究中，他希望提取出 2020 年 1 月 3 日到 1 月 7 日的“open”和“close”两列的数据。则他可以使用的代码为（ ）？

- A. `stock_prices.loc['2020-01-03', '2020-01-07', 'open', 'close']`
- B. `stock_prices.loc['2020-01-03': '2020-01-08', 'open': 'high']`
- C. `stock_prices.iloc[1:4, 0:2]`
- D. `stock_prices.iloc[1:3, 0:1]`

**参考答案：C**

**解析：**C 选项可以提取出 2020 年 1 月 3 日到 1 月 7 日的“open”和“close”两列的数据。

1.14. 李明，某量化基金经理，使用爬虫技术对微博、twitter 上的博文分析并作出投资决策。他获取的数据类型为（ ）？

- A. 价量数据
- B. 行情数据
- C. 结构化数据
- D. 非结构化数据

**参考答案：D**

**解析：**博文数据为非结构化数据。

1.15. 以下数据类型中，更新频率最慢的是（ ）？

- A. 十档行情数据
- B. 股票分时图数据
- C. 日成交量数据
- D. 公司财务报表数据

**参考答案：D**

**解析：**公司财务报表数据以季度频率更新，是选项中更新频率最慢的。

1.16. 现有变量 `data`，数据类型为 `numpy.ndarray`，具体数据如下：

```
data = np.array([0.2, 0.2, 1.3, 0.8, 0.6, 1.1])
```

现希望将该一维数组转换为列表数据类型，则可以使用的 `numpy.ndarray` 的方法为（ ）？

- A. `data.all()`
- B. `data.reshape()`
- C. `data.flatten()`
- D. `data.tolist()`

**参考答案：D**

**解析：**`numpy` 中多维数组的 `tolist()` 方法可以将数组转为列表。

1.17. 李明，某量化研究员，将某因子截面数据存储于两个列表，股票代码和因子数据分

别存储于列表 `stock_code` 和 `factor_data` 中，且相同位置索引的股票代码和因子数据相互对应。部分数据如下：

```
stock_code = ['600000', '600002', '600004', ...]
```

```
factor_data = [0.45, 0.66, 0.64, ...]
```

现希望用两个列表的信息创建一个字典，字典键为股票代码，字典值为因子数据，创建后的字典部分数据如下：

```
{'600000': 0.45, '600002': 0.66, '600004': 0.64, ...}
```

下列代码可以实现该功能的有（ ）？

- A. `data_dict = {code:factor for code, factor in zip(stock_code, factor_data)}`
- B. `data_dict = [code:factor for code, factor in zip(stock_code, factor_data)]`
- C. `data_dict = {code:factor for code, factor in stock_code, factor_data}`
- D. `data_dict = [code:factor for code, factor in stock_code, factor_data]`

参考答案：A

1.18. 李明，AQF，某量化基金经理，将某股票的历史收益率数据存储于变量 `data_pool` 中，数据类型为 `pandas.Series`，部分数据如下：

时间	收益率
2005-02-24	0.000000
2005-02-25	0.004566
2005-02-28	-0.009091
2005-03-01	-0.005734
2005-03-02	-0.009227

现在他需要在历史数据中随机抽取 300 天的日收益率计算在险价值 VaR，则他可以在提取数据时可以使用代码为（ ）？

- A. `np.random.choice(data_pool, size=300)`
- B. `np.random.randn(1, 300)`
- C. `np.random.normal(data_pool, size=300)`



D. `np.random.sample(data_pool, size=300)`

参考答案: A

解析: `np.random.choice()`函数可以从指定序列中随机抽取数据。

1.19. 李明, AQF, 某量化基金经理, 正在进行配对交易策略的回测。他选取了 `future1` 和 `future2` 两个期货品种, 将相关价格数据和持仓数据存储在变量 `data` 中, 数据类型为 `pandas.DataFrame`。其中, `future1` 和 `future2` 两列为期货每日收盘价, `future1` 的交易信号保存在列 `position_1` 中, 其中 1 表示开多仓, 0 表示平仓, -1 表示开空仓。变量 `data` 部分数据如下:

date	future1	future2	position_1
2018-01-11	9975	2890	NaN
2018-01-12	9862	2826	NaN
2018-01-15	9749	2785	NaN
2018-01-16	9408	2823	1.0
2018-01-17	9502	2865	NaN
2018-01-18	9675	2808	1.0

李明想要填补 `position_1` 一列中的缺失值, 使得在未收到下一个交易信号前保持原来的持仓, 则他可以使用的代码为 ( ) ?

- A. `data['position_1'] = data['position_1'].fillna(method = 'bfill')`
- B. `data['position_1'] = data['position_1'].fillna(method = 'ffill')`
- C. `data['position_1'] = data['position_1'].fillna(method = 'ffill', inplace = true)`
- D. `data['position_1'] = data['position_1'].fillna(method = 'bfill', inplace = true)`

参考答案: B

解析: `fillna` 方法中 `method` 参数可以指定缺失值填充方法, `method='ffill'` 缺失值使用前面的值填充。

1.20. 李明, AQF, 某量化基金经理, 在量化策略回测的过程中, 为了计算持仓信号, 他

希望可以实现某指标大于 0 的返回 1.0，小于 0 的返回 -1.0，等于 0 的返回 0.0。以下哪个函数可以实现该功能？

- A. `np.sign()`
- B. `np.dot()`
- C. `np.reshape()`
- D. `np.diff()`

参考答案：A

解析：`np.sign()`函数可以取符号。

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. 下列关于债券交易策略的说法中正确的是（ ）？

- A. 卖出凸性策略是购买凸性小的债券，卖出凸性大的债券
- B. 久期管理策略在预期利率下降时，增加组合久期获得收益
- C. 持有到期策略属于被动管理策略
- D. 骑乘收益率曲线的投资者会购买比要求的期限稍短的债券
- E. 套息交易策略的收益来源于利差

**参考答案：ABE**

**解析：**C 选项持有到期策略属于主动管理策略。D 选项骑乘收益率曲线的投资者会购买比要求的期限稍长的债券，然后在债券到期前售出，获取收益。

2.2. 某研究员正在进行日内高频交易策略的研究，以下说法不正确的是（ ）？

- A. 由于高频交易策略的交易频率非常高，因此对手续费等成本非常敏感
- B. 高频交易常用的行情数据包括分钟 K 线、盘口快照、委托队列、日成交持仓排名等
- C. 自动做市商高频交易主要是分析实时订单簿数据提供市场报价
- D. 高频交易的核心是交易速度，交易速度越快收益越高
- E. 部分高频交易策略可以缩小市场买卖价差，提高市场流动性

**参考答案：BD**

**解析：**B 选项，日成交持仓排名为低频数据。D 选项，交易速度是高频交易的核心要素，但交易速度越快不一定收益越高，还需考虑策略思路、交易成本等其他因素影响。

2.3. 在投资过程中，基本面分析和技术分析都是重要的分析方法。某量化交易员正在研究多因子策略，并整理了一系列候选因子如下，请问哪些属于基本面因子类型（ ）？

- A. 市盈率因子
- B. 价值成长因子
- C. 成交量分布因子
- D. 净资产收益率因子
- E. 量价相关性因子

**参考答案：ABD**

**解析：**基本面分析以证券的内在价值为依据，如公司营运能力、盈利能力、偿债能力等。技术分析通过分析关于股票价格和成交量的信息来获取超额收益。C 选项成交量分布因子、E 选项量价相关性因子属于价量因子，属于技术分析的范畴。

2.4. 李明定义了一个变量 `new_list = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]`，现在他希望从中过滤出所有奇数，并储存在列表 `filtered` 中，以下可以实现该功能的代码为（ ）？

- A. `filtered=[]`  
`for num in new_list:`  
`if num % 2 != 0:`  
`filtered.append(num)`
- B. `filtered=[]`  
`for num in new_list:`  
`if num % 2 == 0:`  
`filtered.append(num)`
- C. `filtered = [x for x in new_list if x % 2 != 0]`
- D. `filtered = {x for x in new_list if x % 2 != 0}`
- E. `filtered = list(filter(lambda i: i % 2 != 0, new_list))`

**参考答案：ACE**

2.5. 李明，AQF，通过历史模拟的方法计算 VaR，得出 50ETF 一日 99%和 95%的 VaR 值分别是-0.0518 和-0.0265，分别存储在变量 `VaR_1` 和 `VaR_5` 中，以下打印此结果的代码正确的是（ ）？

- A. '50ETF 一日 95%的 VaR 为{:.2f}%', 99%的  
`VaR{:.2f}%'.format(VaR_5*100,VaR_1*100)`
- B. '50ETF 一日 95%的 VaR 为%0.2f%%, 99%的 VaR 为%0.2f%%' %  
`(VaR_5*100,VaR_1*100)`
- C. '50ETF 一日 95%的 VaR 为%0.2f%%, 99%的 VaR 为%0.2f%%' %  
`(VaR_5*100,VaR_1*100)`
- D. `f'50ETF 一日 95%的 VaR 为{VaR_5*100:.2f}%', 99%的 VaR 为{VaR_1*100:.2f}%'`

E. f'50ETF 一日 95%的 VaR 为 $\{VaR_5*100:.2f\}%%$ , 99%的 VaR 为 $\{VaR_1*100:.2f\}%%$ '

参考答案: ACD

2.6. 李明, 某量化基金经理, 把某股票一周的收益率存储在变量 `arr` 中, `arr = np.array([0.05, -0.13, 0.19, 0.20, -0.03])`, 他现在想要判断这一周的股票收益率是否都大于 0, 如果都大于 0 返回 `True`, 否则返回 `False`。可以实现该功能的代码为 ( ) ?

- A. `np.all(arr>0)`
- B. `np.any(arr>0)`
- C. `np.where(arr>0)`
- D. `np.sign(arr>0)`
- E. `np.all(np.where(arr>0, True, False))`

参考答案: AE

解析: A 选项, `np.all()` 函数判断给定轴向上的所有元素是否都为 `True`, 符合要求。E 选项, `np.where(condition, x, y)` 满足条件 `condition` 输出 `x`, 否则输出 `y`, 则 `np.where(arr>0, True, False)` 输出 `arr>0` 条件判断的结果, `np.all()` 再次判断条件判断的结果是否都为 `True`。

2.7. 数据清洗是量化策略开发过程中不可缺少的一个环节, 其结果质量直接关系到策略回测的准确度。因此, 在数据分析之前, 研究员往往会花费大量的时间来进行数据清洗工作。

以下数据清洗的做法, 不正确的有 ( ) ?

- A. 数据录入过程、数据整合过程都可能会产生重复数据, 可以直接删除
- B. Pandas 中可以使用 `fillna` 方法替换缺失值数据
- C. 可以使用某个变量的样本均值、中位数或众数代替无效值和缺失值
- D. 对于量级相差大的数据, 直接建立线性回归模型拟合数据
- E. 对于数据中的离群值, 单独进行标准化处理

参考答案: DE

解析: D 选项, 对于量级相差大的数据, 应先进行标准化处理。E 选项, 不能单独对数据中的离群值进行标准化处理, 可以采用删除或者离群值处理的方法。

2.8. 某研究员在进行量化交易策略研究过程中，获取某股票的 k 线数据存储在变量 `k_data` 中，数据类型为 `pandas.DataFrame`，部分数据如下：

	date	open	close	high	low	volume	code
0	2018-01-02	17.00	17.47	18.39	16.91	113042.0	002577
1	2018-01-03	17.41	17.25	17.42	17.01	65463.0	002577
2	2018-01-04	17.03	17.00	17.44	16.98	49592.0	002577
3	2018-01-05	17.01	17.16	17.37	17.01	40664.0	002577
4	2018-01-08	17.16	17.03	17.16	16.70	34389.0	002577

他想删除“volume”和“code”两列，只保留日期和高开低收价格数据，则以下代码中可以实现该目标的是（ ）？

- A. `del k_data[-2:]`
- B. `del k_data[['volume','code']]`
- C. `k_data = k_data.iloc[:,5]`
- D. `k_data.drop(['volume','code'],inplace=True)`
- E. `k_data.drop(['volume','code'],axis=1,inplace=True)`

参考答案：CE

解析：

A 选项和 B 选项运行时都会报错，错误类型为 `TypeError`；

D 选项错误，使用 `DataFrame` 的 `drop` 方法删除列时需设置参数 `axis=1`。

2.9. 在量化交易策略的回测过程中，李明发现回测收益往往会出现高于实盘收益，可能导致这种情况的原因有（ ）？

- A. 使用了未来函数
- B. 策略模型过拟合或过度优化
- C. 忽视了策略流动性不足的问题
- D. 在交易过程中考虑到交易成本
- E. 存在幸存者偏差

参考答案：ABCE

2.10. 李明，AQF，某量化基金经理，在多因子策略实现过程中，计算出某股票的日收益率和日收益率的标准差分别存储于变量 `returns` 和 `risk` 中，数据类型为 `pandas.DataFrame`，部分数据如下：

	return		risk
20200102	0.04	20200113	0.51
20200103	0.07	20200114	0.53
20200106	0.03	20200115	0.49
20200107	0.10	20200116	0.53
20200108	0.02	20200117	0.48

他希望将两个 `DataFrame` 按相同日期合并，并且只合并两表均有的日期数据，从而合并后的表格中没有空值，合并后的 `DataFrame` 部分数据如下：

	return	risk
20200113	0.02	0.51
20200114	0.01	0.53
20200115	0.04	0.49
20200116	0.05	0.53
20200117	0.02	0.48

则以下可以实现此功能的代码为（ ）？

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

参考答案：CD

3. 解答题（每题 4 分，本部分总计 40 分）：按步骤、得分点给分。

3.1. 假设去年某公司支付每股股利为 1.80 元，预计在未来日子里该公司股利按每年 5% 的速率增长。假定要求收益率为 11%，目前该公司的股票每股价格为 40 元，你建议当前持有该股票的投资者出售股票吗？（要求写出分析步骤）

参考答案：

利用 DDM 永续增长模型，该公司股票估值为  $1.80 \times (1+0.05) / (0.11-0.05) = 31.50$  元，当前股票价格为 40 元，因此股票被高估，建议持有该股票的投资者出售该股票。

3.2. 在多因子模型的构件中，单因子有效性的检验是至关重要的步骤。请写出至少三个单因子有效性检验的方法，并就其中某一方法做详细解释。

参考答案：

1) 回归法、因子 IC 值、因子单调性

2) 因子单调性：选取三年回测区间，以月为单位分为 36 期，每一期期初将股票池中的股票按照因子值的大小从低到高分 10 组，月末计算每一组的月收益率，重复 36 个月，计算每组的月均收益率，如果各组表现单调性越强，说明该因子选股能力越强。

3.3. 李明，AQF，某量化基金经理，在数据清洗的过程中，利用  $3\sigma$  原则进行因子离群值数据的处理，即：数值分布在  $(\mu-3\sigma, \mu+3\sigma)$  区间之外的视为离群值，其中  $\sigma$  代表标准差， $\mu$  代表均值。李明搜集了股票池中所有股票的换手率（%）和市盈率因子值，存储于变量 data 中，数据类型为 pandas.DataFrame，其中 index 为股票代码，部分数据如下：

	turnover_rate	pe
600000	0.34	5.38
600015	0.48	5.23
600016	0.21	4.52
600036	0.30	14.32
600908	7.68	10.43

为了方便之后批量处理因子离群值数据，李明希望定义一个函数 three\_sigma。该函数有两个参数，函数参数 data 为如上待处理的因子值数据表，数据类型为 pandas.DataFrame，函数参数 factor 为待处理的因子名称，数据类型为字符串。该函数可以对数据表中的某个



因子数据进行离群值处理的操作，即超过上限的数值用上限值替换，超过下限的数值用下限值替换，并将修改后的数据保存于新的一列中，该函数最终返回一个新的 DataFrame。例如，调用函数 `three_sigma(data, 'pe')` 将返回如下结果（部分数据）。

	turnover_rate	pe	new_pe
600000	0.34	5.38	5.38
600015	0.48	5.23	5.23
600016	0.21	4.52	4.52
600036	0.30	14.32	14.32
600908	7.68	10.43	10.43

请写出函数 `three_sigma` 的定义代码。

**参考答案：**

```
def three_sigma(data, factor):
    up = data[factor].mean() + 3 * data[factor].std()
    low = data[factor].mean() - 3 * data[factor].std()
    data[f'new_{factor}'] = np.where(data[factor] > up, up,
data[factor])
    data[f'new_{factor}'] = np.where(data[factor] < low, low,
data[f'new_{factor}'])
    return data
```

3.4. 李明，AQF，某量化基金经理，在策略研究的过程中，生成了一系列日期数据 `dates=pd.Series(['2020年2月1日', '2020年2月2日', '2020年2月3日', ...])`，但是日期的数据类型为字符串，不方便之后时间序列数据的处理，请你帮他把日期格式改为 `datetime` 格式。

**参考答案：**

```
dates = pd.to_datetime(dates, format='%Y年%m月%d日')
```

3.5. 李明，AQF，某量化基金经理，在策略研究过程中，他认为股票的换手率是影响股票价格振幅的重要因素，他获取到某日股票池横截面数据，并将股票换手率和股票价格振幅

数据分别存储于变量 `turnover_rate` 和 `price_range` 中，数据类型为 `pandas.Series`，请你绘制散点图描述这两个变量之间的关系（注明横纵轴刻度、横纵轴标签、标题）。

**参考答案：**

```
import pandas as pd

import matplotlib.pyplot as plt

plt.scatter(turnover_rate, price_range)

plt.xlabel('turnover_rate')

plt.ylabel('price_range')

plt.title('Range-Turnover Scatter')
```

3.6. 李明，AQF，某量化基金经理，正在进行某量化策略的策略回测。他获取得到某股票的 K 线数据，存储在变量 `k_data` 中，数据类型为 `pandas.DataFrame`，部分数据如下：

	open	close	high	low
2019-01-02	5.00	5.01	5.02	4.93
2019-01-03	5.03	5.09	5.10	4.98
2019-01-04	5.10	5.13	5.16	5.06
2019-01-07	5.16	5.16	5.20	5.11
2019-01-08	5.18	5.20	5.21	5.10

他现想计算出收盘价 20 日的滚动波动率，作为每日波动率的参考值，并将计算的结果存储在变量 `k_data` 新的一列“std”中，处理后的 `DataFrame` 部分数据如下：

	open	close	high	low	std
2019-01-28	5.16	5.15	5.19	5.15	NaN
2019-01-29	5.14	5.12	5.15	5.09	0.044988
2019-01-30	5.14	5.14	5.17	5.12	0.030088
2019-01-31	5.16	5.16	5.19	5.13	0.025189
2019-02-01	5.17	5.17	5.19	5.15	0.023946

请写出可以实现该目标的代码。

**参考答案：** `k_data['std']=k_data['close'].rolling(20).std()`

3.7. 李明，AQF，某量化基金经理，在策略研究的过程中，获取了平安银行 2019 年全年的股票日 K 线价格数据存储在变量 `stock_data` 中，数据类型为 `pandas.DataFrame`。其中包括 “open”，“close”，“high”，“low” 四列，Index 为日期。现在他希望筛选出所有开盘价比前日收盘价跌幅超过 2% 的日期并且输出。请你写出可以实现该功能的代码。

**参考答案：**

```
stock_data[(stock_data['open']-
stock_data['close'].shift())/stock_data['close'].shift()<=-0.02].index
```

3.8. 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，他把股票信息存储于变量 `stock_data` 中，数据类型为 `pandas.DataFrame`，部分数据如下：

	pe	industry
stockA	38.45	Information Technology
stockB	47.34	Information Technology
stockC	74.49	Insurance
stockD	28.54	Insurance

现在请你使用 `groupby` 技术，找出 `pe` 均值最小的行业，并且打印该行业所有股票数据。

**参考答案：**

```
pe_data = stock_data.groupby('industry')['pe'].mean()
pe_min_industry = pe_data.idxmin()
pe_min_industry_data= stock_data[stock_data['industry']==pe_min_industry]
```

3.9. 李明，AQF，某量化基金经理，在量化回测的过程中，获取了某股票的每日价格数据，并将其存储在变量 `data` 中，数据类型为 `pandas.DataFrame`，部分数据如下：

date	open	close
2005-02-23	0.891	0.876
2005-02-24	0.876	0.876
2005-02-25	0.877	0.880
2005-02-28	0.878	0.872

现请帮助该研究员编写一段代码用于计算该股票的上涨的天数和上涨的概率，并打印出来。

**参考答案：**

```
data['up']=data['close']-data['open']  
print('上涨天数为: ', len(data[data['up'] > 0]))  
print('上涨概率为: ', len(data[data['up'] > 0])/len(data))
```

3.10. 李明，AQF，某量化基金经理，正在研究动量策略。为了优化策略，他把高频数据用于动量策略，他获取某只股票的 k 线数据，计算出 `position` 和 `returns` 存储于 `data_hf` 变量中，数据类型为 `pandas.DataFrame`，部分数据如下：

	close	position	returns
2020-12-22 13:45	54.6	0	NaN
2020-12-22 13:50	54.9	1	0.005468
2020-12-22 13:55	56.3	1	0.025509
2020-12-22 14:00	55.7	0	-0.01034

其中，`DataFrame` 的行索引为 `DatetimeIndex` 数据类型，“close”列为该股票每日收盘数据，“position”列为策略持仓信号，如果是 1 表示满仓持有该股票，0 表示无持仓，李明现希望计算该动量策略的累计收益，并存储于变量 `data` 新的一列“`cum_return`”中，请写出相应代码。

**参考答案：**

```
data_hf['strategy'] = data_hf['position']*data_hf['returns']  
data_hf.fillna(0,inplace=True)  
data_hf['cum_return']=(1+data_hf['strategy']).cumprod()
```