



总分 162

判断题

得分： 52    总分： 52

1-1    密度函数可以是负的。 (1分)

☐ T        ☒ F

评测结果    答案正确 (1 分)

1-2    如果相关系数为0，则表明两个变量间不存在线性相关。 (1分)

☒ T        ☐ F

评测结果    答案正确 (1 分)

1-3    聚类分析的目的就是把相似的研究对象归类。 (1分)

☒ T        ☐ F

评测结果    答案正确 (1 分)

1-4    因子载荷经正交旋转后,各变量的共性方差和各因子的贡献都发生了变化。 (1分)

☐ T        ☒ F

评测结果    答案正确 (1 分)

1-5    判别分析中,若两个总体的协差阵相等,则 Fisher判别与距离判别等价。 (1分)

☒ T        ☐ F

评测结果    答案正确 (1 分)

1-6    因子分析只能用于研究变量之间的相关关系。 (1分)

☐ T        ☒ F

评测结果    答案正确 (1 分)

1-7    主成分分析是将原来较少的指标扩充为多个新的的综合指标的多元统计方法。 (1分)

☐ T        ☒ F

评测结果    答案正确 (1 分)

1-8    一般而言，不同聚类方法的结果不完全相同 (1分)

☒ T        ☐ F

评测结果    答案正确 (1 分)

1-9    在系统聚类过程中，聚合系数越大，合并的两类差异越小。 (1分)

☐ T        ☒ F

评测结果    答案正确 (1 分)

1-10    因子载荷量是指因子结构中原始变量在因子分析时抽取出的公共因子的相关程度。 (1分)

☒ T        ☐ F



评测结果 答案正确 (1 分)

1-11 最长距离法中，选择最小的距离作为新类与其他类之间的距离，然后将类间距离最小的两类进行合并，一直合并到只有一类为止。 (1分)

☐ T ☒ F

评测结果 答案正确 (1 分)

1-12 一个行列式中某一行（列）所有元素的公因子可以提到行列式符号的外边。 (1分)

☒ T ☐ F

评测结果 答案正确 (1 分)

1-13 主成分的数目大大少于原始变量的数目。 (1分)

☒ T ☐ F

评测结果 答案正确 (1 分)

1-14 numpy中产生全1的矩阵使用的方法是empty。 (1分)

☐ T ☒ F

评测结果 答案正确 (1 分)

1-15 K均值聚类分析中，样品一旦划入某一类就不可改变。 (1分)

☐ T ☒ F

评测结果 答案正确 (1 分)

1-16 相关关系是指变量间不确定性的依存关系。 (1分)

☒ T ☐ F

评测结果 答案正确 (1 分)

1-17 pandas中head(n)的意思是获取最后的n行数据。 (1分)

☐ T ☒ F

评测结果 答案正确 (1 分)

1-18 名义尺度的指标用一些类来表示，这些类之间有等级关系，但没有数量关系。 (1分)

☐ T ☒ F

评测结果 答案正确 (1 分)

1-19 因子分析把变量分成公共因子和独立因子两部分因素。 (1分)

☐ T ☒ F

评测结果 答案正确 (1 分)

1-20 Numpy的ndarray(数组)中，使用shape()来表示数组的维度尺寸。 (1分)

☒ T ☐ F

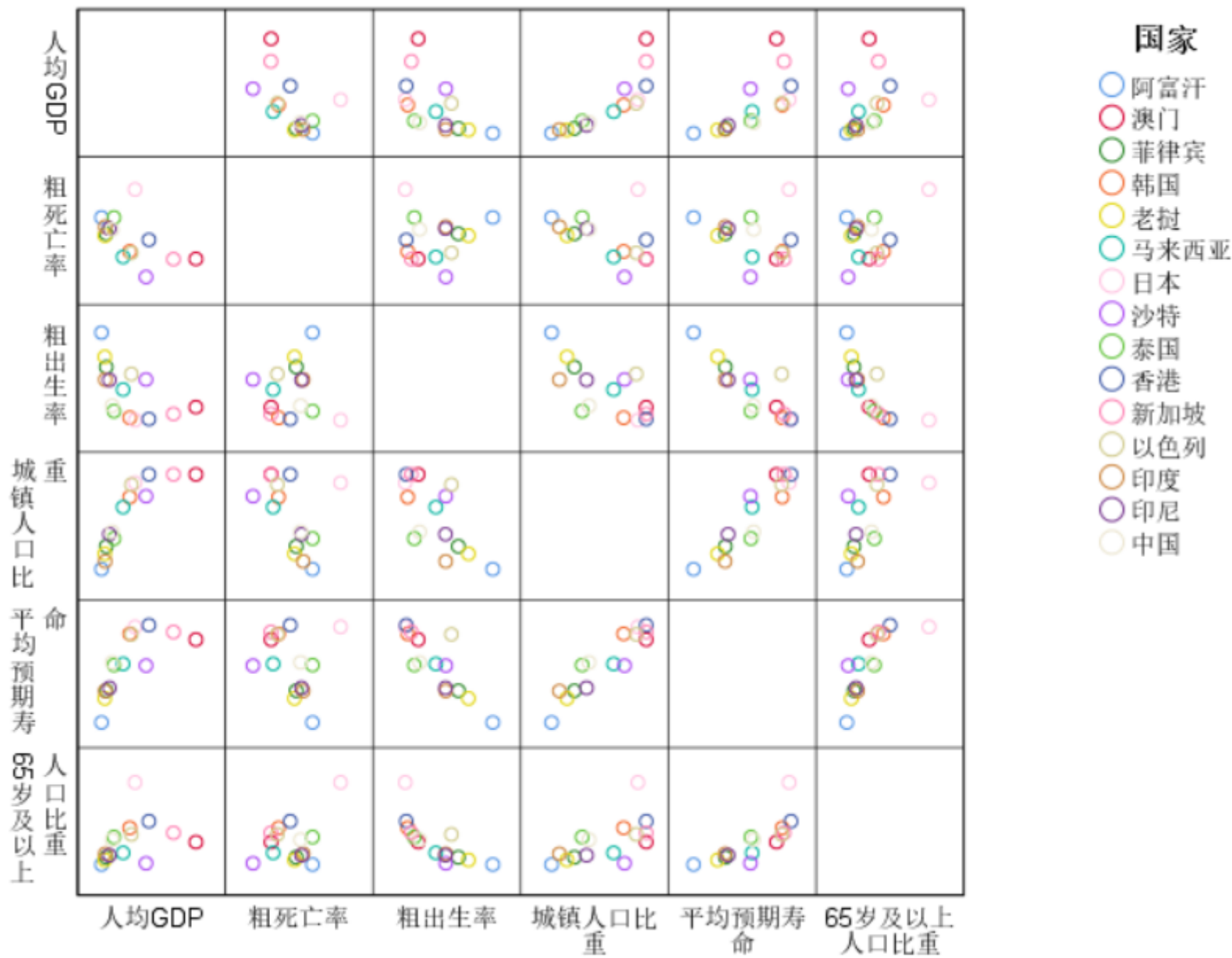
评测结果 答案正确 (1 分)

1-21 Python语言是非开源的语言。 (1分)

☐ T ☒ F



1-22 下图为应用SPSS软件制作的散点图矩阵，图中城镇人口比重与平均预期寿命存在强线性相关性。(1分)



☒ T ☐ F

评测结果 答案正确 (1 分)

1-23 脸谱图是用脸部特征表达变量间的相关性。(1分)

☐ T ☒ F

评测结果 答案正确 (1 分)

1-24 通过对多变量的脸谱图分析，可以直观地对原始数据资料进行归类或比较研究。(1分)

☒ T ☐ F

评测结果 答案正确 (1 分)

1-25 设要分析的资料共有p个变量，当p值较大时一张雷达图也可以清晰表达各观测之间的接近程度。(1分)

☐ T ☒ F

评测结果 答案正确 (1 分)

1-26 雷达图是目前应用较为广泛的多元资料进行作图的方法，利用雷达图可以很方便地研究个样本点之间的关系并对样品进行归类。(1分)

☒ T ☐ F

评测结果 答案正确 (1 分)

1-27 星图和星座图很相似，甚至有的文献把两者看成是一回事。(1分)

☐ T ☒ F



评测结果 答案正确 (1 分)

1-28 利用星座图可以方便地对样本点进行分类，再星座图上比较靠近的样本点比较相似，可分为一类。(1分)

☒ T ☐ F

评测结果 答案正确 (1 分)

1-29 多变量的图表示法使资料的呈现方式更直观、更形象，可以作为定量分析的研究结果并形成结论。(1分)

☐ T ☒ F

评测结果 答案正确 (1 分)

1-30 只要变量的指标数目不变，对应脸谱图的特征就不变。(1分)

☐ T ☒ F

评测结果 答案正确 (1 分)

1-31 马氏距离在协差阵为单位阵时退化为欧氏距离。(1分)

☒ T ☐ F

评测结果 答案正确 (1 分)

1-32 马氏距离受单位的影响。(1分)

☐ T ☒ F

评测结果 答案正确 (1 分)

1-33 相关系数度量了两个随机变量之间依赖关系的强弱。(1分)

☒ T ☐ F

评测结果 答案正确 (1 分)

1-34  $Cov(X,Y)=0$ , 称X与Y是不相关的。(1分)

☒ T ☐ F

评测结果 答案正确 (1 分)

1-35 随机向量X的协方差阵是对称矩阵。(1分)

☒ T ☐ F

评测结果 答案正确 (1 分)

1-36 若p维随机向量X的协方差阵存在,且每个分量的方差大于零，则X的相关阵的元素计算公式为: $r_{ij}=(cov(X_i,X_j))/(D(X_i)D(X_j)), i,j=1,2,...,p$ 。(1分)

☐ T ☒ F

评测结果 答案正确 (1 分)

1-37 设两个随机向量X和Y是相互独立的， $F(x,y)$ 为(X,Y)的联合分布函数， $G(x)$ 和 $H(y)$ 分别为X和Y的分布函数，则 $F(x,y)=G(x)H(y)$ 。(1分)

☒ T ☐ F

评测结果 答案正确 (1 分)

1-38 设两个随机向量X和Y是相互独立的， $f(x,y)$ 为(X,Y)的密度函数， $g(x)$ 和 $h(y)$ 分别为X和Y的密度函数，则 $f(x,y)=g(x)+h(y)$ 。(1分)

☐ T ☒ F



评测结果 答案正确 (1 分)

1-39 正态分布的条件分布仍为条件分布。

(1分)

☒ T ☐ F

评测结果 答案正确 (1 分)

1-40 相关系数不会取负值。

(1分)

☐ T ☒ F

评测结果 答案正确 (1 分)

1-41 相关系数的绝对值不会大于1。

(1分)

☒ T ☐ F

评测结果 答案正确 (1 分)

1-42 单纯依靠相关与回归分析，无法判断事物之间存在的因果关系

(1分)

☒ T ☐ F

评测结果 答案正确 (1 分)

1-43 相关系数为0表明两个变量之间不存在任何关系

(1分)

☐ T ☒ F

评测结果 答案正确 (1 分)

1-44 距离判别法考虑到了误判之后所造成的损失的差异。

(1分)

☐ T ☒ F

评测结果 答案正确 (1 分)

1-45 对于两组的判别，最大后验概率法的判别规则可使两个误判概率之和达到最小。

(1分)

☐ T ☒ F

评测结果 答案正确 (1 分)

1-46 对于两组皆为正态组及协差阵相同的情形下，两组先验概率相同及两个误判代价也相等时的贝叶斯判别等价于距离判别，也等价于费希尔判别。

(1分)

☒ T ☐ F

评测结果 答案正确 (1 分)

1-47 密度函数可以是负的。

(1分)

☐ T ☒ F

评测结果 答案正确 (1 分)

1-48 判别分析最基本要求是分组类型在两组以上且解释变量必须是可测量的。

(1分)

☒ T ☐ F

评测结果 答案正确 (1 分)

1-49 当总体G1和G2为正态总体且协方差相等时，选用马氏距离。

(1分)

☒ T ☐ F



评测结果 答案正确 (1 分)

1-50 标准化随机变量的协方差阵与原变量的相关系数相同。

(1分)

☒ T ☐ F

评测结果 答案正确 (1 分)

1-51 样本相关系数 $r \in (-1, 1)$

(1分)

☐ T ☒ F

评测结果 答案正确 (1 分)

1-52 在回归分析中，变量间的关系若是非确定关系，那么因变量不能由自变量唯一确定。

(1分)

☒ T ☐ F

评测结果 答案正确 (1 分)

## 单选题

得分：58 总分：58

2-1 关于 Python 语言的特点，以下选项中描述错误的是

(1分)

- ☒ A. Python 语言是非开源语言
- ☐ B. Python 语言是跨平台语言
- ☐ C. Python 语言是多模型语言
- ☐ D. Python 语言是脚本语言

评测结果 答案正确 (1 分)

2-2 使用pandas时需要导入下面哪个？

(1分)

- ☒ A. import pandas as pd
- ☐ B. import sys
- ☐ C. import numpy
- ☐ D. import matplotlib

评测结果 答案正确 (1 分)

2-3 Numpy中矩阵转换为向量用下面哪个函数？

(1分)

- ☐ A. areshape
- ☒ B. resize
- ☐ C. arange
- ☐ D. random

评测结果 答案正确 (1 分)

2-4 numpy中向量转成矩阵使用下面哪个函数？

(1分)

- ☒ A. reshape
- ☐ B. reval
- ☐ C. arange
- ☐ D. random

评测结果 答案正确 (1 分)

2-5 下列变量中，属于负相关的是

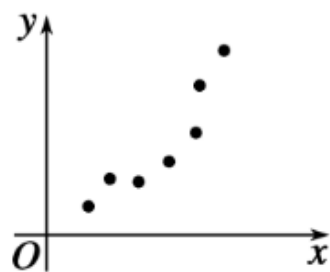
(1分)

- ☐ A. 收入增加，储蓄额增加
- ☐ B. 产量增加，生产费用增加
- ☐ C. 收入增加，支出增加
- ☒ D. 价格下降，消费增加

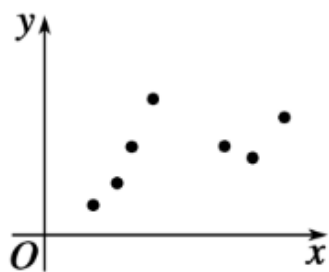
**评测结果** 答案正确 (1 分)

2-6 在以下四个散点图中，适用于做线性回归的是？

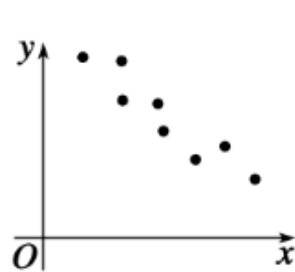
(1分)



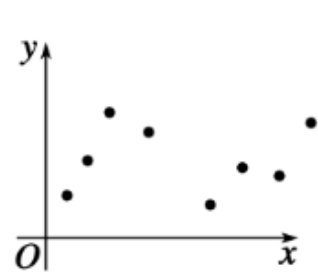
①



②



③



④

- ☐ A. ①②
- ☒ B. ①③
- ☐ C. ②③
- ☐ D. ③④

**评测结果**      答案正确 (1 分)

2-7 下面哪一项不是判别分析的方法?

(1分)

- ☐ A. 距离判别
- ☐ B. 贝叶斯判别
- ☒ C. 协方差阵判别
- ☐ D. 费歇判别

**评测结果** 答案正确 (1 分)

2-8 如果正态随机向量 $X$ 的各分量是相互独立的随机变量,那么随机变量 $X$ 的协方差阵是下面哪个?

(1分)

- ☐ A. 零矩阵
- ☒ B. 对角阵
- ☐ C. 对称阵
- ☐ D. 单位阵

**评测结果**      答案正确 (1 分)

2-9 利用主成分分析得到的各个主成分之间为下面哪个?

(1分)

- ☐ A. 互不独立
- ☐ B. 线性相关
- ☐ C. 不完全相关
- ☒ D. 互不相关

**评测结果** 答案正确 (1 分)

2-10 因子分析把变量分成下列两部分因素。

(1分)

- A. 相关因子和不相关因子



- ☐ B. 向量因子和非向量因子
- ☐ C. 公共因子和独立因子
- ☒ D. 公共因子和特殊因子

评测结果    答案正确 (1 分)

2-11    利用主成分分析得到的各个主成分之间为下面哪个选项？ (1分)

- ☐ A. 0
- ☐ B. 1
- ☐ C. 一定数值
- ☒ D. 最小

评测结果    答案正确 (1 分)

2-12    决策树中不包含以下哪种结点？ (1分)

- ☐ A. 根结点 (root node)
- ☐ B. 内部结点 (internal node)
- ☒ C. 外部结点 (external node)
- ☐ D. 叶结点 (leaf node)

评测结果    答案正确 (1 分)

2-13    多元正态分布是以下哪项的推广。 (1分)

- ☒ A. 一元正态分布
- ☐ B. 二元正态分布
- ☐ C. 三元正态分布
- ☐ D. 四元正态分布

评测结果    答案正确 (1 分)

2-14    因子分析中对于因子载荷的求解最不常用的方法是。 (1分)

- ☐ A. 主成分法
- ☐ B. 主轴因子法
- ☐ C. 极大似然法
- ☒ D. 极小似然法

评测结果    答案正确 (1 分)

2-15    当不知道数据所带标签时，可以使用哪种技术促使带同类标签的数据与带其他标签的数据相分离？ (1分)

- ☐ A. 分类
- ☒ B. 聚类
- ☐ C. 关联分析
- ☐ D. 隐马尔可夫链

评测结果    答案正确 (1 分)

2-16    Numpy的ndarray(数组)中，使用下面哪个来表示数组的维度尺寸。 (1分)

- ☐ A. size





- ☐ B. ndim
- ☒ C. shape
- ☐ D. itemsize

评测结果 答案正确 (1 分)

2-17 第k个主成分 $y_k$ 的系数向量是下面哪个? (1分)

- ☐ A. 第k个特征根
- ☒ B. 第k个特征根所对应的特征向量
- ☐ C. 第k个特征根所对应的方差贡献率
- ☐ D. 第k个特征根所对应的累计方差贡献率

评测结果 答案正确 (1 分)

2-18 假如你在训练一个线性回归模型，有下面两句话： (1分)

1. 如果数据量较少，容易发生过拟合。
2. 如果假设空间较小，容易发生过拟合。

关于这两句话，下列说法正确的是？

- ☐ A. 1 和 2 都错误
- ☒ B. 1 正确, 2 错误
- ☐ C. 1 错误, 2 正确
- ☐ D. 1 和 2 都正确

评测结果 答案正确 (1 分)

2-19 有数组  $n = \text{np.arange}(24). \text{reshape}(2,-1,2,2)$ ， $n.\text{shape}$ 的返回结果是什么? (1分)

- ☒ A. (2,3,2,2)
- ☐ B. (2,2,2,2)
- ☐ C. (2,4,2,2)
- ☐ D. (2,6,2,2)

评测结果 答案正确 (1 分)

2-20 常用统计函数中， $\text{mean}$ 计算下面的哪个? (1分)

- ☐ A. 数组的和
- ☐ B. 数组的最大值
- ☒ C. 数组的均值
- ☐ D. 数组的最小值

评测结果 答案正确 (1 分)

2-21 设，则 $x$ 离 $\mu$ 越远（近）密度越小（大）。度量这里远近的距离应是: (1分)

- ☐ A. 明氏距离
- ☐ B. 欧氏距离
- ☒ C. 马氏距离
- ☐ D. 曼哈顿距离

评测结果 答案正确 (1 分)

2-22 聚类分析方法中，下列哪种方法也称为快速聚类法。(1分)

☐ A. 系统聚类

☒ B. Kmeans聚类

☐ C. 分层聚类

☐ D. 密度聚类

评测结果 答案正确 (1 分)

2-23 下面哪一个选项不是判别分析的方法：(1分)

☐ A. 二次判别

☐ B. logistic回归

☐ C. Fisher判别

☒ D. 协方差阵判别

评测结果 答案正确 (1 分)

2-24 多变量的图表示法不包括(1分)

☐ A. 塑像图

☐ B. 轮廓图

☒ C. 箱线图

☐ D. 树形图

评测结果 答案正确 (1 分)

2-25 将所有样本点都点在一个半圆里面，根据样本点的位置可以直观地对样本点之间的相关性进行分析的多变量表示图是(1分)

☒ A. 星座图

☐ B. 雷达图

☐ C. 树形图

☐ D. 脸谱图

评测结果 答案正确 (1 分)

2-26 利用其可以很方便的研究个样本点之间的关系并对样本进行分类的是(1分)

☐ A. 散点图矩阵和星图

☐ B. 散点图矩阵和脸谱图

☒ C. 雷达图和脸谱图

☐ D. 雷达图和箱线图

评测结果 答案正确 (1 分)

2-27 在脸谱图中，脸谱的绘制顺序是脸庞高度、脸庞宽度、脸庞轮廓、嘴唇高度、嘴唇宽度、笑容曲线、眼睛高度、眼睛宽度、头发高度、头发宽度、头发造型、鼻子高度、鼻子宽度、耳朵宽度、耳朵高度共计15项，输入的数据变量依次为X1~X10共计10个变量，则X1表示的脸部特征是(1分)

☐ A. 脸庞高度

☒ B. 脸庞高度和头发造型

☐ C. 脸庞高度和耳朵高度

☐ D. 不能确定

评测结果 答案正确 (1 分)

- 2-28 将圆平分并由圆心连接各分点，将所得的p条线段作为坐标轴，根据各变量的取值对各坐标轴做适当刻度，对任意样本点可分别在p轴上确定其坐标，在各坐标轴上点出其坐标并依次连接p个点。这样做出的图是
- ☐ A. 星座图
  - ☒ B. 雷达图
  - ☐ C. 树形图
  - ☐ D. 脸谱图

评测结果    答案正确 (1 分)

---

- 2-29 设 $X \sim N(\mu, \Sigma)$ ，则X离 $\mu$ 越近（远）密度越大（小），度量这里远近的距离应是
- ☐ A. 其他距离
  - ☒ B. 马氏距离
  - ☐ C. 各变量标准化后的欧氏距离
  - ☐ D. 欧氏距离

评测结果    答案正确 (1 分)

---

- 2-30 设随机变量X和Y都服从正态分布, 且它们不相关, 则
- ☐ A.  $X+Y$ 服从一维正态分布
  - ☐ B.  $X$ 与 $Y$ 一定独立
  - ☐ C.  $X$ 和 $Y$ 都服从正态分布
  - ☒ D.  $X$ 与 $Y$ 未必独立

评测结果    答案正确 (1 分)

---

- 2-31 用样本的信息推断总体，样本应该是
- ☐ A. 总体中的典型部分
  - ☐ B. 总体中的有意义的一部分
  - ☒ C. 总体中随机抽取的一部分
  - ☐ D. 总体中随便抽取的一部分

评测结果    答案正确 (1 分)

---

- 2-32 抽样调查的目的是
- ☐ A. 研究总体信息
  - ☐ B. 研究样本信息
  - ☐ C. 研究典型案例
  - ☒ D. 样本推断总体特征

评测结果    答案正确 (1 分)

---

- 2-33 设A,B,C均为常数矩阵，随机向量X和Y的均值特征不包括
- ☒ A.  $E(AX) = AE(X)A'$
  - ☐ B.  $E(AXB) = AE(X)B$
  - ☐ C.  $E(AX+BY) = AE(X) + BE(Y)$
  - ☐ D.  $E(AXB+C) = AE(X)B + C$

评测结果    答案正确 (1 分)

---



2-34 设A,B,C均为常数矩阵，随机向量X和Y的协差阵D[X]和D[Y]性质包括

- ☒ A.  $D(AX)=AD(X)A'$
- ☐ B.  $cov(AX,BY)=Acov(X,Y)B$
- ☐ C.  $D(XY)=D(X)*D(Y)$
- ☐ D.  $D(X+BY)=D(X)+BD(Y)$

评测结果 答案正确 (1 分)

2-35 就大部分统计问题而言，欧氏距离存在缺点不包括

(1分)

- ☐ A. 与变量所用的单位有关
- ☐ B. 与变量所用的单位有关
- ☐ C. 没有考虑变量间的相关性
- ☒ D. 与变量的分布特征有关

评测结果 答案正确 (1 分)

2-36 T分布的均值是

(1分)

- ☒ A. 0
- ☐ B. 1
- ☐ C. 2
- ☐ D. 不确定

评测结果 答案正确 (1 分)

2-37 假设检验的一般步骤中不包括

(1分)

- ☐ A. 选定检验方法和计算检验统计量
- ☐ B. 确定P值和作出推断性结论
- ☐ C. 对总体参数的置信区间作出估计
- ☒ D. 建立假设和确定检验水准

评测结果 答案正确 (1 分)

2-38 假设检验是基于何种情况提出来的？

(1分)

- ☒ A.  $H_0$ 成立
- ☐ B.  $H_1$ 成立
- ☐ C.  $H_0$   $H_1$ 均可
- ☐ D.  $H_1$  成立还是 $H_1$ 成立看情况

评测结果 答案正确 (1 分)

2-39 比较两组均值向量，检验一般采用哪种方法

(1分)

- ☒ A. t检验
- ☐ B. 方差分析
- ☐ C. 卡方检验
- ☐ D. 秩和

评测结果 答案正确 (1 分)

2-40 进行多元均值检验时，假设 $H_0:\mu=\mu_0$ , $H_1: \mu\neq\mu_0$ ,协方差阵 $\Sigma$ 已知且 $\alpha$ 显著性水平下的分位点为c，拒绝域为：

(1分)



- ☐ A. 统计量  $< c$
- ☒ B. 统计量  $> c$
- ☐ C. 统计量  $\leq c$
- ☐ D. 统计量  $\geq c$

评测结果 答案正确 (1 分)

2-41 进行多元均值检验时，假设  $H_0: \mu = \mu_0, H_1: \mu \neq \mu_0$ , 协方差阵  $\Sigma$  未知且  $\alpha$  显著性水平下的分位点为  $c$ ，拒绝域为： (1分)

- ☐ A. 统计量  $\leq c$
- ☐ B. 统计量  $\geq c$
- ☐ C. 统计量  $< c$
- ☒ D. 统计量  $> c$

评测结果 答案正确 (1 分)

2-42 在进行两总体的均值比较时，假设  $H_0: \mu_1 = \mu_2, H_1: \mu_1 \neq \mu_2$ , 协方差阵  $\Sigma$  相等时，且  $\alpha$  显著性水平下的分位点为  $c$ ，拒绝域为： (1分)

- ☐ A. 统计量  $\leq c$
- ☐ B. 统计量  $\geq c$
- ☐ C. 统计量  $< c$
- ☒ D. 统计量  $> c$

评测结果 答案正确 (1 分)

2-43 在进行两总体的均值比较时，假设  $H_0: \mu_1 = \mu_2, H_1: \mu_1 \neq \mu_2$ , 协方差阵  $\Sigma$  不相等时，且  $\alpha$  显著性水平下的分位点为  $c$ ，拒绝域为： (1分)

- ☐ A. 统计量  $< c$
- ☒ B. 统计量  $> c$  C 统计量  $\leq c$
- ☐ C. 统计量  $\geq c$

评测结果 答案正确 (1 分)

2-44 在进行两总体的均值比较时，假设  $H_0: \mu_1 = \mu_2, H_1: \mu_1 \neq \mu_2$ , 协方差阵  $\Sigma$  不相等时，且  $\alpha$  显著性水平下的分位点为  $c$ ，拒绝域为： (1分)

- ☒ A. 统计量  $T^2 > T_{(p, n_1 + n_2 - 2)}^2(\alpha)$
- ☐ B. 统计量  $T^2 < T_{(p, n_1 + n_2 - 2)}^2(\alpha)$
- ☐ C. 统计量  $T^2 \geq T_{(p, n_1 + n_2 - 2)}^2(\alpha)$
- ☐ D. 统计量  $T^2 \leq T_{(p, n_1 + n_2 - 2)}^2(\alpha)$

评测结果 答案正确 (1 分)

2-45 在假设检验中，如果  $p$  值大于显著性水平  $\alpha$ ，就认为总体均值与检验值之间？ (1分)

- ☐ A. 无关
- ☐ B. 不知道
- ☐ C. 有显著差异
- ☒ D. 无显著差异

评测结果 答案正确 (1 分)

2-46 两个变量与  $x$  的回归模型中，通常用  $R^2$  来刻画回归的效果，则正确的叙述是 (1分)



- ☐ A.  $R^2$  越小，残差平方和越小
- ☐ B.  $R^2$  越大，残差平方和越大
- ☐ C.  $R^2$  与残差平方和无关
- ☒ D.  $R^2$  越小，残差平方和越大

评测结果    答案正确 (1 分)

2-47    下列说法中正确的是 (1分)

- ☐ A. 任何两个变量都具有相关关系
- ☒ B. 人的知识与其年龄具有相关关系
- ☐ C. 散点图中的各点是分散的没有规律
- ☐ D. 根据散点图求得的回归直线方程都是有意义的

评测结果    答案正确 (1 分)

2-48    一位母亲记录了儿子3~9岁的身高，由此建立的身高与年龄的回归直线方程为 $y' = 7.19x + 73.93$ ，据此可以预测这个孩子10岁时的身高，则正确的叙述是 (1分)

- ☐ A. 身高一定是145.83cm
- ☐ B. 身高超过146.00cm
- ☐ C. 身高低于145.00cm
- ☒ D. 升高在145.83cm左右

评测结果    答案正确 (1 分)

2-49    在画两个变量的散点图时，下面哪个叙述是正确的 (1分)

- ☐ A. 因变量在x轴上，自变量在y轴上
- ☒ B. 自变量在x轴上，因变量在y轴上
- ☐ C. 可以选择两个变量中任意一个变量在x轴上
- ☐ D. 可以选择两个变量中任意一个变量在y轴上

评测结果    答案正确 (1 分)

2-50    在对两个变量x, y进行线性回归分析时，有下列步骤：①对所求出的回归直线方程作出解释；②收集数据  $(x_i, y_i)$ ,  $i=1,2,\dots,n$ ；③求线性回归方程；④求未知参数；⑤根据所搜集的数据绘制散点图。如果根据可行性要求能够作出变量x, y具有线性相关结论，则在下列操作中正确的是 (1分)

- ☐ A. ①②⑤③④
- ☐ B. ③②④⑤①
- ☐ C. ②④③①⑤
- ☒ D. ②⑤④③①

评测结果    答案正确 (1 分)

2-51    线性回归中普通最小二乘估计的缺点是 (1分)

- ☐ A. 方差大的项在平方和中所起的作用小；
- ☐ B. 回归直线被拉向方差小的项；
- ☒ C. 是有偏估计；
- ☐ D. 回归直线被拉向方差大的项；



2-52 在回归分析中，代表了数据点和它在回归直线上相应位置的差异的是 (1分)

- ☐ A. 相关指数  $R^2$
- ☒ B. 残差平方和
- ☐ C. 回归平方和
- ☐ D. 总偏差平方和

评测结果 答案正确 (1 分)

2-53 三个总体  $G_1$ ,  $G_2$  和  $G_3$  的分布分别为:  $N(2, 0.5^2)$ ,  $N(0, 2^2)$ ,  $N(3, 1^2)$  现有  
一个样本  $x=2.5$ , 使用马氏距离进行距离判别,  $x$  应该归入总体 (1分)

- ☐ A.  $G_1$
- ☐ B. 无法确定
- ☐ C.  $G_2$
- ☒ D.  $G_3$

评测结果 答案正确 (1 分)

2-54 设有样本  $X(0,3)$ ,  $Y(1,2)$ , 求  $XY$  的欧式距离 (1分)

- ☒ A.  $\sqrt{2}$
- ☐ B.  $\sqrt{3}$
- ☐ C.  $\sqrt{5}$
- ☐ D. 2

评测结果 答案正确 (1 分)

2-55 使用“最大后验概率准则”得到的贝叶斯判别规则为 (1分)

- ☒ A.  $X \in G_l$  如果  $P(G_l|X) = \max_{1 \leq i \leq k} P(G_i|X)$
- ☐ B.  $X \in G_l$  如果  $h_l(x) = \max_j h_j(x) = \max \sum_{i=1}^k q_i C(j|i) f_i(x), j = 1, \dots, k$
- ☐ C.  $X \in G_l$  如果  $P(G_l|X) = \min_{1 \leq i \leq k} P(G_i|X)$
- ☐ D.  $X \in G_l$  如果  $h_l(x) = \min_j h_j(x) = \min \sum_{i=1}^k q_i C(j|i) f_i(x), j = 1, \dots, k$

评测结果 答案正确 (1 分)

2-56 以下哪种判别方法既可用于判别分类也可用于分离各组。 (1分)

- ☐ A. 贝叶斯判别——最大后验概率法
- ☐ B. 距离判别
- ☐ C. 贝叶斯判别——最小期望误判代价法
- ☒ D. 费希尔判别

评测结果 答案正确 (1 分)

2-57 两组情形下的最小期望误判代价法的判别规则包含三个比值, 其中最富有实际意义的是 (1分)





- ☐ A. 三个比值同样富有实际意义
- ☐ B. 概率密度之比
- ☒ C. 误判代价之比
- ☐ D. 先验概率之比

评测结果    答案正确 (1 分)

2-58 对两变量的散点图拟合最好的回归线，必须满足一个基本的条件是 (1分)

- ☒ A.  $\sum(y - \hat{y})^2$ ; 最小
- ☐ B.  $\sum(\hat{y} - \bar{y})^2$ ; 最小
- ☐ C.  $\sum(y - \hat{y})^2$ ; 最大
- ☐ D.  $\sum(\hat{y} - \bar{y})^2$ ; 最大

评测结果    答案正确 (1 分)

填空题

得分：48    总分：48

4-1 主成分分析是利用 降维 (1分) 的思想，在损失很少的信息前提下，把多个指标转化为几个综合指标的多元统计方法。

评测结果    答案正确 (1 分)

测试点得分	序号	结果	得分
	0	答案正确	1

4-2 Numpy中创建全为0的矩阵使用 zeros (1分) 。

评测结果    答案正确 (1 分)

测试点得分	序号	结果	得分
	0	答案正确	1

4-3 pandas中， read\_csv (1分) 用来读取csv文件。

评测结果    答案正确 (1 分)

测试点得分	序号	结果	得分
	0	答案正确	1

4-4 回归分析中，被预测或被解释的变量称为 因变量 (1分) 。

评测结果    答案正确 (1 分)

测试点得分	序号	结果	得分
	0	答案正确	1

4-5 主成分的协方差矩阵为 对角矩阵 (1分) 。

评测结果    答案正确 (1 分)

测试点得分	序号	结果	得分
	0	答案正确	1

4-6 因子分析中因子载荷系数 $a_{ij}$ 的统计意义是第 $i$ 个变量与第 $j$ 个公因子的 相关系数 (1分) 。

评测结果    答案正确 (1 分)

测试点得分	序号	结果	得分
	0	答案正确	1





4-7 判别分析是判别样品所属类型 (1分)的一种统计方法，常用的判别方法有 距离判别 (1分)、Fisher判别 (1分)、Bayes判别 (1分)、逐步判别法。

评测结果 答案正确 (4 分)

测试点得分	序号	结果	得分
	0	答案正确	1
	1	答案正确	1
	2	答案正确	1
	3	答案正确	1

4-8 多元分析研究的是多指标问题 (1分)的统计总体。

评测结果 答案正确 (1 分)

测试点得分	序号	结果	得分
	0	答案正确	1

4-9 多元分析中常用的统计量有 均值向量 (1分)、样本协差阵 (1分)、相关系数阵 (1分)。

评测结果 答案正确 (4 分)

测试点得分	序号	结果	得分
	0	答案正确	1
	1	答案正确	1
	2	答案正确	1
	3	答案正确	1

4-10 聚类分析就是分析如何对样品(或变量)进行量化分类的问题。通常聚类分析分为 Q型 (1分) 聚类 和 R型 (1分)聚类。

评测结果 答案正确 (2 分)

测试点得分	序号	结果	得分
	0	答案正确	1
	1	答案正确	1

4-11 协方差和相关系数仅仅是变量间 离散程度 (1分)的一种度量，并不能刻画变量间可能存在的 关联程度关系 (1分)。

评测结果 答案正确 (2 分)

测试点得分	序号	结果	得分
	0	答案正确	1
	1	答案正确	1

4-12 设 $X=(X_1,X_2,...,X_p)'$  有 $p$ 个分量， $\mu$ 为 $X$ 的均值向量，则 $\mu$ 是  $p$  (1分)维向量。

评测结果 答案正确 (1 分)

测试点得分	序号	结果	得分
	0	答案正确	1

4-13 几何平面上的点 $p=(x_1,x_2)$ 到原点 $O=(0,0)$ 的欧氏距离是  $(x_1^2+x_2^2)^{(1/2)}$  (1分)。

评测结果 答案正确 (1 分)

测试点得分	序号	结果	得分
	0	答案正确	1

4-14 设 $X$ 、 $Y$ 从均值向量为 $\mu$ ，协方差阵为 $\Sigma$ 的总体 $G$ 中抽取的两个样品，定义 $X$ 、 $Y$ 两点之间的马氏距离为  $(X-Y)' \Sigma^{-1} (X-Y)$  (1分)。



评测结果 答案正确 (1 分)

测试点得分	序号	结果	得分
	0	答案正确	1

4-15 如果正态随机向量X的协方差阵 $\Sigma$ 是对角阵，则X的各分量是相互独立 (1分) 的随机变量。

评测结果 答案正确 (1 分)

测试点得分	序号	结果	得分
	0	答案正确	1

4-16 在实际问题中,通常可以假定被研究的对象是多元正态分布 (1分) ,但分布中的参数 $\mu$ 和 $\Sigma$ 是未知的,一般的做法是通过样本来估计。

评测结果 答案正确 (1 分)

测试点得分	序号	结果	得分
	0	答案正确	1

4-17 数理统计中常用的抽样分布卡方分布，在多元统计中,与之对应的分布为Wishart分布 (1分)

评测结果 答案正确 (1 分)

测试点得分	序号	结果	得分
	0	答案正确	1

4-18 多元统计研究的是多指标问题,为了了解总体的特征,通过对总体抽样得到代表总体的样本但因为信息是分散在每个样本上的,就需要对样本进行加工,把样本的信息浓缩到不包含未知量的样本函数中,这个函数称为统计量 (1分)

评测结果 答案正确 (1 分)

测试点得分	序号	结果	得分
	0	答案正确	1

4-19  $\frac{1}{n-1} \sum_{i=1}^n [(x_i - \bar{x})]^2$  是样本方差 (1分) 的计算公式。

评测结果 答案正确 (1 分)

测试点得分	序号	结果	得分
	0	答案正确	1

4-20  $\frac{1}{n} \sum_{i=1}^n x_i$  是样本均值 (1分)

评测结果 答案正确 (1 分)

测试点得分	序号	结果	得分
	0	答案正确	1

4-21 假设  $(X,Y)$  是二元随机变量，则 $E[(X-\mu_X)(Y-\mu_Y)]=E(XY)-\mu_X * \mu_Y$ 是总体协方差 (1分)的计算公式

评测结果 答案正确 (1 分)

测试点得分	序号	结果	得分
	0	答案正确	1

4-22 多元数据的协方差阵检验中，需分析当前的波动幅度与过去的波动情形有无显著差异，此时要检验的假设 $H_0$ 为 $\Sigma=\Sigma_0$  (1分)

评测结果 答案正确 (1 分)

测试点得分	序号	结果	得分
	0	答案正确	1

4-23 多元数据协方差阵检验中，需要了解这多个总体之间的波动幅度有无明显的差异，此时要检验的假设的备择假设 $H_1$ 为 $\Sigma_i$ 不全相等 (1分)



评测结果 答案正确 (1 分)

测试点得分	序号	结果	得分
	0	答案正确	1

4-24 多总体的均值向量检验中，假设r个总体的方差相等，要检验的假设H0是 $\mu_1=\mu_2=\cdots=\mu_r$ ,备择假设H1是至少存在两个均值不等 (1分)

评测结果 答案正确 (1 分)

测试点得分	序号	结果	得分
	0	答案正确	1

4-25 散点图矩阵 (1分) 是借助两变量散点图的作图方法，它可以看作是一个大的图形方阵，其每一个非主对角元素的位置上是对应行的变量与对应列的变量的散点图。

评测结果 答案正确 (1 分)

测试点得分	序号	结果	得分
	0	答案正确	1

4-26 进行单指标检验时，假设H0:  $\mu=\mu_0$ ,H1:  $\mu\neq\mu_0$ ,计算得到统计量的数值为1.833，临界值t为0.45，此时应拒绝 (1分) 原假设

评测结果 答案正确 (1 分)

测试点得分	序号	结果	得分
	0	答案正确	1

4-27 针对连续变量的统计推断方法中，最常用的有T检验和方差分析 (1分)

评测结果 答案正确 (1 分)

测试点得分	序号	结果	得分
	0	答案正确	1

4-28 常用的Minkowski距离公式为 
$$d(x,y)=\left[\sum_{i=1}^p|x_i-y_i|^q\right]^{1/q}$$
，当q=2时，它表示欧氏距离 (1分)

评测结果 答案正确 (1 分)

测试点得分	序号	结果	得分
	0	答案正确	1

4-29 回归分析中从研究对象上可分为一元和多元 (1分)

评测结果 答案正确 (1 分)

测试点得分	序号	结果	得分
	0	答案正确	1

4-30 学习回归分析的目的是对实际问题进行预测 (1分) 和控制。

评测结果 答案正确 (1 分)

测试点得分	序号	结果	得分
	0	答案正确	1

4-31 判别分析适用于被解释变量是非度量 (1分) 变量的情形。

评测结果 答案正确 (1 分)

测试点得分	序号	结果	得分
	0	答案正确	1



4-32 与其他多元线性统计模型类似，判别分析的假设之一是每一个判别变量（解释变量）不能是其他判别变量的线性组合，假设之二是各组变量的协方差矩阵(1分)相等，假设之三是各判别变量遵从多元正态(1分)分布。

评测结果 答案正确 (3分)

测试点得分	序号	结果	得分
	0	答案正确	1
	1	答案正确	1
	2	答案正确	1

4-33 贝叶斯统计的思想是：假定对研究对象已有一定的认识，常用先验概率(1分)分布来描述这种认识，然后我们取得一个样本，用样本来修正已有的认识，得到后验概率(1分)分布，各种统计推断就都可以通过这个分布来进行。

评测结果 答案正确 (2分)

测试点得分	序号	结果	得分
	0	答案正确	1
	1	答案正确	1

4-34 按经典假设，线性回归模型中的解释变量应是非随机变量，且与随机误差项不相关(1分)

评测结果 答案正确 (1分)

测试点得分	序号	结果	得分
	0	答案正确	1

4-35 如果X和Y在统计上独立，则相关系数等于0(1分)。

评测结果 答案正确 (1分)

测试点得分	序号	结果	得分
	0	答案正确	1

4-36 回归分析中定义的解释变量和被解释变量都是随机(1分)变量

评测结果 答案正确 (1分)

测试点得分	序号	结果	得分
	0	答案正确	1

4-37 相关系数r的取值范围是 $-1 \leq r \leq 1$ (1分)。

评测结果 答案正确 (1分)

测试点得分	序号	结果	得分
	0	答案正确	1

编程题

得分：4 总分：4

7-1 统计单词的个数 (4分)

输入一行字符，以回车结束，统计其中单词的个数。单词是中间没有空格的字符序列，各单词之间用空格分隔，单词间空格数可以是多个。

输入格式:

在一行中输入字符

输出格式:

在一行中输出单词的数量

输入样例:

在这里给出一组输入。例如：

Let's go to room 209

输出样例:

在这里给出相应的输出。例如：

```
count = 5
```

编译器 PYTHON3

代码

```
a = input().split()
print('count =',len(a))
```

评测结果 答案正确 (4 分)

测试点得分

测试点	结果	得分	耗时	内存
0	答案正确	1	31.00 ms	3112 KB
1	答案正确	1	26.00 ms	3116 KB
2	答案正确	1	32.00 ms	3116 KB
3	答案正确	1	28.00 ms	3124 KB