

山东大学

二〇一九年招收攻读硕士学位研究生入学考试试题

科目代码 432科目名称 统计学

(答案必须写在答卷纸上, 写在试题上无效)

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)

1. 下面哪种抽样方式不属于概率抽样 ()。
- A. 系统抽样 B. 整群抽样 C. 分层抽样 D. 滚雪球抽样
2. 某市场调查公司为了对一家大型商场作顾客满意度调查, 对不同性别和年龄的顾客按事先规定的人数任意进行了一些调查询问, 这种调查属于 ()。
- A. 任意调查 B. 立意调查 C. 配额抽样 D. 整群抽样
3. 指出下面的陈述哪一个是错误的 ()。
- A. 抽样误差是可以避免的 B. 非抽样误差是可以避免的
C. 抽样误差是不可避免的 D. 抽样误差是可以控制的
4. 下面的哪个图形适合描述顺序数据 ()。
- A. 直方图 B. 茎叶图 C. 累计频数分布图 D. 箱线图
5. 下列叙述中正确的是 ()。
- A. 如果计算每个数据与平均数的离差, 则这些离差的和总是等于零
B. 如果考试成绩的分布是对称的, 平均数为 75, 标准差为 12, 则考试成绩在 63~75 分之间的比例大约为 95%
C. 平均数和中位数相等
D. 中位数大于平均数
6. 设总体 $X \sim N(\mu, \sigma^2)$, μ, σ^2 为未知数, X_1, \dots, X_n 是来自 X 的样本, 则下面结论正确的是 ()。

A. $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ 服从 $\chi^2(n-1)$ 分布

B. $S_1^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ 服从 $\chi^2(n-1)$ 分布

C. $\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2$ 服从 $\chi^2(n-1)$ 分布

D. $\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2$ 服从 $\chi^2(n)$ 分布

7. 有一个样本容量为 10 的样本, 其均值为 1300 小时, 方差为 8175.56。若按放回抽样计算, 则样本均值的标准误差为 ()。
- A. 28.35 小时 B. 28.59 小时 C. 29.61 小时 D. 30.02 小时
8. 根据抽样测定 100 名 4 岁男孩发育情况的资料, 平均身高为 95 厘米, 标准差为 4 厘米, 则有 () 的概率可确信 4 岁男孩平均身高在 93.8~96.2 厘米之间。
- A. 68.37% B. 90% C. 95.45% D. 99.74%
9. 设样本 X_1, \dots, X_n 来自于正态总体 $N(\mu, \sigma^2)$, 则常数 $C = ()$ 时, $C \sum_{i=1}^{n-1} (X_{i+1} - X_i)^2$ 为 σ^2 的无偏估计。
- A. $\frac{1}{2(n-1)}$ B. $\frac{1}{2(n+1)}$ C. $\frac{1}{n+1}$ D. $\frac{1}{n-1}$
10. 一个 95% 的置信区间是指 ()。
- A. 总体参数有 95% 的概率落在这一区间内
B. 总体参数有 5% 的概率未落在这一区间内
C. 在用同样方法构造的总体参数的多个区间中, 有 95% 的区间包含该总体参数
D. 在用同样方法构造的总体参数的多个区间中, 有 95% 的区间不包含该总体参数
11. 从一个正态总体中随机抽取一个容量为 n 的样本, 其均值和标准差分别为 33 和 4。当 $n=5$ 时, 构造总体均值的 95% 的置信区间为 ()。
- A. 33 ± 4.97 B. 33 ± 2.22 C. 33 ± 1.65 D. 33 ± 1.96
12. 为估计自考学生的平均年龄, 随机抽取一个 $n=600$ 的样本, 算得样本均值 $\bar{x} = 25.3$ 岁, 总体方差为 $\sigma^2 = 16$, 总体均值 μ 的 95% 的置信区间为 ()。
- A. (22.29, 24.31) B. (23.29, 25.31) C. (24.29, 26.31) D. (25.29, 27.31)
13. 在总体均值和总体比例的区间估计中, 边际误差由 ()。
- A. 置信水平确定 B. 统计量的抽样标准差确定
C. 置信水平和统计量的抽样标准差确定 D. 统计量的方差确定
14. 在假设检验中, 接受原假设时 ()。
- A. 可能会犯第一类错误 B. 可能会犯第二类错误
C. 可能会犯第一、二类错误 D. 不会犯错误
15. 设样本 X_1, X_2, \dots, X_9 来自总体 $N(\mu, 0.04)$, 在显著性水平 $\alpha = 0.05$ 的条件下, 假设检

验 $H_0: \mu \leq 0.5, H_1: \mu > 0.5$, 若总体均值的真实值为 $\mu = 0.65$, 则此时的取伪概率为 ()。

- A. $\Phi(0.605)$ B. $\Phi(-0.605)$ C. $\Phi(1.65)$ D. $\Phi(-1.65)$

16. 在对两个广告效果的电视评比中, 每个广告在一周的时间内播放 6 次, 然后要求看过广告的人陈述广告的内容。记录的资料如下表:

广告	看过广告的人数	回想起内容的人数
A	150	63
B	200	60

则在 $\alpha = 0.05$ 的显著性水平下, 检验两个广告的回想比例没有差别, 即检验假设

$H_0: \pi_1 - \pi_2 = 0, H_1: \pi_1 - \pi_2 \neq 0$, 得到的结论是 ()。

- A. 拒绝 H_0 B. 不拒绝 H_0
C. 可以拒绝也可以不拒绝 H_0 D. 可能拒绝也可能不拒绝 H_0

17. 下面的各问题中, 哪个不是相关分析要解决的问题 ()。

- A. 判断变量之间是否存在关系
B. 判断一个变量数值的变化对另一个变量的影响
C. 判断变量之间的关系强度
D. 判断样本所反映的变量之间的关系能否代表总体变量之间的关系

18. 根据最小二乘法拟合直线回归方程是使 ()。

- A. $\sum (y_i - \hat{y}_i)^2 = \text{最小}$ B. $\sum (y_i - \hat{y}_i) = \text{最小}$
C. $\sum (y_i - \bar{y}_i)^2 = \text{最小}$ D. $\sum (y_i - \bar{y}_i) = \text{最小}$

19. 某汽车生产商欲了解广告费用 (x) 对销售量 (y) 的影响, 收集了过去 12 年来的有关数据。经计算得到下面的方差分析表 ($\alpha = 0.05$):

变差来源	df	SS	MS	F	Significance F
回归	1	1602708.6	1602708.6	()	2.17E-09
残差	10	40158.07	()	()	
总计	11	1642866.67	()	()	

根据上表计算的相关系数为 ()。

- A. 0.9844 B. 0.9855 C. 0.9866 D. 0.9877

20. 如果两个变量之间完全相关, 则以下结论中不正确的是 ()。

- A. 相关系数 $r = 1$ B. 判定系数 $R^2 = 1$
C. 回归系数 $\beta = 0$ D. 估计标准误差 $s_y = 0$

21. 在多元线性回归分析中, 如果 F 检验表明线性关系显著, 则意味着 ()。

- A. 在多个自变量中至少有一个自变量与因变量的线性关系显著
B. 所有的自变量与因变量之间的线性关系都显著
C. 在多个自变量中至少有一个自变量与因变量的线性关系不显著
D. 所有的自变量与因变量之间的线性关系都不显著

22. 在单因素方差分析中, 涉及的两个变量是 ()。

- A. 数值型变量
B. 分类型变量
C. 一个分类型自变量和一个数值型因变量
D. 一个数值型自变量和一个分类型因变量

23. 有交互作用的双因素方差分析, 是指用于检验的两个因素 ()。

- A. 对因变量的影响是独立的 B. 对因变量的影响是有交互作用的
C. 对自变量的影响是独立的 D. 对自变量的影响是有交互作用的

24. 指数平滑法适合于预测 ()。

- A. 平稳序列 B. 非平稳序列 C. 有趋势成分的序列 D. 有季节成分的序列

25. 对时间序列数据作季节调整的目的是 ()。

- A. 消除时间序列中季节变动的影响 B. 描述时间序列中季节变动的影响
C. 消除时间序列中趋势的影响 D. 消除时间序列中随机变动的影响

26. 设 A, B 为两个事件, 且 $B \subset A$, 则下列各式中正确的是 ()。

- A. $P(A \cup B) = P(A)$ B. $P(AB) = P(A)$
C. $P(B|A) = P(B)$ D. $P(B - A) = P(B) - P(A)$

27. 设 A, B 是两个事件, 且 $P(A) \leq P(A|B)$, 则有 ()。

- A. $P(A) = P(A|B)$ B. $P(B) > 0$
C. $P(A) \geq P(A|B)$ D. 前三者都不一定成立

28. 设 X 是一个随机变量, x_0 为任意的实数, EX 是 X 的数学期望, 则 ()。

- A. $E(X - x_0)^2 = E(X - EX)^2$ B. $E(X - x_0)^2 \geq E(X - EX)^2$
C. $E(X - x_0)^2 < E(X - EX)^2$ D. $E(X - x_0)^2 = 0$

29. 设随机变量 X 服从伯努利分布, 且 $P(X = 0) = 0.3, P(X = 1) = 0.7$, 则 X 的方差等于 ()。

- A. 0.21 B. 0.49 C. 0.7 D. 0.3

30. 随机变量 $X \sim N(2, 4)$, 且 $aX + b \sim N(0, 1)$, 则 ()。

A. $a=2, b=-2$

B. $a=-2, b=-1$

C. $a=\frac{1}{2}, b=-1$

D. $a=\frac{1}{2}, b=1$

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. 常用的概率抽样方法有哪些？各自的含义如何？
2. 相关分析与回归分析有何区别？
3. 在拟合时间序列长期趋势方程时，如何选择合适的趋势线？
4. 联合概率分布函数有哪些性质？

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 20 分，第 3 小题 10 分，共 50 分）。

1. 随机挑选 8 个人，分别测量了他们在早晨起床时和晚上就餐时的身高（cm），得到以下数据：

序号	1	2	3	4	5	6	7	8
早上(x_i)	172	168	180	181	160	163	165	177
晚上(y_i)	172	167	177	179	159	161	166	175

假设各对数据的差 $D_i = X_i - Y_i$ ($i=1, 2, \dots, 8$) 是来自正态总体 $N(\mu_D, \sigma_D^2)$ 的样本， μ_D, σ_D^2 均未知，

- (1) 计算样本均值和样本方差；
- (2) 在显著性水平 $\alpha = 0.05$ 下，是否可以认为早晨的身高比晚上的身高要高？

$$(t_{0.05}(7) = 1.8946, t_{0.025}(7) = 2.36)$$

2. 在单因子方差分析中，因子 A 有三个水平，每个水平各做 4 次重复试验，请完成下列方差分析表，并在显著性水平 $\alpha = 0.05$ 下对因子 A 是否显著做出检验。

方差分析表

来源	平方和	自由度	均方和	F 比
因子 A	4.2			
误差 e	2.5			
总计	6.7			

已知： $F_{0.05}(2, 9) = 4.26$, $F_{0.05}(9, 2) = 19.38$

3. 已知随机变量 X 的概率密度为：

$$f(x) = \begin{cases} ax + b, & 0 < x < 1 \\ 0, & \text{其他} \end{cases}$$

且 $P\left\{X > \frac{1}{2}\right\} = \frac{5}{8}$ ，求：(1) 系数 a, b; (2) $P\left\{\frac{1}{4} < X \leq \frac{1}{2}\right\}$; (3) 分布函数 $F(x)$ 。