

座位号  

国家开放大学2022年秋季学期期末统一考试

# 社会统计学 试题

2023 年 1 月

|     |   |   |   |   |     |
|-----|---|---|---|---|-----|
| 题 号 | 一 | 二 | 三 | 四 | 总 分 |
| 分 数 |   |   |   |   |     |

|     |     |
|-----|-----|
| 得 分 | 评卷人 |
|     |     |

一、单项选择题(每题只有一个正确答案,请将正确答案的字母填写在括号内。每题 2 分,共 20 分)

1. 某班级学生平均每天上网时间可以分为以下六组:1)1 小时及以下;2)1-2 小时;3)2-3 小时;4)3-4 小时;5)4-5 小时;6)5 小时及以上,则 5 小时及以上这一组的组中值近似为( )。

- A. 5 小时                      B. 6 小时  
C. 5.5 小时                  D. 6.5 小时

2. 以下关于条形图的表述,不正确的是( )。

- A. 条形图中矩形的宽度是固定的
- B. 条形图中矩形的长度(或高度)表示各类别频数的多少
- C. 条形图的矩形通常是紧密排列的
- D. 条形图通常是适用于所有类型数据

3. 某校期末考试,全校语文平均成绩为 80 分,标准差为 3 分,数学平均成绩为 87 分,标准差为 5 分。某学生语文得了 83 分,数学得了 97 分,从相对名次的角度看,该生( )的成绩考得更好。

- A. 数学                      B. 语文  
C. 两门课程一样          D. 无法判断

4. 在假设检验中,不拒绝虚无假设意味着( )。

- A. 虚无假设是肯定正确的  
B. 虚无假设肯定是错误的  
C. 没有证据证明虚无假设是正确的  
D. 没有证据证明虚无假设是错误的

5. 在回归方程中,若回归系数等于 0,这表明( )。

- A. 因变量  $y$  对自变量  $x$  的影响是不显著的  
B. 自变量  $x$  对因变量  $y$  的影响是不显著的  
C. 因变量  $y$  对自变量  $x$  的影响是显著的  
D. 自变量  $x$  对因变量  $y$  的影响是显著的

6. 对于线性回归,在因变量的总离差平方和中,如果残差平方和所占比例越大,那么两个变量之间( )。

- A. 相关程度越大                      B. 相关程度越小  
C. 完全相关                              D. 完全不相关

7. 在期中考试中,某班级学生统计学平均成绩为 80 分,标准差为 4 分。如果学生的成绩是正态分布,可以判断成绩在 72 分——88 分之间的学生人数大约占总体的( )。

- A. 95%                      B. 68%
- C. 99%                      D. 90%

8. 根据一个样本均值求出的 96% 的置信区间表明( )。

- A. 总体均值一定落入该区间内  
B. 总体均值有 96% 的概率不会落入该区间内  
C. 总体均值有 96% 的概率会落入该区间内  
D. 总体均值有 4% 的概率会落入该区间内

9. 残差平方和(SSR)反映了  $y$  的总变差中( )。

- A. 由于  $x$  与  $y$  之间的线性关系引起的  $y$  的变化部分  
B. 除了  $x$  对  $y$  的现有影响之外的其他因素对  $y$  变差的影响  
C. 由于  $x$  与  $y$  之间的非线性关系引起的  $y$  的变化部分  
D. 由于  $x$  与  $y$  之间的函数关系引起的  $y$  的变化部分

10. 下表是某单位工作人员年龄分布表,该组数据的众数出现在( )。

| 组别 | 按年龄分组(岁) | 工作人员数(人) |
|----|----------|----------|
| 1  | 20~24    | 6        |
| 2  | 25~29    | 14       |
| 3  | 30~34    | 24       |
| 4  | 35~39    | 18       |
| 5  | 40~44    | 12       |
| 6  | 45~49    | 18       |
| 7  | 50~54    | 14       |
| 8  | 55~59    | 6        |
|    | 合计       | 112      |

- A. 第 3 组                      B. 第 4 组  
C. 第 5 组                      D. 第 6 组

|     |     |
|-----|-----|
| 得 分 | 评卷人 |
|     |     |

二、名词解释(每小题 5 分,共 20 分)

11. 分层抽样

12. 散点图

13. Z 值

14. 卡方检验

|     |     |
|-----|-----|
| 得 分 | 评卷人 |
|     |     |

三、简答题(每题 10 分,共 30 分)

15. 判断以下随机变量是定性变量还是定量变量,如果是定量变量,确定是离散变量还是连续变量。

- (1)网络供应商的姓名
- (2)每月的网络服务费
- (3)每月上网时间
- (4)上网的主要目的
- (5)上月网上购物的次数

16. 简述直方图与条形图的区别。

17. 简要说明卡方的拟合优度检验和独立性检验的含义。

|     |     |
|-----|-----|
| 得 分 | 评卷人 |
|     |     |

四、计算题(每题 15 分,共 30 分)

18. 为估计每个网络用户每天上网的平均时间是多少,抽取了 225 个网络用户的简单随机样本,得到样本均值为 6.5 个小时,样本标准差为 2.5 个小时。

- (1)试用 95%的置信水平,计算网络用户每天平均上网时间的置信区间。
- (2)在所调查的 225 个网络用户中,年龄在 20 岁以下的用户为 90 个。以 95%的置信水平,计算年龄在 20 岁以下的网络用户比例的置信区间。
- 注: $Z_{0.025}=1.96$

19. 某行业管理局所属 40 个企业 2011 年产品销售额数据如下所示。

40 个企业 2011 年产品销售额

| 企业编号 | 销售额 | 企业编号 | 销售额 | 企业编号 | 销售额 | 企业编号 | 销售额 |
|------|-----|------|-----|------|-----|------|-----|
| 1    | 152 | 11   | 105 | 21   | 103 | 31   | 136 |
| 2    | 105 | 12   | 123 | 22   | 103 | 32   | 146 |
| 3    | 117 | 13   | 116 | 23   | 137 | 33   | 127 |
| 4    | 97  | 14   | 115 | 24   | 138 | 34   | 135 |
| 5    | 124 | 15   | 110 | 25   | 91  | 35   | 117 |
| 6    | 119 | 16   | 115 | 26   | 118 | 36   | 113 |
| 7    | 108 | 17   | 100 | 27   | 120 | 37   | 104 |
| 8    | 88  | 18   | 87  | 28   | 112 | 38   | 125 |
| 9    | 129 | 19   | 107 | 29   | 95  | 39   | 108 |
| 10   | 115 | 20   | 119 | 30   | 142 | 40   | 126 |

- 要求:
- (1)对 2011 年销售额按由低到高进行排序,求出众数、中位数和平均数。
- (2)如果按照规定,销售额在 125 万元以上的为先进企业,115 万-125 万之间的为良好企业,105 万-115 万之间的为一般企业,105 万以下的为落后企业,请按先进企业、良好企业、一般企业、落后企业进行分组,编制频数分布表,并计算累积频数和累积频率。

2022年秋季学期考试  
社会统计学 参考答案

2023 年 1 月

一、单项选择题(每题 2 分,共 20 分)

1. C
2. C
3. A
4. D
5. B
6. B
7. A
8. C
9. A
10. A

二、名词解释(每小题 5 分,共 20 分)

11. 分层抽样:先将总体中的所有单位按某种特征或标志(如年龄、性别、职业等)划分为若干类型或层次,然后再在各个类型或层次中采用简单随机抽样或系统抽样的方法抽取一个子样本,最后将这些子样本合起来构成总体样本。(5 分)
12. 散点图:在坐标系中,用 X 轴表示自变量 x,用 Y 轴表示因变量 y,而变量组(x,y)则用坐标系中的点表示,不同的变量组在坐标系中形成不同的散点,用坐标系及其坐标系中的散点形成的二维图就是散点图。(5 分)
13. Z 值:Z 值又称为标准分数,它是以平均数为参照点,以标准差为单位的描述原始数据在总体中相对位置的量数,通过计算 Z 值将一般正态分布转换为标准正态分布,Z 值的计算公式为: $Z=(X-\mu)/\sigma$ 。(5 分)
14. 卡方检验:卡方检验是对样本的频数分布所来自的总体分布是否服从某种理论分布或某种假设分布所作的假设检验,即根据样本的频数分布来推断总体的分布。(5 分)

三、简答题(每题 10 分,共 30 分)

15. 判断以下随机变量是定性变量还是定量变量,如果是定量变量,确定是离散变量还是连续变量。
- (1)网络供应商的姓名:定性(2 分)
- (2)每月的网络服务费:定量、连续变量(2 分)
- (3)每月上网时间:定量、连续变量(2 分)
- (4)上网的主要目的:定性(2 分)
- (5)上月网上购物的次数:定量、离散变量(2 分)
16. 简述直方图与条形图的区别。
- (1)适用范围不同,条形图适用于所有类型的数据;直方图只适用于数值型数据。(3 分)
- (2)条形图中条形的宽度是固定的,是用条形的长度(或高度)表示各类别频数的大小(2 分);直方图则用矩形的面积表示各类频数的大小,矩形的宽度和高度均有意义,宽度表示组距,高度表示每一组的频数或频率。(3 分)
- (3)条形图中各条形是分开排列的;直方图的矩形通常是紧密排列的。(2 分)
17. 简要说明卡方的拟合优度检验和独立性检验的含义。
- (1) $\chi^2$  检验用于分类变量之间关系的检验。当用于检验不同类别的目标量之间是否存在显著差异时,称为拟合优度检验。(4 分)例如,不同职业的人群中对某项改革措施的支持率是否一致。(1 分)
- (2) $\chi^2$  检验还可用于判断两个分类变量之间是否存在联系。如果两个分类变量之间没有关系,则称为独立,我们用  $\chi^2$  判断它们之间是否关联,这时称为独立性检验。(4 分)例如,对性行为的态度是否与受教育程度有关。(1 分)

四、计算题(每题 15 分,共 30 分)

18. (1)已知: $n=225, \bar{x}=6.5, s=2.5, Z_{0.025}=1.96$  (2 分)  
按照 95% 的置信水平,每天平均上网时间的置信区间为:  
$$\bar{x} \pm Z_{\alpha/2} \frac{s}{\sqrt{n}} = 6.5 \pm 1.96 \times \frac{2.5}{\sqrt{225}} = 6.5 \pm 0.33$$
 (4 分)  
即(6.53,6.17) (2 分)
- (2)样本比例: $p = \frac{90}{225} = 0.4$  (2 分)  
按照 95% 的置信水平,年龄在 20 岁以下的网络用户比例的置信区间为:  
$$p \pm Z_{\alpha/2} \sqrt{\frac{P(1-P)}{n}} = 0.4 \pm 1.96 \times \sqrt{\frac{0.4 \times (1-0.4)}{225}} = 0.4 \pm 0.064$$
 (3 分)  
即(33.6%,46.4%) (2 分)
19. (1)销售额由低到高排序: (5 分)  
87、88、91、95、97、100、103、103、104、105、105、107、108、108、110、112、113、115、115、115、116、117、117、118、119、119、120、123、124、125、126、127、129、135、136、137、138、142、146、152  
众数:115 (1 分)  
中位数:115.5 (1 分)  
平均数: $\bar{X} = \frac{X_1 + X_2 + \dots + X_N}{N}$   
$$= (152 + 146 + \dots + 88 + 87) \div 40$$
$$= 4647 \div 40$$
$$= 116.175$$
 (3 分)
- (2)40 个企业分组表 (5 分)

| 按销售额分组(万元) | 企业个数 | 累计频数 | 频率(%) | 累计频率 |
|------------|------|------|-------|------|
| 先进企业       | 11   | 11   | 27.5  | 27.5 |
| 良好企业       | 12   | 23   | 30    | 57.5 |
| 一般企业       | 8    | 31   | 20    | 77.5 |
| 落后企业       | 9    | 40   | 22.5  | 100  |
| 合计         | 40   |      | 100   |      |