一. 单选题（共6题，24分）

1. (单选题)下列选项中，用于实现数据可视化功能的库是（ ）。

A. NumPy

B. pandas

C. Matplotlib

D. NLTK

我的答案: C:Matplotlib;正确答案: C:Matplotlib;

4分

2. (单选题)下列选项中，用来表示数组维度元组的属性是（ ）。

A. ndim

B. shape

C. size

D. dtype

我的答案: B:shape;正确答案: B:shape;

4分

3. (单选题)下列关于Series和DataFrame的描述，错误的是（ ）。

A. Series是一维的数据结构

B. DataFrame是二维的数据结构

C. Series既有行索引又有列索

D. Series和DataFrame都可以重置索引

我的答案: C:Series既有行索引又有列索;正确答案: C:Series既有行索引又有列索;

4分

4. (单选题)下列选项中，可以删除缺失值的是（ ）。

A. isnull()

B. dropna()

C. notnull()

D. fillna()

我的答案: B:dropna();正确答案: B:dropna();

4分

5. (单选题) 在Matplotlib中，用于展示图表的函数是（

A. plot()

B. bar()

C. hist()

D. show()

我的答案: D:show();正确答案: D:show();

4分

6. (单选题)下列选项中，说法错误的是。（ ）

A. distplot()函数用于绘制直方图和核密度估计曲线

B. distplot()函数默认不使用密度观察条

C. stripplot()函数用于绘制直方图

D. swarmplot()函数用于绘制数据点不重叠的散点图

我的答案: C:stripplot()函数用于绘制直方图;正确答案: C:stripplot()函数用于绘制直方图;

4分

二. 判断题（共9题，36分）

7. (判断题)如果要卸载指定环境中的包，则可以直接使用conda remove命令移除。

A. 对

B. 错

我的答案: 对正确答案: 对

4分

8. (判断题)若使用ones()函数创建数组时没有指明元素的数据类型，默认数据类型为float64。

A. 对

B. 错

我的答案: 对正确答案: 对

4分

9. (判断题)DataFrame的每列数据可以看作一个Series类的对象。

A. 对

B. 错

我的答案: 对正确答案: 对

4分

10. (判断题)Matplotlib默认支持中文显示。

A. 对

B. 错

我的答案: 错正确答案: 错

4分

11. (判断题)3σ原则只适用于符合或近似正态分布的数据集。（ ）

A. 对

B. 错

我的答案: 对正确答案: 对

4分

12. (判断题)直方图与柱形图没有任何区别。（ ）

A. 对

B. 错

我的答案: 错正确答案: 错

4分

13. (判断题)饼图在识别异常值方面有一定的优越性。（ ）

A. 对

B. 错

我的答案: 错正确答案: 错

4分

14. (判断题)停用词表中可以随意添加任何词。（ ）

A. 对

B. 错

我的答案: 错正确答案: 错

4分

15. (判断题)当使用stack()方法将有分层索引的DataFrame 进行重塑操作后，一定会返回一个Series 类的对象。（ ）

A. 对

B. 错

我的答案: 错正确答案: 错

4分

三. 填空题（共10题，40分）

16. (填空题)

分析下列程序的输出结果是 。

import numpy as np

arr\_2d = np.array([[11, 20, 13],[14, 25, 16],[27, 18, 9]])

print(arr\_2d[1, :1])

我的答案：

4分

(1) [[14]]

正确答案：

(1) [14]

17. (填空题)

分析下列程序的输出结果是 。

import numpy as np

import pandas as pd

ser\_obj = pd.Series([4, np.nan, 6, 5, -3, 2])

ser\_obj.sort\_values()

我的答案：

0分

(1) 1

正确答案：

(1)

18. (填空题)

分析下列程序的输出结果是 。

from scipy.linalg import \*

import numpy as np

A=np.matrix('[1,2;3,4]')

print(A.T)

我的答案：

0分

(1)

[[1 3]

[2 4]]

正确答案：

(1)

19. (填空题)

分析下列程序的输出结果是 。

import jieba

sentence = '人生苦短，我用Pyhton'

terms\_list = jieba.cut(sentence, cut\_all=True)

print(' '.join(terms\_list))

我的答案：

4分

(1) 人生苦短，我用Pyhton

正确答案：

(1) 人生 苦短 ， 我 用 Pyhton

20. (填空题)pandas中缺失值一般使用None或np.nan表示，统一标记为\_\_\_\_\_\_\_\_。

我的答案：

4分

(1) NaN

正确答案：

(1) NaN

21. (填空题)\_\_\_\_\_\_\_\_ ()方法用于将DataFrame的列索引旋转为行索引。

我的答案：

4分

(1) pivot()

正确答案：

(1) pivot()

22. (填空题)\_\_\_\_\_\_\_\_是指数据集中的个别值明显偏离它所属数据集的其余值。

我的答案：

4分

(1) 异常值

正确答案：

(1) 异常值

23. (填空题)

from pandas import Series

import pandas as pd

from numpy import NaN

series\_obj = Series([None, 4, NaN])

pd.isnull(series\_obj)

执行上述程序后，最终输出的结果为（ ）。

我的答案：

4分

(1) 0 True 1 False 2 True

正确答案：

(1)

0 True

1 False

2 True

24. (填空题) 由一系列高度不等的矩形条或线段组成，用于反映数据的分布和波动情况。

我的答案：

4分

(1) 直方图

正确答案：

(1) 直方图

25. (填空题)Matplotlib的 模块封装了快速绘制图表的函数。

我的答案：

4分

(1) pyplot

正确答案：

1. pyplot

1. (填空题)

在下列程序中填写代码。

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

fig,subs=plt.subplots(2,2)

x=np.arange(1,100)

subs[0][0]. (x,x\*x) # 第1个图：折线图

subs[0][1]. (np.arange(0,10), np.random.rand(10)) # 第2个图：散点图

plt.show()

我的答案：

6.6分

(1) plot

正确答案：

(1) plot;scatter

2. (填空题)

在下列程序中填写代码。

import pandas as pd

import numpy as np

# 对生成的四列随机数的数据类型绘制条形图

df1= pd. (np.random.rand(10,4),columns=['a','b','c','d'])

df1.

我的答案：

6.6分

(1) DataFrame，plot.bar()

正确答案：

(1) DataFrame;plot.bar()

3. (填空题)

from sklearn.datasets import load\_iris #加载数据集

iris=load\_iris()

n\_samples,n\_features= iris.data.shape

print( ) #(150, 4) 表示为150个样本，4个特征

print(iris.target.shape) #(150,)

print("特征值的名字:\n", ) #特征名称

print("鸢尾花的数据集描述:\n",iris['DESCR']) #数据描述

我的答案：

6.6分

(1) iris.data.shape，iris.feature\_names

正确答案：

(1) iris.data.shape;iris.feature\_names

4. (填空题)

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

# 生成100个成标准正态分布的随机数

x = np.random. (size=100)

# kde=True，进行核密度估计

sns. (x,kde=True) #密度曲线KDE

plt.show()

我的答案：

6.6分

(1) normal;distplot

正确答案：

(1) normal;distplot

5. (填空题)

词形归一化包括词干提取和\_\_\_\_\_\_\_\_，它们会将派生词转化为基本形式。

我的答案：

6.6分

(1) 词形还原

正确答案：

(1)

词形还原

6. (填空题)\_\_\_\_\_\_\_\_是指为单词标注一个正确的词性。

我的答案：

6.7分

(1) 词性标注

正确答案：

(1) 词性标注

7. (填空题)\_\_\_\_\_\_\_\_是在处理自然语言文本之前或之后会自动过滤掉的没有具体意义的字或词。

我的答案：

6.7分

(1) 停用词

正确答案：

(1)

停用词

8. (填空题)\_\_\_\_\_\_\_\_研究的是计算机怎么模拟或实现人类的学习行为。

我的答案：

6.7分

(1) 机器学习

正确答案：

(1) 机器学习

9. (填空题)

数据集的每一行数据称为一个\_\_\_\_\_\_\_\_。

我的答案：

6.7分

(1) 样本

正确答案：

(1) 样本

10. (填空题)\_\_\_\_\_\_\_\_是利用一组已知类别的样本作为训练集，调整分类器的参数，使其达到所要求性能的过程。

我的答案：

6.7分

(1) 监督学习

正确答案：

(1) 监督学习

二. 判断题（共5题，33.5分）

11. (判断题)假设数据集有10行数据，说明它里面有10个样本。（ ）

A. 对

B. 错

我的答案: 对正确答案: 对

6.7分

12. (判断题)特征是机器学习最终要预测的结果。（ ）

A. 对

B. 错

我的答案: 错正确答案: 错

6.7分

13. (判断题)KNeighborsClassifier类封装了线性回归的相关功能。（ ）

A. 对

B. 错

我的答案: 错正确答案: 错

6.7分

14. (判断题)为了消除评价指标之间的量纲影响，需要对数据进行归一化处理。（ ）

A. 对

B. 错

我的答案: 对正确答案: 对

6.7分

15. (判断题)

超参数的值默认是最好的，无需进行调参处理。（ ）

A. 对

B. 错

我的答案: 错正确答案: 错

1. (填空题)在NumPy中，可以使用数组对象\_\_\_\_\_\_\_\_执行一些科学计算。

我的答案：

5分

(1) ndarray

正确答案：

(1) ndarray

2. (填空题)如果ndarray.ndim执行的结果为2，则表示创建的是\_\_\_\_\_维数组。

我的答案：

5分

(1) 二

正确答案：

(1) 二

3. (填空题)

NumPy中\_\_\_\_\_\_\_\_是由一个类型名和元素位长的数字组成。

我的答案：

5分

(1) 数据类型

正确答案：

(1) 数据类型

4. (填空题)

如果两个数组的形状不同，则它们进行算术运算时可能会触发\_\_\_\_\_\_\_\_机制。

我的答案：

5分

(1) 广播

正确答案：

(1) 广播

5. (填空题)

NumPy中\_\_\_\_\_\_\_\_是指用整数数组或整数列表作为数组的索引。

我的答案：

5分

(1) 花式索引

正确答案：

(1) 花式索引

6. (填空题)

由一系列高度不等的矩形条或线段组成，用于反映数据的分布和波动情况。

我的答案：

5分

(1) 直方图

正确答案：

(1) 直方图

7. (填空题)箱形图利用最小值、第一四分位数、 、第三四分位数和最大值­这几个统计量描述一组数据。

我的答案：

5分

(1) 1．中位数

正确答案：

(1)

1．中位数

8. (填空题)

Matplotlib的 模块封装了快速绘制图表的函数。

我的答案：

5分

(1) pyplot

正确答案：

(1) pyplot

9. (填空题) 二维直方图主要用于显示落在六边形区域内的观察值的计数。

我的答案：

5分

(1) 六边形

正确答案：

(1) 六边形

10. (填空题)

是一个用于生成Echarts图表的开源库。

我的答案：

5分

(1) Pyecharts

正确答案：

(1) Pyecharts

二. 判断题（共10题，50分）

11. (判断题)

通过empty()函数创建的数组，该数组中没有任何的元素。（ ）

A. 对

B. 错

我的答案: 错正确答案: 错

5分

12. (判断题)

若使用ones()函数创建数组时没有明确地指明元素的数据类型，则默认数据类型为float64。（ ）

A. 对

B. 错

我的答案: 对正确答案: 对

5分

13. (判断题)数组之间的任何算术运算都会将运算应用到数组的每个元素上。（ ）

A. 对

B. 错

我的答案: 对正确答案: 对

5分

14. (判断题)

NumPy 中，可以混合使用切片与整数索引来获取二维数组的元素。（ ）

A. 对

B. 错

我的答案: 对正确答案: 对

5分

15. (判断题)

当通过布尔索引操作数组时，返回的数据是布尔数组中False对应位置的值。（ ）

A. 对

B. 错

我的答案: 错正确答案: 错

5分

16. (判断题)

Seaborn是基于Matplotlib的绘图库。（ ）

A. 对

B. 错

我的答案: 对正确答案: 对

5分

17. (判断题)直方图与柱形图没有任何区别。（ ）

A. 对

B. 错

我的答案: 错正确答案: 错

5分

18. (判断题)

饼图在识别异常值方面有一定的优越性。（ ）

A. 对

B. 错

我的答案: 错正确答案: 错

5分

19. (判断题)Matplotlib默认支持中文显示。（ ）

A. 对

B. 错

我的答案: 错正确答案: 错

5分

20. (判断题)

Matplotlib不能绘制2D图表。（ ）

A. 对

B. 错

我的答案: 错正确答案: 错

5分

1. (单选题)下列选项中，用来表示数组维度元组的属性是（ ）。

A. ndim

B. shape

C. size

D.

dtype

我的答案: B:shape;正确答案: B:shape;

10分

2. (单选题)下面选项中，创建的是3行3列数组的是（ ）。

A. arr = np.array([1, 2, 3])

B. arr = np.array([[1, 2, 3], [4, 5, 6]])

C. arr = np.array([[1, 2], [3, 4]])

D. np.ones((3, 3))

我的答案: D:np.ones((3, 3));正确答案: D:np.ones((3, 3));

10分

3. (单选题)

请阅读下面一段程序：

arr\_2d = np.array([[11, 20, 13],[14, 25, 16],[27, 18, 9]])

print(arr\_2d[1, :1])

执行上述程序后，最终输出的结果为（ ）。

A. [14]

B. [25]

C. [14, 25]

D. [20, 25]

我的答案: A:[14];正确答案: A:[14];

10分

4. (单选题)

请阅读下面一段程序：

arr = np.arange(6).reshape(1, 2, 3)

print(arr.transpose(2, 0, 1))

执行上述程序后，最终输出的结果为（ ）。

A.

Plain Text

[[[2 5]]

Plain Text

[[0 3]]

[[1 4]]]

B.

Plain Text

[[[1 4]]

Plain Text

[[0 3]]

[[2 5]]]

C.

Plain Text

[[[0 3]]

Plain Text

[[1 4]]

Plain Text

[[2 5]]]

D.

Plain Text

[[[0]

Plain Text

[3]]

Plain Text

[[1]

Plain Text

[4]]

Plain Text

[[2]

Plain Text

[5]]]

我的答案: C:[[[0 3]] [[1 4]] [[2 5]]] ;正确答案: C:[[[0 3]] [[1 4]] [[2 5]]] ;

10分

5. (单选题)下列选项中，用于表示矢量化三元表达式的是（ ）。

A. where()

B. cumsum()

C. sort()

D. unique()

我的答案: A:where();正确答案: A:where();

10分

6. (单选题)在Matplotlib中，用于展示图表的函数是（ ）。

A. plot()

B.

A. bar()

C. hist()

D. show()

我的答案: D:show();正确答案: D:show();

10分

7. (单选题)

下列选项中，关于Pyecharts库的描述正确的是（ ）。

A. Pyecharts库生成图表不能与用户交互

B. Pyecharts库无需安装便可以直接使用

C. Pyecharts库是基于Seaborn进行封装的

D. Pyecharts库可以绘制柱形图

我的答案: D:Pyecharts库可以绘制柱形图;正确答案: D:Pyecharts库可以绘制柱形图;

10分

8. (单选题)下列选项中，说法错误的是。（ ）

A. distplot()函数用于绘制直方图和核密度估计曲线

B. distplot()函数默认不使用密度观察条

C. stripplot()函数用于绘制直方图

D. swarmplot()函数用于绘制数据点不重叠的散点图

我的答案: C:stripplot()函数用于绘制直方图;正确答案: C:stripplot()函数用于绘制直方图;

10分

9. (单选题)在Matplotlib中，用于给图表添加标题的函数是（ ）。

A. xlabel()

B. title()

C. legend()

D. text()

我的答案: B:title();正确答案: B:title();

10分

10. (单选题)下列选项中，关于图表的描述错误的是（ ）。

A. 气泡图用于展示三个变量之间的关系

B. 核密度曲线图通过线段的倾斜程度显示数据的变化情况

C. 饼图显示一个数据序列中的各项的大小与各项总和的比例

D. 柱形图使用矩形条的高度反映不同分类数据之间的差异情况

我的答案: B:核密度曲线图通过线段的倾斜程度显示数据的变化情况;正确答案: B:核密度曲线图通过线段的倾斜程度显示数据的变化情况;  
  
以下老师说考的比较少  
1. (单选题)

Jupyter Notebook不具备的功能是（ ）。

A. Jupyter Notebook可以直接生成一份交互式文档

B. Jupyter Notebook可以安装Python库

C. Jupyter Notebook可以导出HTML文件

D. Jupyter Notebook可以将文件分享给他人

我的答案: B:Jupyter Notebook可以安装Python库;

10分

2. (单选题)

关于数据分析的说法，下列描述错误的是（）。

A. 数据分析可以通过计算机工具和数学知识处理数据

B. 可以做出具有针对性的决策

C. 数据分析没有实际的使用意义

D. 在大数据环境下，数据分析能够挖掘出更有价值的信息

我的答案: C:数据分析没有实际的使用意义;

10分

3. (单选题)

下列关于数据分析概念的描述错误的是（ ）。

A. 使用适当的统计分析方法对收集来的大量数据进行分析

B. 数据分析可以从数据中提炼出有用的信息并形成结论

C. 数据分析可以对数据进行更深层的研究

D. 数据分析不能够在杂乱无章的数据中提取有用的数据

我的答案: D:数据分析不能够在杂乱无章的数据中提取有用的数据;

10分

二. 多选题（共3题，30分）

4. (多选题)

下列关于Jupyter Notebook的描述错误的是（ ）。

A. Jupyter Notebook有两种模式

B. Jupyter Notebook有两种单元形式

C. Jupyter Notebook Markdown无法使用LaTeX语法

D. Jupyter Notebook仅仅支持Python语言

我的答案: ABCD:Jupyter Notebook有两种模式; Jupyter Notebook有两种单元形式; Jupyter Notebook Markdown无法使用LaTeX语法; Jupyter Notebook仅仅支持Python语言;

0分

5. (多选题)

下列选项中，属于数据处理的目的是（ ）。

A. 使数据更加适用于分析

B. 对无意义的数据进行清洗

C. 整合数据

D. 提高数据质量

我的答案: ABCD:使数据更加适用于分析; 对无意义的数据进行清洗; 整合数据; 提高数据质量;

10分

6. (多选题)

下列关于Python数据分析库的描述错误的是（ ）

A. NumPy的在线安装不需要其它任何辅助工具

B. SciPy的主要功能是可视化图表

C. pandas能够实现数据的整理工作

D. scikit-learn包含所有算法

我的答案: ABD:NumPy的在线安装不需要其它任何辅助工具; SciPy的主要功能是可视化图表; scikit-learn包含所有算法;

10分

三. 填空题（共2题，20分）

7. (填空题)

数据分析是指使用适当的【】对收集来的大量数据进行分析，从中提取有用信息和形成结论，并加以详细研究和概括总结的过程。

我的答案：

10分

(1) 统计分析方法

8. (填空题)

在统计学领域，数据分析可以划分为【】、【】和【】。

我的答案：

0分

(1) 描述性分析

(2) 推断性分析

(3) 因果性分析

四. 判断题（共2题，20分）

9. (判断题)

Python是一门胶水语言，可以轻松地操作其它语言编写的库。

A. 对

B. 错

我的答案: 对

10分

10. (判断题)

数据分析是一个有目的地收集和整合数据的过程

A. 对

B. 错

我的答案: 错