

## 22-23-2 学期 Python 数据分析复习提纲

### ■ 填空题与单选题：10+10 每题 1.5 分共 30 分

熟练掌握测试 1、2、3、4 中共 20+15+15+20=70 题的内容  
试题大多基于或者直接为原测试题

### ■ 简答题 4：每题 5 分共 20 分

字符串格式化的三种方法（PPT S1-41，Notebook 基础 5，给出整数、浮点数、字符数，能够使用三种方式格式化为指定的表达式）

什么是可变对象和不可变对象，举几种类型的例子（PPT S1-65）

try 语句的结构和执行过程（PPT S1-91,92）

~~with 语句的结构和执行过程（PPT S1-99,100）~~

模块导入的几种方式（PPT S2-17）

Django 项目与应用的关系（PPT S2-29）

~~如何使用通用显示视图（PPT S2-71~75，urls.py，views.py，templates）~~

Numpy 数组的五种属性及其关系（PPT S3-5）

可视化常见的图形和对应的应用场景，以及 Matplotlib 对应的绘制函数（PPT S3-53~59）

### ■ 计算题 4：每题 5 分共 20 分

（Notebook Numpy 与 PPT S3）

#### ✧ 基本技能

创建数组：array, zeros, ones, arange, full, eye, identity（Notebook 3,4,6, PPT 6,7,9）  
*从数组 全 0 全 1 范围 全？ 对称 对称矩阵*

数组重塑：reshape（PPT 22）  
*方阵*

索引和切片：低维数组的索引切片（PPT 17，Notebook 11）、布尔索引（Notebook 13）、神奇索引（Notebook 14），参考测验

#### ✧ 重点掌握以下三种类型的计算

（1）数组统计运算（PPT 37，Notebook 19）

✓ 考察两种运算（sum 与 mean），特别注意 axis 参数的用法

注意 mean 函数结果为浮点数

练习：np.arange(20).reshape(4, 5).sum(1)

np.arange(20).reshape(4, 5).sum(0)

✓ （2）数组排序（PPT 39，Notebook 21）

特别注意 axis 参数的用法

练习：arr = np.random.rand(4, 5), arr.sort(0), arr.sort(1)

✓ （3）数组算术运算

包括两种情形：当形状相同时，直接逐元素运算（PPT 13，Notebook 8）

当形状不同时，先广播再逐元素运算（PPT 23~26）

arr+brr 三步解题法：先确定结果数组的形状，再将 arr 和 brr 补成结果的形状，最后逐元素计算

练习：PPT 26 前 3 个蓝色

np.arange(24).reshape(3, 4, 2) + np.arange(6).reshape(3, 1, 2)

*Warning: 切片与索引存在的细微差异*

0 1	8 9	16 17	0 1	2 3	4 5
2 3	10 11	18 19	+ 向维度1广播后逐元		
4 5	12 13	20 21			
6 7	14 15	22 23			

0 2	10 12	20 22
2 4	12 14	22 24
4 6	14 16	24 26
6 8	16 18	26 28

0 1	8 9	16 17	0	+ 向维度0,2广播
2 3	10 11	18 19	1	
4 5	12 13	20 21	2	
6 7	14 15	22 23	3	
0 1	8 9	16 17	=	
3 4	11 12	19 20		
6 7	14 15	22 23		
9 10	17 18	25 26		

0 1 | 2 3 | 4 5 +

0	1
2	3
4	5
6	7
8	9

↓ 向1轴广播

→ 向0轴广播

0	1	2	3	4	5
0	1	2	3	4	5
0	1	2	3	4	5
0	1	2	3	4	5
0	1	2	3	4	5

+

0	1	0	1	0	1
2	3	2	3	2	3
4	5	4	5	4	5
6	7	6	7	6	7
8	9	8	9	8	9

[[0	2]	[[2	4]	4	6
2	4	4	6	6	8
4	6	6	8	8	10
6	8	8	10	10	12
8	10]	10	12]	12	14

伟大的生命  
不论以何种形式  
将在宇宙中永存  
:)

可广播: 两个维度同长或其中一个的长度为1

若两个矩阵不同维, 则numpy会相应的补1

结果为两个矩阵每一个维度长度的最大值

numpy中的五种属性:

ndim 维度

shape 形状

size 总个数

dtype 数据类型

itemsize 元素大小

matplotlib 常见的图形:

(条形图)

折线图: 趋势 plot

直方图: 差异(方块图) hist

箱图: 差异 barh, bar

饼图: 占比 pie

散点图: 分布, 相关性 scatter

箱形图: 识别异常 boxplot

```
np.arange(24).reshape(3, 4, 2) + np.arange(4).reshape(4, 1)
np.arange(6).reshape(3, 1, 2) + np.arange(10).reshape(5, 2)
```

#### ■ 程序题 6: 每题 5 分共 30 分

简单编程 (2 个)

字符串格式化 (同简答题, 会用, Notebook 基础 5)

基本控制结构 (Notebook 基础 14~17)

元组、列表、字典的使用 (Notebook 基础 20, 21, 22, 23, 25, 26)

包括索引切片、新增元素、字典默认值

字符串的常用方法, 包括 title、join、split (如 split 和 join 组合使用)

函数基本用法 (Notebook 基础 44, 46, 48)

文件操作 (Notebook 基础 79)

生成器 (Notebook 基础 62): 斐波那契数列、阶乘

类 (Notebook 基础 71, 73, 实例变量与方法) 与类的继承 (Notebook 基础 75)

✓ 1 单次随机漫步 (Notebook Numpy 26, 27)

✓ 1 折线图的绘制和标注: 添加图片标题, x 轴名称, y 轴名称, 设置线的标记和类型 (Notebook 绘图 1.2 设置标签和图例 In [11], 1.4 设置颜色、标记和线类型 In [22], In [23])

图片标题 plt.title('')

x, y 轴名称 plt.xlabel('')

线样式: plt.plot(, color='', marker='', linestyle='')

linestyle: — 实线, -- 长虚线, -. 点划线, : 短虚线

marker: o 实心圆圈, . 点, s 正方形, \* 星形

v 倒三角, ^ 正三角, < 左三角, > 右三角

+ 加号, p 五边形, q 八边形, h 六边形

H 六边形 2, D 梯形

图例: legend(['']).

## 单次随机漫步:

1. random 库解法:

```
import random
```

```
distance = [0]
```

```
for i in range(1000):
```

```
    distance.append(distance[-1] + random.choice([-1, 1]))
```

```
print(min(distance), max(distance), distance.index(30))
```

2. 使用 numpy 求解:

```
import numpy as n
```

```
steps = n.random.randint(0, 2, (1000,))
```

```
steps = n.where(steps == 0, -1, 1)
```

```
distance = steps.cumsum(axis=0)
```

```
print(min(distance), max(distance))
```

```
morethan30 = n.where(n.abs(distance) >= 30)
```

```
print(morethan30.argmax())
```