

○-○-○

考点名称:

姓名:

学号:

○-○-○

试卷代号:11318

座位号

国家开放大学2024年秋季学期期末统一考试

社会统计学 试题

2025 年 1 月

注意事项:

1. 将你的学号、姓名及考点名称填写在试题和答题纸的规定栏内。考试结束后,把试题和答题纸放在桌上。试题和答题纸均不得带出考场。待监考人员收完试题和答题纸后方可离开考场。
2. 仔细阅读题目的说明,并按题目要求答题。所有答案必须写在答题纸的指定位置上,写在试题上的答案无效。
3. 用蓝、黑圆珠笔或钢笔(含签字笔)答题,使用铅笔答题无效。

一、单项选择题(每题只有一个正确答案。每题 2 分,共 20 分)

1. 某社会工作班级有 100 名学生,为了解该班级学生的消费水平,将所有学生按照学号排序后,在前十位学生中随机抽出学号为 003 名的学生,后面依次选出学号为 013、023、033、043、053、063、073、083、093 的九名同学进行调查。这种调查方法属于()。
A. 简单随机抽样 B. 整群抽样
C. 分层抽样 D. 系统抽样
2. 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
3. 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

4. 在期中考试中,某班级学生社会统计学平均成绩为 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 76 分—84 分之间的学生大约占总体的()。

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

5. 对于左偏分布,平均数、中位数和众数之间的关系是()。

- A. 平均数>中位数>众数
B. 中位数>平均数>众数
C. 众数>中位数>平均数
D. 众数>平均数>中位数

6. 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

7. 回归平方和(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

8. 对消费的回归分析中,学历、年龄、户口、性别、收入都是自变量,其中收入的回归系数为 1.6,这表明()。

- A. 收入每增加 1 元,消费增加 1.6 元
B. 消费每增加 1 元,收入增加 1.6 元
C. 收入与消费的相关系数为 1.6
D. 收入对消费影响的显著性为 1.6

9. 下列哪种情况不适合用方差分析()。

- A. 性别对收入的影响
B. 专业对收入的影响
C. 年龄对收入的影响
D. 行业对收入的影响

10. 下表是某单位工作人员年龄分布表,该单位工作人员的平均年龄是()岁。

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. 37
- B. 35
- C. 36
- D. 39

二、名词解释(每小题 5 分,共 20 分)

11. 非概率抽样
12. 折线图
13. Z 值
14. 参数与统计量

三、简答题(每题 10 分,共 30 分)

15. 等距分组和不等距分组有什么区别? 请举例说明。
16. 简述离散系数的作用与特点。
17. 简要举例说明在分析双变量的关系时,t 检验和卡方检验的主要区别。

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. 甲单位人均月收入 4500 元,标准差 1200 元。乙单位月收入分布如下所示:

乙单位月收入分布表

按收入分组(元)	人数(个)
3000 以下	120
3000~4000	420
4000~5000	540
5000~6000	420
6000 以上	300
合 计	1800

- 要求:
- (1)计算乙单位员工月收入的均值和标准差。
- (2)比较甲单位和乙单位哪个单位员工月收入的离散程度更大?

一、单项选择题(每题 2 分,共 20 分)

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

二、名词解释(每小题 5 分,共 20 分)

11. 非概率抽样:根据主观意愿、实际情况等进行抽样,而不依据随机原则进行抽样,这些不符合概率抽样要求的抽样都称为非概率抽样。(5 分)
12. 折线图:以直方图为基础,把直方图顶部的中点用直线依次连接起来,然后把原有的直方图抹掉后得出的图形就是折线图,也可称为频数多边形图。(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. 参数与统计量:参数是研究者想要了解的总体的某种特征值(2 分)。统计量是根据样本数据计算出来的一个量。(3 分)

三、简答题(每题 10 分,共 30 分)

15. 等距分组和不等距分组有什么区别? 请举例说明。
(1)在对数据进行分组时,如果各组组距相等,则称为等距分组。(2 分)例如,分析某班同学期末统计课成绩时,假如最低分为 73 分,最高分为 98 分,以 5 分为组距进行分组,分为 70—75 分,75—80 分,80—85 分,85—90 分,90—95 分,95—100 分。(3 分)
(2)如果各组组距不相等,则称为不等距分组。(2 分)例如,在分析人口时,往往将人口分为婴幼儿组(0—6 岁),少年儿童组(7—17 岁),中青年组(18—59 岁),老年人组(60 岁及以上),该分类中各组组距不相等,这就是不等距分组。(3 分)
16. 简述离散系数的作用与特点。
(1)离散系数是为了消除变量值高低的影响和计量单位不同对离散程度测度值的影响,主要用来分析数据离散程度。(3 分)离散系数是一组数据的标准差与该组数据均值之比,也称为变异系数。(2 分)
(2)离散系数是以标准差和均值为基础,因而离散系数包括所有数据信息,其代表性较好,但是只要其中一个数据变化,离散系数就跟随变化。(2 分)离散系数消除了变量值水平高低和计量单位不同对离散程度的影响,它的主要作用是比较不同总体或样本数据的离散程度。(3 分)

17. 简要举例说明在分析双变量的关系时,t 检验和卡方检验的主要区别。
分析双变量关系时,t 检验和卡方检验都是主要用于检验这两个变量之间是否存在显著关系。(2 分)t 检验主要用于对一个为数值型变量、另一个为分类变量且只有两个类别的变量的双变量关系的统计显著性检验。(2 分)卡方检验主要用于对两个分类变量之间的相关性进行统计检验,判断变量之间是否存在显著关系。(2 分)
例如,我们想考察收入与性别是否存在关系,或者两性的收入是否存在显著差异,可以用两独立样本 t 检验。(2 分)如果我们想考察职业与性别是否存在关系,而职业和性别都是分类变量,那么可以用卡方检验考察不同性别之间职业是否存在显著差异。(2 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:

$$\bar{x} \pm Z_{\alpha/2} \frac{s}{\sqrt{n}} = 6.5 \pm 1.96 \times \frac{2.5}{\sqrt{225}} = 6.5 \pm 0.33 \quad (4 \text{ 分})$$

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:

$$P \pm Z_{\alpha/2} \sqrt{\frac{P(1-P)}{n}} = 0.4 \pm 1.96 \times \sqrt{\frac{0.4 \times (1-0.4)}{225}} = 0.4 \pm 0.064 \quad (3 \text{ 分})$$

即(33.6%, 46.4%) (2 分)

19. (1)均值:

$$\begin{aligned} \bar{X} &= \frac{f_1 X_1 + f_2 X_2 + \dots + f_k X_k}{f_1 + f_2 + \dots + f_k} \\ &= (2500 \times 120 + 3500 \times 420 + 4500 \times 540 + 5500 \times 420 + 6500 \times 300) \div 1800 \\ &= 8460000 \div 1800 \\ &= 4700 \quad (4 \text{ 分}) \end{aligned}$$

方差:

$$\begin{aligned} \sigma^2 &= \frac{\sum_{i=1}^k (X_i - \bar{X})^2 f_i}{N} \\ &= \left[(2500 - 4700)^2 \times 120 + (3500 - 4700)^2 \times 420 + (4500 - 4700)^2 \times 540 \right. \\ &\quad \left. + (5500 - 4700)^2 \times 420 + (6500 - 4700)^2 \times 300 \right] \div 1800 \\ &= 2448000000 \div 1800 = 1360000 \end{aligned}$$

标准差: $\sigma = \sqrt{\sigma^2} = \sqrt{1360000} = 1166.19$ (4 分)

- (2)甲单位人均月收入的离散系数为:

$$v_{\text{甲}} = \frac{S_{\text{甲}}}{\bar{X}_{\text{甲}}} = 1200 \div 4500 = 0.2667 \quad (3 \text{ 分})$$

乙单位人均月收入的离散系数为:

$$v_{\text{乙}} = \frac{S_{\text{乙}}}{\bar{X}_{\text{乙}}} = 1166.19 \div 4700 = 0.2481 \quad (3 \text{ 分})$$

$v_{\text{甲}} > v_{\text{乙}}$,所以说甲单位人均月收入的离散程度大于乙单位人均月收入。 (1 分)