

国家开放大学2024年春季学期期末统一考试

社会统计学 试题

2024年7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. 为了调查某校社会工作专业200名学生的实务能力,需要抽取50人进行调查。调查员首先找到这200名学生的名单,并将所有学生从1到200编号,用200张小纸条分别写上001、002、003、...、200,然后将这些小纸条放入一个空盒子中,搅拌后,随意抽出50张小纸条,然后按照号码找到对应的学生进行调查。这种抽样方式是()。
 - A. 简单随机抽样
 - B. 系统抽样
 - C. 分层抽样
 - D. 多阶段抽样
2. 为了解某地区8万户家庭的消费情况,从该地区随机抽取5000户家庭进行调查,其中80%的家庭月消费在3000元以上,20%的家庭每月用于通讯、网络的费用在300元以上,则该研究的样本是()。
 - A. 8万户家庭
 - B. 5000户家庭
 - C. 80%家庭的月消费
 - D. 20%家庭的月消费
3. 某班级学生平均每学期的实习时长可以分为以下六组:1)50小时及以下;2)50—60小时;3)60—70小时;4)70—80小时;5)80—90小时;6)90小时及以上,该分组的组距是()。
 - A. 5
 - B. 10
 - C. 15
 - D. 20
4. 对于右偏分布,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

5. 有甲、乙两人同时打靶,各打10靶。甲平均每靶为8环,标准差为2;乙平均每靶9环,标准差为3,以下关于甲、乙两人打靶的稳定性水平表述正确的是()。

- A. 甲的离散程度小,稳定性水平低
- B. 甲的离散程度大,稳定性水平高
- C. 乙的离散程度小,稳定性水平低
- D. 乙的离散程度大,稳定性水平低

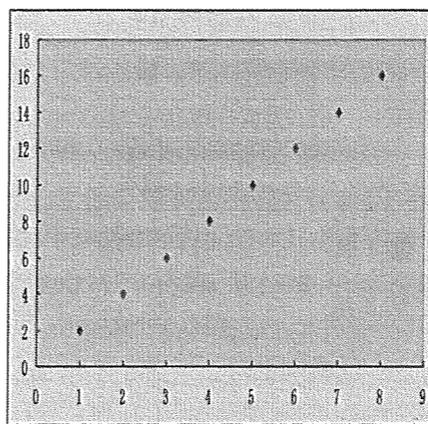
6. 在回归方程中,若回归系数等于0,这表明()。

- 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

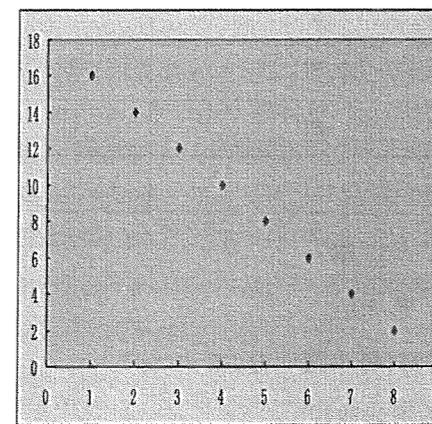
7. 对消费的回归分析中,学历、年龄、户口、性别、收入都是自变量,其中收入的回归系数为1.6,这表明()。

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

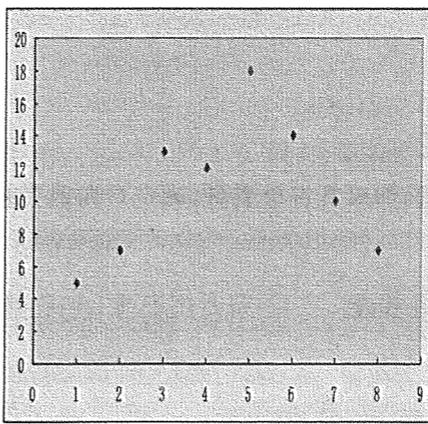
8. 以下四种不同特征的散点图,哪一个属于负线性相关()。



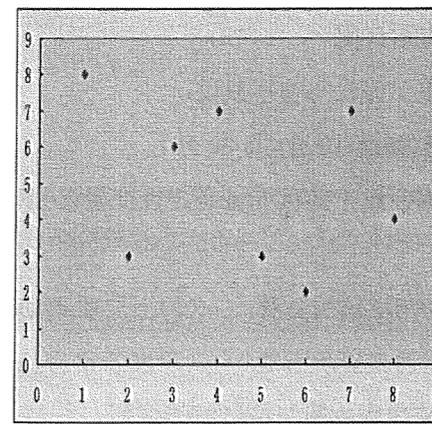
A



B



C



D

9. 残差平方和(SSE)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,该组数据的中位数出现在()。

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 5 组
- B. 第 4 组
- C. 第 3 组
- D. 第 2 组

二、名词解释(每小题 5 分,共 20 分)

- 11. 滚雪球抽样
- 12. 中心极限定理
- 13. 置信区间
- 14. 虚无假设与替换假设

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)使用电脑的品牌
- (2)上周收到的电子邮件数量
- (3)每月用于网上购物的金额
- (4)上网是否玩游戏
- (5)发表论文的数量

- 16. 简述直方图与条形图的区别。
- 17. 简述相关系数的取值与意义。

四、计算题(每题 15 分,共 30 分)

18. 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。

19. 为估计某地区每个家庭日均生活用水量是多少,抽取了 450 个家庭的简单随机样本,得到样本均值为 200 升,样本标准差为 50 升。

- (1)试用 95% 的置信水平,计算该地区家庭日均用水量的置信区间。
- (2)在所调查的 450 个家庭中,女性为户主的为 180 个。以 95% 的置信水平,计算女性为户主的家庭比例的置信区间。

注: $Z_{0.025} = 1.96$

国家开放大学2024年春季学期期末统一考试

社会统计学 试题答案及评分标准

(供参考)

2024年7月

一、单项选择题(每题2分,共20分)

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

二、名词解释(每小题5分,共20分)

11. 滚雪球抽样:先从总体中少数人员入手调查,并向他们询问其他符合条件的人,再去调查这些人并继续询问其他人的方法,像滚雪球一样。(5分)

12. 中心极限定理:不论总体分布是否服从正态分布,从均值为 μ 、方差为 σ^2 的总体中,抽取容量为 n 的随机样本,当 n 充分大时(通常要求 $n \geq 30$),样本均值的抽样分布近似服从均值为 μ 、方差为 σ^2/n 的正态分布。(5分)

13. 置信区间:在区间估计中,由样本统计量所构造的总体参数的估计区间,它由估计量加减抽样误差构成,我们将区间的最小值称为置信下限,区间的最大值称为置信上限。(5分)

14. 虚无假设与替换假设:我们将需要通过样本信息来推断其正确与否的命题称为虚无假设,也成为原假设或者零假设。(2分)如果虚无假设不成立,我们就拒绝虚无假设,需要在另一个假设中进行选择,这就是替换假设。(3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)使用电脑的品牌 定性变量(2分)
- (2)上周收到的电子邮件数量 离散变量(2分)
- (3)每月用于网上购物的金额 连续变量(2分)
- (4)上网是否玩游戏 定性变量(2分)
- (5)发表论文的数量 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. 简述相关系数的取值与意义。

相关系数的取值在 $-1 \sim 1$ 之间(2分);相关系数的正负号表示两个变量相关关系的方向,“+”表示正相关,“-”表示负相关(2分);相关系数的绝对值表示相关关系的程度,绝对值越大,相关程度越大,即 r 越接近1(2分);反之,绝对值越小,及 r 越接近0,相关程度越弱(2分);相关系数 $r=0$ 时,只能说变量之间不存在线性相关,而不能说它们之间不相关(2分)。

四、计算题(每题15分,共30分)

18. (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)

$$\begin{aligned} \text{平均数: } \bar{X} &= \frac{X_1 + X_2 + \dots + X_N}{N} = (152 + 146 + \dots + 88 + 87) \div 40 \\ &= 4647 \div 40 \\ &= 116.175 \quad (3 \text{分}) \end{aligned}$$

(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	

19. (1)已知: $n=225, \bar{x}=6.5, s=2.5, Z_{0.025}=1.96$ (2分)

用户每天平均用水量的95%的置信区间为:

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

所以,网络用户每天平均上网时间的置信区间为(6.17,6.83)。(2分)

(2)样本比例: $P = \frac{90}{225} = 0.4$ (2分)

户主为女性的家庭比例的95%的置信区间为:

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

所以,年龄在20岁以下的网络用户比例的置信区为(33.6%,46.4%)。(2分)