

2013 年华东师范大学 432 应用统计考研

真题



启航龍圖
SAILING EDUCATION GROUP

2013 年华东师范大学 432 应用统计考研真题

一、单项选择题（本题包括 1~30 题，共 30 个小题，每小题 2 分，共 60 分。在每小题给出的四个选项中，只有一个符合题目要求。先在答题纸上写上序号，再把所选项前的字母填在相应的序号后。）

1. 以下关于抽样误差的说法错误的是（ ）。

- A. 抽样误差是指由于抽选样本的随机性所导致的误差
- B. 由绝对误差和置信度可显示抽样调查方案设计者预期的抽样误差
- C. 由总体目标量的点估计值可显示实际抽样误差的估计值
- D. 总体目标量的估计量的方差越大，表明抽样误差越大

2. 设有一个抽样调查方案，其抽样步骤如下：

- （1）在全国每一个省都独立地实施抽样调查；
- （2）在每一省中，随机抽取 2 个市；
- （3）在每一被抽取中的市中随机抽取 5 个调查小区；
- （4）对每一被抽中的调查小区中，调查该小区全体居民。

该方案采用的抽样方法是（ ）

- A. 二阶抽样
- B. 二阶整群抽样
- C. 分层二阶抽样
- D. 分层二阶整群抽样

3. 右偏分布的特征是（ ）

- A. 偏态系数大于 0
- B. 偏态系数小于 0
- C. 偏态系数等于 0
- D. 只能从直方图上判断，不能用统计量进行描述

4. 以下关于变异系数的说法错误的是（ ）

- A. 变异系数等于标准差除以均值
- B. 变异系数没有单位、不受数据量纲的影响
- C. 变异系数是稳健的统计量
- D. 变异系数反应了数据的波动情况

5. 以下关于数据类型的说法错误的是（ ）。

- A. 温度属于定距类型的数据
- B. 考试成绩（百分制）属于定比类型的数据
- C. 性别属于定类类型的数据
- D. 评级（如：优、良、中、差）属于定序类型的数据

6. 以下说法错误的是（ ）。

- A. 适用于定类数据的图形, 则同样适用于定序数据
- B. 由定序数据可计算标准差
- C. 箱线图能大致反应数据的中心、波动和对称情况
- D. 几何平均值主要用于计算平均比率

7. 已知一总体服从正太分布, 其均值为 μ , 方差为 $\sigma^2(>0)$, 取本 x_1, \dots, x_n , 得到样本均值和

样本方差为 $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ 和 $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$, 以下说法错误的是 ()

- A. μ 的矩估计和极大似然估计都是 \bar{x}
- B. \bar{x} 是 μ 的无偏估计
- C. σ^2 的矩估计和极大似然估计都是 s^2
- D. s^2 是 σ^2 的无偏估计

8. 样本中位数和样本均值可以作为总体中心的估计, 那么, 使得 $\sum_{i=1}^n |x_i - a|$ 达到最小的 a 应是

(), 使得 $\sum_{i=1}^n (x_i - b)^2$ 达到最小的 b 应是 ()

- A. 样本中位数, 样本均值
- B. 样本均值, 样本中位数
- C. 样本中位数, 样本中位数
- D. 样本均值, 样本均值

9. 对原假设和备择假设的选取, 以下说法错误的是 ()

- A. 带等号的假设一般作为原假设
- B. 需要保护的假设一般作为备择假设
- C. 想要证明的假设一般作为备择假设
- D. 选定原假设后, 原假设的对立假设就是备择假设

10. 对于方差已知为 σ_0^2 的正态总体均值的假设检验问题: $H_0: \mu = \mu_0$ vs. $H_1: \mu < \mu_0$, 显著性水平取为 α , 样本量为 n 。当真实的值为 $\bar{\mu} (\bar{\mu} < \mu_0)$ 时, 检验的势为 ()

- A. $\phi\left(\mu_a + \frac{\mu_0 - \bar{\mu}}{\sigma_0} \sqrt{n}\right)$
- B. $\phi\left(\frac{\mu_a}{2} + \frac{\mu_0 - \bar{\mu}}{\sigma_0} \sqrt{n}\right)$
- C. $\phi\left(\mu_a + \frac{\mu_0 - \bar{\mu}}{\sigma_0}\right)$
- D. $\phi\left(\frac{\mu_a}{2} + \frac{\mu_0 - \bar{\mu}}{\sigma_0}\right)$

11. 下列关于回归模型的说法正确的是 ()。

- A. 回归系数的绝对值越大, 表明其对应的自变量对因变量的影响越大

B.采用因变量预测值的区间估计公式不需要用到正态分布的假设

C.建立一个回归模型发现有 2 个自变量不显著,应将二者同时删去然后重新建立回归模型

D.F 比服从 F 分布是基于 ε_i , 独立正态的假设得来的

12.两变量的线性相关系数为 r , 对二者建立一元回归模型 $y = \beta_0 + \beta_1 x + \varepsilon$, 其中的系数均采用最小二乘估计。则以下关系式不成立的是 ()

$$\text{A. } r = \beta_1 \sqrt{\frac{l_{yy}}{l_{xx}}} \quad \text{B. } r^2 = \frac{F}{F + (n - 2)} \quad \text{C. } S_R = r l_{xy} \sqrt{\frac{l_{yy}}{l_{xx}}} \quad \text{D. } r^2 = \frac{S_R}{S_r}$$

13.在单因子方差分析中,因子 A 有 4 个水平,每个水平下各重复 3 次试验,现已求得每个下水平试验结果的样本标准差分别为 1、2、1、4,则误差平方和为 ()。

A.8 B.16 C.22 D.44

14.单因子方差分析中,已知 $S_A = 156$, $S_r = 240$, $f_A = 2$, $f_r = 15$, F 为 ()

A.0.08 B.0.93 C.1.86 D.12.07

15.一个时间序列 $\{x_t\}$ 存在 AR (1) 模型关系: $x_t - 10 = -0.2(x_{t-1} - 10) + \varepsilon_t$ 下说法错误的是 ()

A.相邻的 x_t 之间是负相关关系,不相邻的 x_t 之间不存在相关关系

B. ε_t 之间存在相关关系

C. x_t 的期望值是 10

D. ε_t 的期望值是 0

16.对自变量 x 和因变量 y 建立一元回归模型。若 y 不服从正态分布,应如何处理? 以下有三个说法:

(1) 通常的回归模型中, y 不需要服从正态分布,只需要 ε 服从正态分布

(2) 对 y 作变换为 \bar{y} , 若 \bar{y} 服从正态分布就可将 x 和 \bar{y} 建立通常的回归模型

(3) 若 y 为定型数据,应建立 LOGISTIC 回归模型

那么,这三个说法是否正确? ()

A. (1) 正确, (2) (3) 错误

B. (2) 正确, (1) (3) 错误

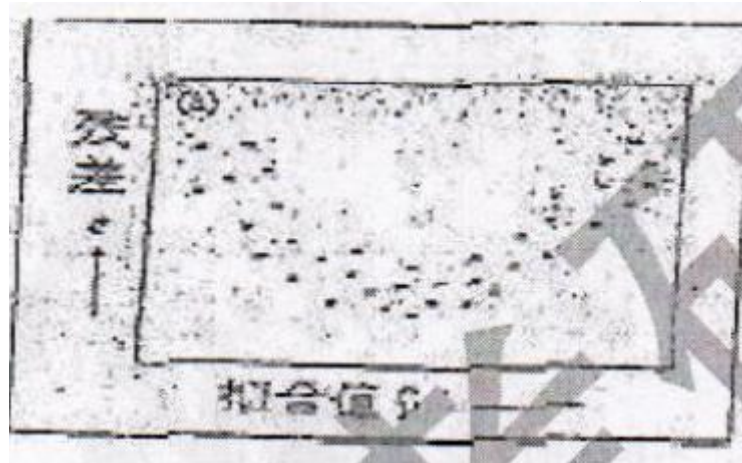
C. (2) (3) 正确, (1) 错误

D. (1) (2) (3) 都错误

17.对自变量 s 和因变量 z 建立曲线回归方程 $\hat{z} - 100 = a \times \exp(-b/s)$ (a 和 b 为回归模型参数估计值, $a > 0, b > 0$), 通常是将曲线回归方程转化为线形回归方程 $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ 后进行参数估计。如何转化? 以下说法正确的是 ()

- A. 应将 z 看做 y , 将 $\frac{1}{s}$ 看做 x
- B. 应将 $z-100$ 看做 y , 将 $esp(-b/s)$ 看做 x
- C. 应将 $\ln(z)$ 看做 y , 将 $\frac{1}{s}$ 看做 x
- D. 应将 $\ln(z-100)$ 看做 y , 将 $\frac{1}{s}$ 看做 x

18. 右图是某一元回归模型的拟合值与残差的图形, 它表明该回归模型存在什么问题? ()



- A. ε 的方差不齐性
 - B. ε 不独立
 - C. ε 不服从正态分布
 - D. x 和 y 之间不是线性模型关系
19. 检验一个总体是否服从正态分布, 可通过以下哪种检验方法实现? ()
- A. 拟合优度检验
 - B. 单位根检验
 - C. 失拟性检验
 - D. 随机游程检验
20. 以下关于抽样分布的说法错误的是 ()。
- A. 抽样分布主要用于评价估计量的效果以及构造置信区间和拒绝域
 - B. 常用的三大抽样分布 (t 分布、 χ^2 分布、 F 分布) 都是基于总体得来的
 - C. 当精确的抽样分布不易推导时, 可以通过随机模拟等方法的近似的抽样分布
 - D. 抽样分布一般与样本量无关
21. 对两个正态总体方差的检验 $H_0: \sigma_1^2 \leq \sigma_2^2$ vs $H_1: \sigma_1^2 > \sigma_2^2$, 有以下三个说法:
- (1) 当 μ_1 和 μ_2 均未知时, 应采用 F 作为检验统计量
 - (2) 当 μ_1 和 μ_2 均已知时, 应采用 F 作为检验统计量

- (3) 当 μ_1 和 μ_2 其中一个已知、另一个未知时, 应采用 F 作为检验统计量
那么, 它们是否正确? ()
- A. (1) 正确, (2) (3) 错误
B. (1) 和 (2) 正确, (3) 错误
C. (2) 和 (3) 正确, (1) 错误
D. (1) (2) (3) 都正确
22. 对一组样本数据: 5、11、11、23、24、35、35、47, 其众数和中位数分别是 ()。
- A. 众数是 11, 中位数是 23.5
B. 众数是 35, 中位数是 24
C. 众数是 11 和 35, 中位数是 23.5
D. 众数是 11 和 35, 中位数是 24
23. 对 X_i 作变换为: $y_i = \frac{x_i}{7} + 1, i=1, \dots, 20$, 那么 ()
- A. x 与 y 的线性相关系数 r 满足: $0 < r < 1$
B. x 的变异系数小于等于 y 的变异系数
C. x 的均值大于 y 的均值
D. x 的标准差大于等于 y 的标准差
24. 设 $[\theta_L, \theta_U]$ 是 θ 的 0.95 置信区间, 以下说法正确的是 ()
- A. 0.95 置信区间是惟一的
B. 置信区间 $[\theta_L, \theta_U]$ 是随机区间
C. 一次抽样后由样本数据计算所得的 $[\theta_L, \theta_U]$ 包含真值 θ 的概率为 0.95
D. 若重复 1000 抽样, 那么所得的 1000 个 $[\theta_L, \theta_U]$ 中一定有 950 个区间包含真值 θ
25. 设 x_1, \dots 是独立同分布的随机变量, 服从标准正态分布, 记 $\bar{x} = \frac{x_1 + \dots}{20}$, 已知标准正态分布的 0.95 分位数 $u_{0.95} = 1.645$, 则 \bar{x} 的分布的 0.95 分位数是 ()
- A. $\frac{1.645}{\sqrt{20}}$ B. $\frac{1.645}{20}$ C. $1.645 \times \sqrt{20}$ D. 1.645×20
26. 设某型号电子元件的寿命 X (单位: 小时) 具有以下密度函数: $p(x) = \begin{cases} \frac{2000}{x^2}, & x > 2000 \\ 0, & x \leq 2000 \end{cases}$,
若一个元件已工作到 3000 小时尚未失效, 则它还能工作 1000 小时的概率是 ()
- A. $\frac{1}{2}$ B. $\frac{2}{3}$ C. $\frac{3}{4}$ D. 信息不足, 无法确定

27. 设 x_1, \dots, x_n ($n > 0$) 的方差都为 1, $Cov(x_i, x_j) = 1$ ($\forall i \neq j$), 那么 x_1, \dots 的均值 \bar{x} 的方差为 ()

- A. $\frac{1}{n}$ B. $\frac{2}{n}$ C. $\frac{1}{2}$ D. 1

28. 设有 10 个产品, 其中 2 个产品为不合格品。从 10 个产品中随机抽取 2 个产品, 则这 2 个产品均为合品的概率是 ()。

- A. 0.02 B. 0.25 C. 0.62 D. 0.64

29. 已知 X 服从参数为 2 的泊松分布, 则 $2X+1$ 的方差为 ()

- A. 5 B. 8 C. 9 D. 16

30. 某项考试为 10 道单项选择题, 每题的选项为 5 项。若考生完全不具备该项考试的知识、只是乱猜, 可答对的题目数期望值是 ()。

- A. 0.2 B. 1.0 C. 2.0 D. 2.5

二、简要回答下列问题 (本题包括 1~4 题共 4 个小题, 每小题 10 分, 共 40 分)

1. 简述点估计优劣的评价标准。
2. 简述似然比检验的统计思想和做法。
3. 给出顺序统计量的定义, 若总体密度函数为 f , 给出第 i 个顺序统计量的密度。
4. 简述方差分析的基本假设, 并回答方差分析主要解决什么统计问题?

三、计算与分析题 (本题包括 1~3 题共 3 个小题, 第 1 小题和第 2 小题每题 15 分, 第 3 小题 20 分, 共 50 分)

1. 研究品牌服装销售额与销售人员数量及商店营业面积之间的关系, 希望建立回归模型:

$y = aX_1 + bX_2 + c + e$, 假设 e 服从 $N(0, a^2)$, 设观察数据为 (Y_k, X_{1k}, X_{2k}) , $k=1, 2, \dots$ 。

(1) 求参数 a, b, c 的最小二乘估计 $\hat{a}, \hat{b}, \hat{c}$ 及其分布。

(2) 若由观察值算出如下数据:

$$\bar{Y} = 783, \bar{X}_1 = 25, \bar{X}_2 = 6.6, \sum_{k=1}^{18} X_{1k}^2 = 12600, \sum_{k=1}^{18} X_{2k}^2 = 800,$$

$$\sum_{k=1}^{18} Y_k^2 = 1160000, \sum_{k=1}^{18} X_{1k} X_{2k} = 3000, \sum_{k=1}^{18} X_{1k} Y_k = 363000, \sum_{k=1}^{18} X_{2k} Y_k = 93000$$

求出回归方程, 并给出其拟合度。

2. 有四个品牌 (因素 1) 的彩电在五个地区 (因素 2) 销售, 对每个品牌在各地区的销售量随机抽取了一些样本, 利用统计软件中的无重复双因素分析功能, 运行后取得以下数据:

UAMMARY	观测数	求和	平均	方差
行 1	5	1721	344.20	233.70
行 2	5	1739	347.80	295.70
行 3	5	1685	337.00	442.50
行 4	5	1424	284.80	249.20

列 1	4	1356	389.00	1224.67
列 2	4	1321	330.23	1464.25
列 3	4	1357	339.25	822.92
列 4	4	1273	318.25	1592.92
列 5	4	1262	315.50	241.67

方差分析

差异源	SS	df	MS	F	P-value	F crit
行	13004.55		4334.85	18.11	0.00	3.49
列	2011.7		502.92	B	0.14	3.26
误差	2872.7	A	239.39			
总计	1788.95	19				

请回答下面几个问题（取显著性水平位 0.05）

- (1) 写出该问题中的原假设与备择假设;
 - (2) 计算上述输出表格中缺少的数据 A、B 的值;
 - (3) 从上述输出结果看, 您得到什么统计结论?
3. 从任意连续分布中抽取独立同分布的随机样本 x_1, x_2, \dots , 记

$p_1 = p(x_1 > 0), p_2 = p(x_1 + x_2 > 0), p_3 = p(x_1 + x_2 > 0, x_1 > 0)$, 请详细推导出用 p_1, p_2 表示 p_3 的关系式。

哎呀这里只有部分真题

加群 779335571

可获取全部真题答案资料及相应答疑

你还在等什么?

启航 2020 应用统计考研交流群 779335571