

## 4.4 作业

2023 年 9 月 18 日

### 作业-Pandas

#### 1.DataFrame 的操作 (1 分 +1 分 +1 分 +1 分 =4 分)

要求:

- 创建一个 pandas.DataFrame 对象, 存储以上元素;
- 添加一个新列为'd', 对应行 1, 2, 3 的值为 13, 14, 15;
- 使用选取操作, 选取'a' 和'd' 列;
- 使用切片操作, 选取第 0 行和第 1 行;

[ ]:

#### 2. 描述性统计方法 (2 分)

内容: 描述性统计、数据集中度量、数据离散状况。

要求:

使用 pandas.read\_csv 读取路径"datasets/world-happiness-report-2021.csv" 文件, 对"East Asia" 东亚地区进行描述性统计分析, 包括: - 最大值 - 最小值 - 均值 - 标准差 - 分位数

[ ]:

#### 3. 商业场景: 巧克力品牌的市场占有率 (1 分 +1 分 +1 分 +1 分 =4 分)

市场份额亦称“市场占有率”。指某企业某一产品（或品类）的销售量（或销售额）在市场同类产品（或品类）中所占比重。反映企业在市场上的地位。通常市场份额越高，竞争力越强。

基本测算方法:

(1) 总体市场份额, 指某企业销售量（额）在整个行业中所占比重。

- (2) 目标市场份额，指某企业销售量（额）在其目标市场，即其所服务的市场中所占比重。
- (3) 相对市场份额，指某企业销售量与市场上最大竞争者销售量之比，若高于 1，表明其为这一市场的领导者。

样本数据文件'Chocolate.csv' 从 JD 平台采集，并清洗后的数据集，包含每个商品的价格、评价数、品牌、巧克力类别。因为销量数据涉及到商业机密和合规问题，电商平台只会披露类似“10 万+”、“1 千+”甚至不提供。

通常这种情况下，可以使用“评价数”这个指标，作为代理变量，表示销量的高低。

要求：1. 使用 pandas 库来读取相关的数据集文件：'datasets/Chocolate.csv' 2. 估算德芙巧克力的价格区间（最小值、最大值）3. 估算德芙巧克力的总体市场份额 4. 计算各个品牌其哪种口味最畅销

[ ]:

### 3. 金融场景：夏普比率（1 分 +1 分 +2 分 =4 分）

夏普比率（Sharpe Ratio）是用来衡量投资组合表现的指标，它是投资组合超额收益率与风险之间的比率。其公式如下：

$$SharpeRatio = (R_p - R_f) / \sigma_p$$

其中， $R_p$  表示投资组合的平均收益率； $R_f$  表示无风险利率； $\sigma_p$  表示投资组合收益率的标准差。

夏普比率越高，表示投资组合每承担一单位的风险能够获得更高的超额收益，因此更为优秀。需要注意的是，夏普比率只适用于对正态分布收益率的投资组合进行评估，而不适用于对非正态分布的投资组合进行评估。此外，夏普比率也不能作为唯一的评估指标，还应该结合其他指标来全面评估投资组合的表现。

IBM 的股票代码为 IBM，标普指数的代码为 SP，可口可乐的股票代码是 KO，三者的月度收益率序列存储在'datasets/ibmspko.csv' 文件中。- 请使用 pandas 读取收益率序列 - 计算持有可口可乐（代码 KO）的夏普比率是多少？ - 计算持有三者总价值为 1,000,000\$ 的股票策略，不考虑无风险利率的情况下，其夏普比率是多少？

```
[21]: import pandas as pd
df = pd.read_csv('datasets/ibmspko.csv', index_col=0)
df
```

```
[21]:          ibm          sp          ko
date
```

1961-01-31	0.072513	0.063156	0.009331
1961-02-28	0.062500	0.026870	0.103236
1961-03-30	0.029630	0.025536	0.012291
1961-04-28	0.027338	0.003843	-0.050000
1961-05-31	0.027521	0.019139	0.087719
...	...	...	...
2011-08-31	-0.050536	-0.056791	0.035877
2011-09-30	0.017218	-0.071762	-0.034351
2011-10-31	0.055813	0.107723	0.011249
2011-11-30	0.022315	-0.005059	-0.009075
2011-12-30	-0.021915	0.008533	0.040756

[612 rows x 3 columns]

[ ]: