

试卷代号:1318

座位号

国家开放大学(中央广播电视大学)2016年秋季学期“开放本科”期末考试

社会统计学 试题(半开卷)

2017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. 以下关于因变量与自变量的表述不正确的是()。
 - A. 自变量的变化是以因变量的变化为前提
 - B. 因变量是由于其他变量的变化而导致自身发生变化的变量
 - C. 自变量是引起其他变量变化的变量
 - D. 因变量的变化不以自变量的变化为前提
2. 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
3. 以下关于条形图的表述,不正确的是()。
 - A. 条形图中条形的宽度是固定的
 - B. 条形图的矩形通常是紧密排列的
 - C. 条形图中条形的长度(或高度)表示各类别频数的多少
 - D. 条形图通常是适用于所有数据类型

8. 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

9. 中心极限定理认为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_____、方差为_____的正态分布。()

- A. μ, σ^2
- B. $\mu/n, \sigma^2/n$
- C. $\mu, \sigma^2/n$
- D. $\mu/n, \sigma^2$

10. 下列哪种情况不适合用方差分析?()

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

得 分	评卷人

二、名词解释(每题 5 分,共 20 分)

- 11. 整群抽样
- 12. Z 值
- 13. 二维表
- 14. 误差减少比例

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)上网的目的；
- (5)上月网购次数。
16. 简述什么是回归分析？其作用是什么？
17. 如何对配对样本进行 t 检验。

得 分	评卷人

四、计算题(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. 某农科院使用 4 种方法培育稻米,为确定哪种方法生产效率最高,随机划出 40 块试验田,并指定每块试验田使用其中的一种方法。通过对每块试验田的产量进行分析得到下面的方差分析表。请完成方差分析表。

变差来源	SS	df	MS	F	Sig.
组间	A	C	320	F	0.000
组内	6048	D	E	—	—
总计	B	39	—	—	—

试卷代号:1318

国家开放大学(中央广播电视大学)2016年秋季学期“开放本科”期末考试

社会统计学 试题答案及评分标准(半开卷)

(供参考)

2017年1月

一、单项选择题(每题2分,共20分)

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

二、名词解释(每题5分,共20分)

11. 整群抽样:先将总体按照某种标志或特征划分为一些子群体,然后从总体中随机抽取一些子群体,再将这些抽出的若干小群体内的所有元素构成总体样本的方法。(5分)

12. 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)

13. 二维表:二维表就是行列交叉的表格,(1分)将两个变量一个分行排放,一个分列排放,(1分)行列交叉处就是同属于两个变量的不同类的数据,也称为列联表。(3分)

14. 误差减少比例:在预测变量Y的值时,知道变量X的值时所减少的误差(E_1-E_2)与总误差 E_1 的比值称为误差减少比例,简称PRE。(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)上网的目的:定性(2分);
- (5)上月网购次数:定量、离散变量。(2分)

16. 简述什么是回归分析?其作用是什么?

(1)回归分析是通过一定的数学表达式将变量间的关系进行描述,确定一个变量或几个变量的变化对另一个特定变量的影响,是进行估计或预测的一种方法,侧重于考察变量之间的数量伴随关系。(4分)

(2)回归分析的作用包括:

①从已知数据出发,确定变量之间的数学关系式;(2分)

②对变量间的关系式进行统计检验,并从影响某一变量的多个变量中找出影响显著的变量;(2分)

③利用所求出的关系式,根据一个变量或多个变量的取值估计或预测另一个特定变量的取值。(2分)

17. 如何对配对样本进行 t 检验。(10分)

配对样本检验主要是判断不同的处理或试验结果是否有差异。配对样本的 t 检验用于检验两个相关的样本是否来自具有相同均值的总体。(2分)在对配对样本进行 t 检验时,首先计算两个样本中每个对应变量之间的差值;(3分)然后再检验其差值的均值是否为零,(1分)如果差值的均值接近零(在给定的置信区间内),(2分)说明两个总体均值在给定的置信水平上没有差异,如果差值的均值在置信区间外,则说明两个总体均值在给定的置信水平上有差异。(2分)

四、计算题(共 30 分)

18. (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:

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

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)

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

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)

19. 960、(3分)7008、(2分)3、(3分)36、(2分)168、(3分)1.90、(2分)

变差来源	SS	df	MS	F	Sig.
组间	960	3	320	1.90	0.000
组内	6048	36	168	—	—
总计	7008	39	—	—	—

试卷代号:1318

座位号

--	--

国家开放大学(中央广播电视大学)2017年春季学期“开放本科”期末考试

社会统计学 试题(半开卷)

2017年6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. 某班级有60名男生,40名女生,为了解学生购书支出,从男生中随机抽取12名学生,从女生中随机抽取8名学生进行调查。这种抽样方法属于()。

- A. 系统抽样
B. 分层抽样
C. 整群抽样
D. 简单随机抽样

2. 某地区家庭年均收入可以分为以下六组:1)1500元及以下;2)1500—2500元;3)2500—3500元;4)3500—4500元;5)4500—5500元;6)5500元及以上,则该分组的组距近似为()。

- A. 500元
B. 1500元
C. 1250元
D. 1000元

3. 为了解某地区的消费,从该地区随机抽取5000户进行调查,其中30%回答他们的月消费在5000元以上,40%回答他们每月用于通讯、网络的费用在300元以上。此处5000户是()。

- A. 样本
B. 总体
C. 变量
D. 统计量

4. 期中考试,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分之间的学生大约占总体的()。

- A. 95%
B. 68%
C. 89%
D. 90%

得 分	评卷人

四、计算题(共 30 分)

18. 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: $Z_{0.975} = 1.96$

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 分以下	120
3000~4000	420
4000~5000	540
5000~6000	420
6000 以上	300
合计	1800

(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?

试卷代号:1318

国家开放大学(中央广播电视大学)2017年春季学期“开放本科”期末考试

社会统计学 试题答案及评分标准(半开卷)

(供参考)

2017年6月

一、单项选择题(每题2分,共20分)

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

二、名词解释(每题5分,共20分)

11. 中心极限定理:中心极限定理具体内容为: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)

12. 整群抽样:先将总体按照某种标志或特征划分为一些子群体,然后从总体中随机抽取一些子群体,再将这些抽出的若干小群体内的所有元素构成总体样本的方法。(5分)

13. 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)

14. 二维表:二维表就是行列交叉的表格,(1分)将两个变量一个分行排放,一个分列排放,(1分)行列交叉处就是同属于两个变量的不同类的数据,也称为列联表。(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。(10分)

- (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. 简要说明卡方的拟合优度检验和独立性检验的含义。

(1) χ^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) χ^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 χ^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)

17. 如何对配对样本进行 t 检验。(10分)

配对样本检验主要是判断不同的处理或试验结果是否有差异。配对样本的 t 检验用于检验两个相关的样本是否来自具有相同均值的总体。(2分)在对配对样本进行 t 检验时,首先计算两个样本中每个对应变量之间的差值;(3分)然后再检验其差值的均值是否为零,(1分)如果差值的均值接近零(在给定的置信区间内),(2分)说明两个总体均值在给定的置信水平上没有差异,如果差值的均值在置信区间外,则说明两个总体均值在给定的置信水平上有差异。(2分)

四、计算题(共 30 分)

18. (1) 已知: $n = 450, \bar{x} = 200, s = 50, Z_{0.025} = 1.96$ (2分)

用户每天平均用水量的 95% 的置信区间为:

$$\bar{x} \pm Z_{\alpha/2} \frac{s}{\sqrt{n}} = 200 \pm 1.96 \times \frac{50}{\sqrt{450}} = 200 \pm 4.62 \text{ (4分)}$$

即(195.38, 204.62) (2分)

(2) 样本比例: $P = \frac{180}{450} = 0.4$ (2分)

户主为女性的家庭比例的 95% 的置信区间为:

$$P \pm Z_{\alpha/2} \sqrt{\frac{P(1-P)}{n}} = 0.4 \pm 1.96 \times \sqrt{\frac{0.4 \times (1-0.4)}{450}} = 0.4 \pm 0.045 \text{ (3分)}$$

即(35.5%, 44.5%) (2分)

19. (1) 均值:

$$\bar{X} = \frac{f_1 X_1 + f_2 X_2 + \cdots + f_k X_k}{f_1 + f_2 + \cdots + f_k}$$

$$= (2500 \times 120 + 3500 \times 420 + 4500 \times 540 + 5500 \times 420 + 6500 \times 300) \div 1800$$

$$= 8460000 \div 1800$$

$$= 4700 \quad (4分)$$

方差:

$$\begin{aligned} \sigma^2 &= \frac{\sum_{i=1}^k (X_i - \bar{X})^2 f_i}{N} \\ &= \frac{[(250 - 4700)^2 \times 120 + (3500 - 4700)^2 \times 420 + (4500 - 4700)^2 \times 540] + (5500 - 4700)^2 \times 420 + (6500 - 4700)^2 \times 300}{1800} \\ &= 2448000000 \div 1800 = 1360000 \end{aligned}$$

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:

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

乙单位人均月收入的离散系数为:

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

$V_{\text{甲}} > V_{\text{乙}}$, 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)

试卷代号:1318

座位号

--	--

国家开放大学(中央广播电视大学)2017年秋季学期“开放本科”期末考试

社会统计学 试题(半开卷)

2018 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. 某班级有 100 名学生,为了了解学生消费水平,将所有学生按照学习成绩排序后,在前十名学生中随机抽出成绩为第 3 名的学生,后面以 10 为间隔依次选出第 13、23、33、43、53、63、73、83、93 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. 以下关于因变量与自变量的表述不正确的是()。
A. 自变量是引起其他变量变化的变量
B. 因变量是由于其他变量的变化而导致自身发生变化的变量
C. 自变量的变化是以因变量的变化为前提
D. 因变量的变化不以自变量的变化为前提
3. 某地区 2001—2010 年人口总量(单位:万人)分别为 98,102,103,106,108,109,110,111,114,115,下列哪种图形最适合描述这些数据?()
A. 茎叶图
B. 环形图
C. 饼图
D. 线图
4. 以下关于条形图的表述,不正确的是()。
A. 条形图中条形的宽度是固定的
B. 条形图中条形的长度(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

5. 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

6. 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. 甲的离散程度小,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,稳定性水平高

7. 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第()组。

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

8. 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

9. 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
10. 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

得 分	评卷人

二、名词解释(每题 5 分,共 20 分)

- 11. 非概率抽样
- 12. 二维表
- 13. 置信水平
- 14. 卡方检验

得 分	评卷人

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. 简述相关系数的取值与意义。
- 17. 简述什么是回归分析? 其作用是什么?

得 分	评卷人

四、计算题(共 30 分)

18. 一项关于大学生体重状况的研究发现,男生的平均体重为 60 千克,标准差为 5 千克;女生的平均体重为 50 千克,标准差为 5 千克。请问:
- (1)是男生体重差异大还是女生体重差异大? 为什么?
 - (2)男生中有多少比重的人体重在 55 千克—65 千克之间?

(3)女生中有多少比重的人体重在 40 千克—60 千克之间?

备注: $\phi(1)=0.8413, \phi(2)=0.9772$

19. 为研究某种商品的价格(x)对其销售量(y)的影响,收集了 12 个地区的有关数据。通过分析得到以下方差分析表:

变差来源	SS	df	MS	F	Sig.
回归	1602708.6	B	D	F	0.000
残差	40158.08	C	E	—	—
总计	A	11	—	—	—

要求:

(1)计算上面方差分析表中 A、B、C、D、E、F 处的值。

(2)商品销售量的变差中有多少是由价格的差异引起的?

试卷代号:1318

国家开放大学(中央广播电视大学)2017年秋季学期“开放本科”期末考试

社会统计学 试题答案及评分标准(半开卷)

(供参考)

2018年1月

一、单项选择题(每题2分,共20分)

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

二、名词解释(每题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. 二维表:二维表就是行列交叉的表格,(1分)将两个变量一个分行排放,一个分列排放,(1分)行列交叉处就是同属于两个变量的不同类的数据,也称为列联表。(3分)

13. 置信水平:置信水平就是将构造置信区间的步骤重复很多次,置信区间包含总体参数真值的次数所占的比例。(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. 简述按照测量水平区分的四类变量,并举例说明。

(1)定类变量:当变量值的含义仅表示个体的不同类别,而不能说明个体的大小、程度等其它特征时,这种变量称为定类变量。(2分)例如:性别。(0.5分)

(2)定序变量:当变量值的含义不仅表示个体的不同类别,还可以区分个体之间大小、程度等序次差异时,这种变量称为定序变量。(2分)例如:学历。(0.5分)

(3)定距变量:当变量值不仅可以个体区分为不同类型并进行排序,而且可以确定不同类别之间的数量差别和间隔差距时,这样的变量称为定距变量。(2分)例如:智商。(0.5分)

(4)定比变量:除了上述三种变量的全部特征外,还可以计算两个变量值之间的比值时,这样的变量称为定比变量。(2分)例如:收入。(0.5分)

16. 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)

17. 简述什么是回归分析? 其作用是什么?

(1)回归分析是通过一定的数学表达式将变量间的关系进行描述,确定一个变量或几个变量的变化对另一个特定变量的影响,是进行估计或预测的一种方法,侧重于考察变量之间的数量伴随关系。(4分)

(2)回归分析的作用包括:

①从已知数据出发,确定变量之间的数学关系式;(2分)

②对变量间的关系式进行统计检验,并从影响某一变量的多个变量中找出影响显著的变量;(2分)

③利用所求出的关系式,根据一个变量或多个变量的取值估计或预测另一个特定变量的取值。(2分)

四、计算题(共30分)

18. (1)

$$V_{\text{男}} = \frac{S_{\text{男}}}{\bar{X}_{\text{男}}} = \frac{5}{60} = 0.083(3\text{分}) \quad V_{\text{女}} = \frac{S_{\text{女}}}{\bar{X}_{\text{女}}} = \frac{5}{50} = 0.1(3\text{分})$$

$V_{\text{男}} < V_{\text{女}}$, 所以女生体重差异比男生大。(1分)

$$\begin{aligned}(2) P(55 < X < 65) &= \Phi[(65-60)/5] - \Phi[(55-60)/5] \\ &= \Phi(1) - \Phi(-1) = \Phi(1) - [1 - \Phi(1)] \\ &= 2\Phi(1) - 1(3\text{分})\end{aligned}$$

由题得, $\Phi(1) = 0.8413$, 所以 $2\Phi(1) - 1 = 0.6826$, 即 68% 的男生体重在 55 千克—65 千克之间。(1分)

$$\begin{aligned}(3) P(55 < X < 65) &= \Phi[(60-50)/5] - \Phi[(40-50)/5] \\ &= \Phi(2) - \Phi(-2) = \Phi(2) - [1 - \Phi(2)] \\ &= 2\Phi(2) - 1(3\text{分})\end{aligned}$$

由题得, $\Phi(2) = 0.9772$, 所以 $2\Phi(2) - 1 = 0.9544$, 95% 的女生体重在 40 千克—60 千克之间。(1分)

19. (1)方差分析表:(12 分)

变差来源	SS	df	MS	F	Sig.
回归	1602708.6	1	1602708.6	399.1	0.000
残差	40158.08	10	4015.808	—	—
总计	1642866.68	11	—	—	—

$$(2)R^2 = \frac{SSR}{SST} = \frac{1602708.6}{1642866.68} = 0.9756$$

即商品销售量的变差中有 97.56%是由广告费用引起的。(3 分)

试卷代号:1318

座位号

国家开放大学(中央广播电视大学)2018年春季学期“开放本科”期末考试

社会统计学 试题(半开卷)

2018年7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. 为了解某地区的消费,从该地区随机抽取8000户进行调查,其中80%回答他们的月消费在3000元以上,20%回答他们每月用于通讯、网络的费用在300元以上。此处8000户是()。

- A. 样本
B. 总体
C. 变量
D. 统计量

2. 某地区家庭年均收入可以分为以下六组:1)1500元及以下;2)1500—2500元;3)2500—3500元;4)3500—4500元;5)4500—5500元;6)5500元及以上,则该分组的组距近似为()。

- A. 500元
B. 1500元
C. 1250元
D. 1000元

3. 先将总体按某种特征或标志分为不同的类别或层次,然后在各个类别中采用简单随机抽样或系统抽样的方式抽取子样本,最后将所有子样本合起来作为总样本,这样的抽样方式称为()。

- A. 简单随机抽样
B. 系统抽样
C. 整群抽样
D. 分层抽样

4. 正态分布中, σ 值越小,则()。

- A. 离散趋势越小
B. 离散趋势越大
C. 曲线越低平
D. 变量值越分散

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. 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. 甲的离散程度小,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,稳定性水平高

7. 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()。

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

8. 某单位认为该厂第一加工车间残品率估计高于 13%,而该车间主任认为该比例偏高,如果要检验该说法是否正确,则假设形式应该为()。

- A. $H_0: \pi \geq 0.13; H_1: \pi < 0.13$ B. $H_0: \pi \leq 0.13; H_1: \pi > 0.13$
C. $H_0: \pi = 0.13; H_1: \pi \neq 0.13$ D. $H_0: \pi > 0.13; H_1: \pi \leq 0.13$

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. 从两个总体中共选取了 8 个观察值,得到组间平方和为 432,组内平方和为 426,则组间均方和组内均方分别为()。

- A. 432,71 B. 216,71
C. 432,426 D. 216,426

得 分	评卷人

二、名词解释(每题 5 分,共 20 分)

11. 散点图
12. 中心极限定理
13. 置信水平
14. 离散系数

得 分	评卷人

三、简答题(每题 10 分,共 30 分)

15. 简述按照测量水平区分的四类变量,并举例说明。
16. 简述相关系数的取值与意义。
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。

得 分	评卷人

四、计算题(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表所示:

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. 某汽车生产商欲了解广告费用(万元)对销售量(辆)的影响。收集了过去 12 年的有关数据,通过分析得到:方程的截距为 363,回归系数为 1.42,回归平方和 $SSR=1600$,残差平方和 $SSE=450$ 。

要求:

(1)写出销售量 y 与广告费用 x 之间的线性回归方程。

(2)假如明年计划投入广告费用为 50 万,根据回归方程估计明年汽车销售量。

(3)计算判定系数 R^2 ,并解释它的意义。

试卷代号:1318

国家开放大学(中央广播电视大学)2018年春季学期“开放本科”期末考试

社会统计学 试题答案及评分标准(半开卷)

(供参考)

2018年7月

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

二、名词解释(每题5分,共20分)

11. 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)

12. 中心极限定理:中心极限定理具体内容为: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. 置信水平:置信水平就是将构造置信区间的步骤重复很多次,置信区间包含总体参数真值的次数所占的比例。(5分)

14. 离散系数:离散系数是一组数据的标准差与该数据均值之比,也称为变异系数。(5分)

三、简答题(每题10分,共30分)

15. 简述按照测量水平区分的四类变量,并举例说明。

(1)定类变量:当变量值的含义仅表示个体的不同类别,而不能说明个体的大小、程度等其它特征时,这种变量称为定类变量。(2分)例如:性别(0.5分)。

(2)定序变量:当变量值的含义不仅表示个体的不同类别,还可以区分个体之间大小、程度等序次差异时,这种变量称为定序变量。(2分)例如:学历(0.5分)。

(3)定距变量:当变量值不仅可以将个体区分为不同类型并进行排序,而且可以确定不同类别之间的数量差别和间隔差距时,这样的变量称为定距变量。(2分)例如:智商(0.5分)。

(4)定比变量:除了上述三种变量的全部特征外,还可以计算两个变量值之间的比值时,这样的变量称为定比变量。(2分)例如:收入(0.5分)。

16. 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)。

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(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)

$$\text{平均数:}=\bar{X}=\frac{X_1+X_2+\cdots+X_N}{N}$$

$$=(152+146+\cdots+88+87)\div 40$$

$$=4647\div 40$$

$$=116.175(3\text{分})$$

(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	

19. (1)回归方程为: $y=363+1.42x$ (4 分)

(2)当 $x=50$ 时, $y=363+1.42 \times 50=434$ (辆)(4 分)

(3)判定系数

$$R^2 = \frac{SSR}{SST} = \frac{SSR}{SSR+SSE}$$

$$=1600 \div (1600+450)$$

$$=0.7805(4 \text{ 分})$$

在汽车销售量的总变差中,有 78.05%可以由回归方程解释,说明回归方程的拟合程度很高。(3 分)

试卷代号:1318

座位号

--	--

国家开放大学(中央广播电视大学)2018年秋季学期“开放本科”期末考试

社会统计学 试题(半开卷)

2019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. 某班级有100名学生,为了解学生消费水平,将所有学生按照学习成绩排序后,在前十名学生中随机抽出成绩为第3名的学生,后面依次选出第13、23、33、43、53、63、73、83、93九名同学进行调查。这种抽样方法属于()。

- A. 简单随机抽样 B. 整群抽样
C. 分层抽样 D. 系统抽样

2. 以下关于因变量与自变量的表述不正确的是()。

- A. 自变量是引起其他变量变化的变量
B. 因变量是由于其他变量的变化而导致自身发生变化的变量
C. 自变量的变化是以因变量的变化为前提
D. 因变量的变化不以自变量的变化为前提

3. 某地区2001—2010年人口总量(单位:万人)分别为98,102,103,106,108,109,110,111,114,115,下列哪种图形最适合描述这些数据?()

- A. 茎叶图 B. 环形图
C. 饼图 D. 折线图

4. 以下关于条形图的表述,不正确的是()。

- A. 条形图中条形的宽度是固定的
B. 条形图中条形的长度(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

5. 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

6. 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. 甲的离散程度小,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,稳定性水平高

7. 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第()组。

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

8. 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

9. 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

10. 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

得 分	评卷人

二、名词解释(每题 5 分,共 20 分)

11. 非概率抽样
12. 二维表
13. 置信水平
14. 卡方检验

得 分	评卷人

三、简答题(每题 10 分,共 30 分)

15. 简述按照测量水平区分的四类变量,并举例说明。
16. 简述相关系数的取值与意义。
17. 简述什么是简单回归分析? 其作用是什么?

得 分	评卷人

四、计算题(每题 15 分,共 30 分)

18. 一项关于大学生体重状况的研究发现,男生的平均体重为 60 千克,标准差为 5 千克;女生的平均体重为 50 千克,标准差为 5 千克。请问:

- (1)是男生体重差异大还是女生体重差异大? 为什么?
- (2)男生中有多少比重的人体重在 55 千克—65 千克之间?
- (3)女生中有多少比重的人体重在 40 千克—60 千克之间?

备注: $\phi(1)=0.8413$, $\phi(2)=0.9772$

19. 为研究某种商品的价格(x)对其销售量(y)的影响,收集了 12 个地区的有关数据。通过分析得到以下方差分析表:

变差来源	SS	df	MS	F	Sig.
回归	1602708.6	B	D	F	0.000
残差	40158.08	C	E	—	—
总计	A	11	—	—	—

要求:

- (1) 计算上面方差分析表中 A、B、C、D、E、F 处的值。
- (2) 商品销售量的变差中有多少是由价格的差异引起的?

试卷代号:1318

国家开放大学(中央广播电视大学)2018年秋季学期“开放本科”期末考试

社会统计学 试题答案及评分标准(半开卷)

(供参考)

2019年1月

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

二、名词解释(每题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. 二维表:二维表就是行列交叉的表格,(1分)将两个变量一个分行排放,一个分列排放,(1分)行列交叉处就是同属于两个变量的不同类的数据,也称为列联表。(3分)

13. 置信水平:置信水平就是将构造置信区间的步骤重复很多次,置信区间包含总体参数真值的次数所占的比例。(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. 简述按照测量水平区分的四类变量,并举例说明。

(1)定类变量:当变量值的含义仅表示个体的不同类别,而不能说明个体的大小、程度等其它特征时,这种变量称为定类变量。(2分)例如:性别(0.5分)。

(2)定序变量:当变量值的含义不仅表示个体的不同类别,还可以区分个体之间大小、程度等序次差异时,这种变量称为定序变量。(2分)例如:学历(0.5分)。

(3)定距变量:当变量值不仅可以个体区分为不同类型并进行排序,而且可以确定不同

类别之间的数量差别和间隔差距时,这样的变量称为定距变量。(2分)例如:智商(0.5分)。

(4)定比变量:除了上述三种变量的全部特征外,还可以计算两个变量值之间的比值时,这样的变量称为定比变量。(2分)例如:收入(0.5分)。

16. 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)。

17. 简述什么是简单回归分析? 其作用是什么?

(1)回归分析是通过一定的数学表达式将变量间的关系进行描述,确定一个变量或几个变量的变化对另一个特定变量的影响,是进行估计或预测的一种方法,侧重于考察变量之间的数量伴随关系。(4分)

(2)回归分析的作用包括:

①从已知数据出发,确定变量之间的数学关系式;(2分)

②对变量间的关系式进行统计检验,并从影响某一变量的多个变量中找出影响显著的变量;(2分)

③利用所求出的关系式,根据一个变量或多个变量的取值估计或预测另一个特定变量的取值。(2分)

四、计算题(每题 15 分,共 30 分)

$$18. (1) v_{男} = \frac{S_{男}}{x_{男}} = \frac{5}{60} = 0.083 (3 \text{ 分}) \quad v_{女} = \frac{S_{女}}{x_{女}} = \frac{5}{50} = 0.1 (3 \text{ 分})$$

$v_{男} < v_{女}$, 所以女生体重差异比男生大。(1分)

$$\begin{aligned} (2) P(55 < X < 65) &= \Phi[(65-60)/5] - \Phi[(55-60)/5] \\ &= \Phi(1) - \Phi(-1) = \Phi(1) - [1 - \Phi(1)] \\ &= 2\Phi(1) - 1 (3 \text{ 分}) \end{aligned}$$

由题得, $\Phi(1) = 0.8413$, 所以 $2\Phi(1) - 1 = 0.6826$, 即 68% 的男生体重在 55 千克—65 千克之间。(1分)

$$\begin{aligned}
 (3) P(55 < X < 65) &= \Phi[(60-50)/5] - \Phi[(40-50)/5] \\
 &= \Phi(2) - \Phi(-2) = \Phi(2) - [1 - \Phi(-2)] \\
 &= 2\Phi(2) - 1 \quad (3 \text{ 分})
 \end{aligned}$$

由题得, $\Phi(2) = 0.9772$, 所以 $2\Phi(2) - 1 = 0.9544$, 95% 的女生体重在 40 千克—60 千克之间。(1 分)

19. (1) 方差分析表: (12 分)

变差来源	SS	df	MS	F	Sig.
回归	1602708.6	1	1602708.6	399.1	0.000
残差	40158.08	10	4015.808	—	—
总计	1642866.68	11	—	—	—

$$(2) R^2 = \frac{SSR}{SST} = \frac{1602708.6}{1642866.68} = 0.9756$$

即商品销售量的变差中有 97.56% 是由广告费用引起的。(3 分)

试卷代号:1318

座位号

--	--

国家开放大学2019年春季学期期末统一考试

社会统计学 试题(半开卷)

2019年7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. 以下关于因变量与自变量的表述不正确的是()。
 - A. 自变量的变化是以因变量的变化为前提
 - B. 因变量是由于其他变量的变化而导致自身发生变化的变量
 - C. 自变量是引起其他变量变化的变量
 - D. 因变量的变化不以自变量的变化为前提
2. 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
3. 以下关于条形图的表述,不正确的是()。
 - A. 条形图中条形的宽度是固定的
 - B. 条形图的矩形通常是紧密排列的
 - C. 条形图中条形的长度(或高度)表示各类别频数的多少
 - D. 条形图通常是适用于所有类型数

4. 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第()组。

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

5. 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

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. 根据一个样本均值求出的90%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90%的概率不会落入该区间内
C. 总体均值有90%的概率会落入该区间内
D. 总体均值有10%的概率会落入该区间内

8. 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

9. 中心极限定理认为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()、方差为()的正态分布。

A. μ, σ^2

B. $\mu/n, \sigma^2/n$

C. $\mu, \sigma^2/n$

D. $\mu/n, \sigma^2$

10. 下列哪种情况不适合使用方差分析? ()

A. 性别对收入的影响

B. 年龄对收入的影响

C. 专业对收入的影响

D. 行业对收入的影响

得 分	评卷人

二、名词解释(每题 5 分,共 20 分)

11. 整群抽样

12. Z 值

13. 二维表

14. 误差减少比例

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)上网的目的;(5)上月网购次数。

16. 简述什么是简单回归分析? 其作用是什么?

17. 如何对配对样本进行 t 检验。

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. 某农科院使用 4 种方法培育稻米,为确定哪种方法生产效率最高,随机划出 40 块试验田,并指定每块试验田使用其中的一种方法。通过对每块试验田的产量进行分析得到下面的方差分析表:

变差来源	SS	df	MS	F	Sig.
组间	A	C	320	F	0.000
组内	6048	D	E	—	—
总计	B	39	—	—	—

要求:

计算上面方差分析表中 A、B、C、D、E、F 处的值。

试卷代号:1318

国家开放大学2019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(半开卷)

(供参考)

2019年7月

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

二、名词解释(每题5分,共20分)

11. 整群抽样:先将总体按照某种标志或特征划分为一些子群体,然后从总体中随机抽取一些子群体,再将这些抽出的若干小群体内的所有元素构成总体样本的方法。(5分)

12. 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)

13. 二维表:二维表就是行列交叉的表格,(1分)将两个变量一个分行排放,一个分列排放,(1分)行列交叉处就是同属于两个变量的不同类的数据,也称为列联表。(3分)

14. 误差减少比例:在预测变量Y的值时,知道变量X的值时所减少的误差($E_1 - E_2$)与总误差 E_1 的比值称为误差减少比例,简称PRE。(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)上网的目的:定性(2分)

(5)上月网购次数:定量、离散变量(2分)

16. 简述什么是简单回归分析? 其作用是什么?

(1)回归分析是通过一定的数学表达式将变量间的关系进行描述,确定一个变量或几个变量的变化对另一个特定变量的影响,是进行估计或预测的一种方法,侧重于考察变量之间的数量伴随关系。(4分)

(2)回归分析的作用包括:

①从已知数据出发,确定变量之间的数学关系式;(2分)

②对变量间的关系式进行统计检验,并从影响某一变量的多个变量中找出影响显著的量;(2分)

③利用所求出的关系式,根据一个变量或多个变量的取值估计或预测另一个特定变量的取值。(2分)

17. 如何对配对样本进行t检验。

配对样本检验主要是判断不同的处理或试验结果是否有差异。配对样本的t检验用于检验两个相关的样本是否来自具有相同均值的总体。(2分)在对配对样本进行t检验时,首先计算两个样本中每个对应变量之间的差值;(3分)然后再检验其差值的均值是否为零,(1分)如果差值的均值接近零(在给定的置信区间内),(2分)说明两个总体均值在给定的置信水平上没有差异,如果差值的均值在置信区间外,则说明两个总体均值在给定的置信水平上有差异。(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)

用户每天平均用水量的95%的置信区间为:

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

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：

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

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)

19.960(3 分)、7008(2 分)、3(3 分)、36(2 分)、168(3 分)、1.90(2 分)

变差来源	SS	df	MS	F	Sig.
组间	960	3	320	1.90	0.000
组内	6048	36	168	—	—
总计	7008	39	—	—	—

试卷代号:1318

座位号

国家开放大学2019年秋季学期期末统一考试

社会统计学 试题(半开卷)

2020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. 先将总体按某标志分为不同的类别或层次,然后在各个类别中采用简单随机抽样或系统抽样的方式抽取子样本,最后将所有子样本合起来作为总样本,这样的抽样方式称为()。

- A. 简单随机抽样 B. 系统抽样
C. 整群抽样 D. 分层抽样

2. 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

3. 中心极限定理认为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()、方差为()的正态分布。

- A. μ σ^2 B. μ/n σ^2/n
C. μ σ^2/n D. μ/n σ^2

4. 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. 甲的离散程度小,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,稳定性水平高

5. 在标准正态分布中, σ 值越小, 则()。

- A. 离散趋势越大
B. 离散趋势越小
C. 曲线越低平
D. 变量值越分散

6. 对于线性回归, 在因变量的总离差平方和中, 如果回归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. 从两个总体中共选取了 8 个观察值, 得到组间平方和为 432, 组内平方和为 426, 则组间均方和组内均方分别为()。

- A. 432, 426
B. 216, 71
C. 432, 71
D. 216, 426

8. 某单位对该厂第一加工车间残品率估计高于 13%, 而该车间主任认为该比例偏高。如果要检验该说法是否正确, 则假设形式应该为()。

- A. $H_0: \pi \geq 0.13; H_1: \pi < 0.13$
B. $H_0: \pi \leq 0.13; H_1: \pi > 0.13$
C. $H_0: \pi = 0.13; H_1: \pi \neq 0.13$
D. $H_0: \pi > 0.13; H_1: \pi \leq 0.13$

9. 某次社会统计学考试中学生平均成绩 82, 标准差为 6, 某同学考试成绩为 80, 将这个分数转化为 Z 值为()。

- A. -0.33
B. +0.33
C. -3
D. +3

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第()组。

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. 折线图
12. 二维表
13. 统计推断
14. 自变量与因变量

得 分	评卷人

三、简答题(每题 10 分,共 30 分)

15. 简述按照测量水平区分的四类变量,并举例说明。
16. 简述什么是简单一元线性回归分析? 其作用是什么?
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。

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. 为研究某种商品的价格(x)对其销售量(y)的影响,收集了 12 个地区的有关数据。通过分析得到以下方差分析表:

变差来源	SS	df	MS	F	Sig.
回归	1602708.6	B	D	F	0.000
残差	40158.08	C	E	—	—
总计	A	11	—	—	—

要求:计算上面方差分析表中 A、B、C、D、E、F 处的值。

试卷代号:1318

国家开放大学2019年秋季学期期末统一考试

社会统计学 试题答案及评分标准(半开卷)

(供参考)

2020年1月

一、单项选择题(每题2分,共20分)

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

二、名词解释(每小题5分,共20分)

11. 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)

12. 二维表:二维表就是行列交叉的表格,(1分)将两个变量一个分行排放,一个分列排放,(1分)行列交叉处就是同属于两个变量的不同类的数据,也称为列联表。(3分)

13. 统计推断:根据统计量的分布和概率理论,由样本统计量来推断总体参数的过程,包括参数估计和假设检验两部分内容。(5分)

14. 自变量与因变量:引起其他变量变化的变量称为自变量(2分),由于其他变量的变化而导致自身发生变化的变量成为因变量。(3分)

三、简答题(每题10分,共30分)

15. 简述按照测量水平区分的四类变量,并举例说明。

(1)定类变量:当变量值的含义仅表示个体的不同类别,而不能说明个体的大小、程度等其它特征时,这种变量称为定类变量。(2分)例如:性别(0.5分)

(2)定序变量:当变量值的含义不仅表示个体的不同类别,还可以区分个体之间大小、程度等序次差异时,这种变量称为定序变量。(2分)例如:学历(0.5分)

(3)定距变量:当变量值不仅可以个体区分为不同类型并进行排序,而且可以确定不同类别之间的数量差别和间隔差距时,这样的变量称为定距变量。(2分)例如:智商(0.5分)

(4)定比变量:除了上述三种变量的全部特征外,还可以计算两个变量值之间的比值时,这样的变量称为定比变量。(2分)例如:收入(0.5分)

16. 简述什么是简单一元线性回归分析? 其作用是什么?

简单回归分析是通过一定的数学表达式将两个变量间的线性关系进行描述,确定自变量的变化对因变量的影响,是进行估计或预测的一种方法,侧重于考察变量之间的数量伴随关系。(或者简单回归分析是对具有线性相关关系的两个变量之间(其中一个为自变量,另一个为因变量)数量变化的一般关系进行分析,确定相应的数学关系式,以便进行估计或预测。)

(4分)

其作用包括:

(1)从已知数据出发,确定变量之间的数学关系式;(2分)

(2)对变量间的关系式进行统计检验,考察自变量是否对因变量有显著影响;(2分)

(3)利用所求出的关系式,根据自变量的取值估计或预测因变量的取值。(2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:

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

即(33.6%, 46.4%) (2分)

19. A: 1642866.68 (2分)

B: 1 (3分)

C: 10 (3分)

D: 1602708.6 (2分)

E: 4015.808 (2分)

F: 399.1 (3分)

座位号

--	--

社会统计学 试题

题 号	一	二	三	四	总 分
分 数					

得 分	评卷人

A. 甲的离散程度小,稳定性水平低 B. 甲的离散程度小,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,稳定性水平高

5. 在标准正态分布中, σ 值越小, 则()。

- A. 离散趋势越小
- B. 离散趋势越大
- C. 曲线越低平
- D. 变量值越分散

6. 某班级学生期末统计学考试平均成绩为 82 分, 标准差为 5 分。如果已知这个班学生的考试分数服从正态分布, 可以判断成绩在 72—92 之间的学生大约占全班学生的()。

- A. 68%
- B. 89%
- C. 90%
- D. 95%

7. 某单位对该厂第一加工车间残品率估计约为 13%, 而该车间主任认为该比例不符合实际情况, 如果要检验该说法是否正确, 则假设形式应该为()。

- A. $H_0: \pi \geq 0.13; H_1: \pi < 0.13$
- B. $H_0: \pi \leq 0.13; H_1: \pi > 0.13$
- C. $H_0: \pi = 0.13; H_1: \pi \neq 0.13$
- D. $H_0: \pi > 0.13; H_1: \pi \leq 0.13$

8. 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

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()。

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

得 分	评卷人

二、名词解释(每小题 5 分,共 20 分)

11. 概率抽样
12. 中心极限定理
13. 误差减少比例
14. 独立样本与配对样本

得 分	评卷人

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. 简述相关系数的取值与意义。
17. 简述什么是简单一元线性回归分析? 其作用是什么?

得 分	评卷人

四、计算题(每题 15 分,共 30 分)

18. 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。请问:

- (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: $Z_{0.025} = 1.96$

19. 一项关于大学生体重状况的研究发现,男生的平均体重为 60 千克,标准差为 5 千克;女生的平均体重为 50 千克,标准差为 5 千克。请问:

- (1)是男生体重差异大还是女生体重差异大? 为什么?
- (2)男生中有多少比重的人体重在 55 千克—65 千克之间?
- (3)女生中有多少比重的人体重在 40 千克—60 千克之间?

备注: $\Phi(1) = 0.8413, \Phi(2) = 0.9772$

试卷代号:1318

国家开放大学2020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)

2020年7月

一、单项选择题(每题2分,共20分)

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

二、名词解释(每小题5分,共20分)

11. 概率抽样:按照随机原则进行的抽样,总体中每个个体都是一定、非零的概率入选样本,并且入选样本的概率都是已知或可以计算的。(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. 误差减少比例:在预测变量 Y 的值时,知道变量 X 的值时所减少的误差($E_1 - E_2$)与总误差 E_1 的比值称为误差减少比例,简称PRE。(5分)

14. 独立样本与配对样本:独立样本是指我们得到的样本是相互独立的。(2分)配对样本就是一个样本中的数据与另一个样本中的数据相对应的两个样本。(1分)配对样本可以消除由于样本指定的不公平造成的差异。(2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71-75分,76-80分,81-85分,86-90分,91-95分,96-100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)。

17. 简述什么是简单一元线性回归分析? 其作用是什么?

简单回归分析是通过一定的数学表达式将两个变量间的线性关系进行描述, 确定自变量的变化对因变量的影响, 是进行估计或预测的一种方法, 侧重于考察变量之间的数量伴随关系。(或者简单回归分析是对具有线性相关关系的两个变量之间(其中一个为自变量, 另一个为因变量)数量变化的一般关系进行分析, 确定相应的数学关系式, 以便进行估计或预测。)

(4 分)

其作用包括:

(1) 从已知数据出发, 确定变量之间的数学关系式; (2 分)

(2) 对变量间的关系式进行统计检验, 考察自变量是否对因变量有显著影响; (2 分)

(3) 利用所求出的关系式, 根据自变量的取值估计或预测因变量的取值。 (2 分)

四、计算题(每题 15 分, 共 30 分)

18. (1) 已知: $n=450, \bar{x}=200, s=50, Z_{0.025}=1.96$ (2 分)

用户每天平均用水量的 95% 的置信区间为:

$$\bar{x} \pm Z_{\alpha/2} \frac{S}{\sqrt{n}} = 200 \pm 1.96 \times \frac{50}{\sqrt{450}} = 200 \pm 4.62 \quad (4 \text{ 分})$$

即(195.38, 204.62) (2 分)

(2) 样本比例: $p = \frac{180}{450} = 0.4$ (2 分)

户主为女性的家庭比例的 95% 的置信区间为:

$$p \pm Z_{\alpha/2} \sqrt{\frac{P(1-P)}{n}} = 0.4 \pm 1.96 \times \sqrt{\frac{0.4 \times (1-0.4)}{450}} = 0.4 \pm 0.045 \quad (3 \text{ 分})$$

即(35.5%, 44.5%) (2 分)

19. (1)

$$V_{\text{男}} = \frac{S_{\text{男}}}{x_{\text{男}}} = \frac{5}{60} = 0.083 \quad (3 \text{ 分}) \quad V_{\text{女}} = \frac{S_{\text{女}}}{x_{\text{女}}} = \frac{5}{50} = 0.1 \quad (3 \text{ 分})$$

$V_{\text{男}} < V_{\text{女}}$, 所以女生体重差异比男生大。 (1 分)

$$\begin{aligned} (2) P(55 < X < 65) &= \Phi[(65-60)/5] - \Phi[(55-60)/5] \\ &= \Phi(1) - \Phi(-1) = \Phi(1) - [1 - \Phi(-1)] \\ &= 2\Phi(1) - 1 \quad (3 \text{ 分}) \end{aligned}$$

由题得, $\Phi(1) = 0.8413$, 所以 $2\Phi(1) - 1 = 0.6826$, 即 68% 的男生体重在 55 千克—65 千克之间。 (1 分)

$$\begin{aligned} (3) P(55 < X < 65) &= \Phi[(60-50)/5] - \Phi[(40-50)/5] \\ &= \Phi(2) - \Phi(-2) = \Phi(2) - [1 - \Phi(-2)] \\ &= 2\Phi(2) - 1 \quad (3 \text{ 分}) \end{aligned}$$

由题得, $\Phi(2) = 0.9772$, 所以 $2\Phi(2) - 1 = 0.9544$, 95% 的女生体重在 40 千克—60 千克之间。 (1 分)

试卷代号:1318

座位号

--	--

国家开放大学2020年春季学期期末统一考试

社会统计学 试题

2020 年 9 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. 学校后勤集团想了解学校 42000 学生的每月生活费用,从中抽取 3200 名学生进行调查,以推断所有学生的每月生活费用水平。这项研究的样本是()。
- A. 42000 名学生
B. 3200 名学生
C. 42000 名学生的每月生活费用
D. 3200 名学生的每月生活费用
2. 关于因变量与自变量的表述,不正确的是()。
- A. 自变量是引起其他变量变化的变量
B. 因变量是由于其他变量的变化而导致自身发生变化的变量
C. 自变量的变化是以因变量的变化为前提
D. 因变量的变化是以自变量的变化为前提
3. 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
4. 某地区家庭年均收入可以分为以下六组:1)1500 元及以下;2)1500—2500 元;3)2500—3500 元;4)3500—4500 元;5)4500—5500 元;6)5500 元及以上,则该分组的组距近似为()。
- A. 500 元
B. 1500 元
C. 1250 元
D. 1000 元

5. 社会工作一班男生的平均体重为 55 千克,社会工作二班男生的平均体重为 52 千克,两个班级男生体重的标准差均为 5 千克,那么()。

- A. 一班学生体重的离散程度大 B. 二班学生体重的离散程度大
C. 两个班学生体重的离散程度相同 D. 无法判断

6. 某班级有 100 名学生,为了了解学生消费水平,将所有学生按照学习成绩排序后,在前十名学生中随机抽出成绩为第 3 名的学生,后面依次选出第 13、23、33、43、53、63、73、83、93 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

7. 在期中考试,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 68 分—92 分之间的学生大约占总体的()。

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

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. 从两个总体中共选取了 8 个观察值,得到组间平方和为 432,组内平方和为 426,则组间均方和组内均方分别为()。

- A. 432,71 B. 216,71
C. 432,426 D. 216,426

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第()组。

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. 整群抽样
12. Z 值
13. 置信水平
14. 虚无假设与替换假设

得 分	评卷人

三、简答题(每题 10 分,共 30 分)

15. 简述按照测量水平区分的四类变量,并举例说明。
16. 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。
17. 简述直方图与条形图的区别。

得 分	评卷人

四、计算题(每题 15 分,共 30 分)

18. 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 分以下	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?

19. 某农科院使用 4 种方法培育稻米,为确定哪种方法生产效率最高,随机划出 40 块试验田,并指定每块试验田使用其中的一种方法。通过对每块试验田的产量进行分析得到下面的方差分析表,请完成方差分析表。

变差来源	SS	df	MS	F	Sig.
组间	A	C	320	F	0.000
组内	6048	D	E	—	—
总计	B	39	—	—	—

试卷代号:1318

国家开放大学2020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)

2020年9月

一、单项选择题(每题2分,共20分)

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

二、名词解释(每小题5分,共20分)

11. 整群抽样:先将总体按照某种标志或特征划分为一些子群体,然后从总体中随机抽取一些子群体,再将这些抽出的若干小群体内的所有元素构成总体样本的方法。(5分)

12. 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)

13. 置信水平:置信水平就是将构造置信区间的步骤重复很多次,置信区间包含总体参数真值的次数所占的比例。(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,需要在另一个一假设中进行选择,这就是替换假设。(3分)

三、简答题(每题10分,共30分)

15. 简述按照测量水平区分的四类变量,并举例说明。

(1)定类变量:当变量值的含义仅表示个体的不同类别,而不能说明个体的大小、程度等其它特征时,这种变量称为定类变量。(2分)例如:性别(0.5分)

(2)定序变量:当变量值的含义不仅表示个体的不同类别,还可以区分个体之间大小、程度等序次差异时,这种变量称为定序变量。(2分)例如:学历(0.5分)

(3)定距变量:当变量值不仅可以个体区分为不同类型并进行排序,而且可以确定不同类别之间的数量差别和间隔差距时,这样的变量称为定距变量。(2分)例如:智商(0.5分)

(4)定比变量:除了上述三种变量的全部特征外,还可以计算两个变量值之间的比值时,这样的变量称为定比变量。(2分)例如:收入(0.5分)

16. 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)

17. 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)

四、计算题(每题 15 分,共 30 分)

18. (1)均值:

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

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

(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)

19. A:960(3分)

B:7008(2分)

C:3(3分)

D:36(2分)

E:168(3分)

F:1.90(2分)

试卷代号:1318

座位号

--	--

国家开放大学2020年秋季学期期末统一考试

社会统计学 试题

2021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. 关于条形图的表述,不正确的是()。

- A. 条形图中条形的宽度是固定的
B. 条形图中条形的长度(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第()组。

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. 离散系数
14. 离散变量与连续变量

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. 某汽车生产商欲了解广告费用(万元)对销售量(辆)的影响。收集了过去 12 年的有关数据,通过分析得到:方程的截距为 363,回归系数为 1.42,回归平方和 $SSR=1600$,残差平方和 $SSE=450$ 。要求:

- (1)写出销售量 y 与广告费用 x 之间的线性回归方程。
 - (2)假如明年计划投入广告费用为 50 万,根据回归方程估计明年汽车销售量。
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。

试卷代号:1318

国家开放大学2020年秋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)

2021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. 离散系数:离散系数是一组数据的标准差与该数据均值之比,也称为变异系数。

(5分)

14. 离散变量与连续变量:如果一个变量的变量值是间断的,可以一一列举的,这种变量称为离散变量;(2分)如果一个变量的变量值是连续不断的,即可以取无数多个数值,这种变量称为连续变量。(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)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) χ^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) χ^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 χ^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: $\beta_0 = 363$ (2分) $\beta_1 = 1.42$ (2分)

所以,回归方程为: $y = 363 + 1.42x$ (5分)

(2)当 $x = 50$ 时,

$y = 363 + 1.42 \times 50 = 434$ (辆) (6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	