

# 2012 年华东师范大学 432 应用统计考研

## 真题



启航教育

2012 年华东师范大学 432 应用统计考研真题

一、单项选择（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，现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. 以下关于调查的说法，错误的是（ ）。
  - A. 全面调查不会产生误差
  - B. 抽样调查具有抽样误差，也可能有非抽样误差
  - C. 正式调查前最好先进行预调查，一方面完善问卷，另一方面可搜集一些数据用于抽样设计
  - D. 实施调查前应先设计调查方案，明确调查目的、内容、抽样方案、调查方法、数据分析方法等
2. 以下属于非概率抽样方法的是（ ）
  - A. 简单随机抽样
  - B. 分层抽样
  - C. 整群抽样
  - D. 滚雪球抽样
3. 关于数据的预处理，以下说法错误的是（ ）
  - A. 数据的审核是数据预处理的一项内容，对二手数据的审核主要有适用性审核和时效性审核
  - B. 数据的筛选是数据预处理的一项内容，可以筛选出符合条件的数据
  - C. 数据的排序是数据预处理的一项内容。也可以作为数据分析的一项内容
  - D. 数据变换是数据预处理的一项内容、著名的 BOX-COX 变换可改善方差齐性，但无法改善正态性
4. 以下不适用于展示定性数据的图表是（ ）。
  - A. 频数频率分布表
  - B. 柱状图
  - C. 环形图
  - D. 直方图
5. 箱线图（box plot）中不包括以下哪个统计量（ ）。
  - A. 中位数
  - B. 均值
  - C. 上四分位差
  - D. 变异系数
6. 为比较一个班级中男生成绩波动与女生成绩波动的大小，以下哪个统计量最合适（ ）。
  - A. 极差
  - B. 标准差
  - C. 四分位差
  - D. 变异系数
7. 以下关于参数估计的说法，错误的是（ ）。
  - A. 无偏性是指估计量的期望等于被估参数
  - B. 矩估计的原理是建立总体矩与被估参数的关系，然后用样本矩去代替总体矩，从而得到估计量

- C.均方误差越大,说明估计量的效果越好
- D.已知 $\hat{\theta}$ 是 $\theta$ 的极大似然估计,则 $\hat{\theta}^3$ 是 $\theta^3$ 的极大似然估计
- 8.已知总体均值为 $\mu$ ,总体方差为 $\sigma^2 (>0)$ ,取样本 $x_1, \dots, x_n$ ,得到样本均值和样本方差分别为 $\bar{x}$ 和 $s^2$ ,以下说法错误的是( )
- A. $\bar{x}$ 的期望值是 $\mu$
- B. $s^2$ 的期望值是 $\sigma^2$
- C. $\bar{x}$ 的方差是 $\sigma^2 / (n-1)$
- D. $\bar{x}$ 的均方误差是 $\sigma^2 / n$
- 9.从一个正态总体 $N(\mu, \sigma^2)$ ( $\mu$ 和 $\sigma^2$ ( $\sigma^2 > 0$ )均未知)中抽取一个容量为 $n$ ( $n > 1$ )的样本,得到样本均值和样本方差分别为 $\bar{x}$ 和 $s^2$ ,则 $\sigma^2$ 的置信度为95%的区间估计是( )
- A.  $[(n-1)s^2 / \chi_{0.025}^2(n-1), (n-1)s^2 / \chi_{0.975}^2(n-1)]$
- B.  $[(n-1)s^2 / \chi_{0.95}^2(n-1), (n-1)s^2 / \chi_{0.05}^2(n-1)]$
- C.  $[ns^2 / \chi_{0.975}^2(n), ns^2 / \chi_{0.025}^2(n)]$
- D.  $[ns^2 / \chi_{0.95}^2(n), ns^2 / \chi_{0.05}^2(n)]$
- 10.下列关于假设检验的说法,正确的是( )。
- A.检验的势是当备择假设为真时接受原假设的概率
- B.p值越小,数据多提供的否定原假设的证据就越强
- C.显著性检验控制了第一类错误概率和第二类错误概率
- D.抽取样本,得到p值为0.03,则应拒绝原假设
- 11.下列关于相关系数r的说法,错误的是( )
- A.  $|r|$ 小于等于1
- B.  $|r|$ 越大,表示两个变量的线性相关程度越大
- C.  $r=0$ 表示两个变量无任何关系
- D.  $r < 0$ 表示两个变量负线性相关
- 12.一元线性回归分析中,方差 $\sigma^2$ 的估计为( )
- A.  $SST / f_r$       B.  $SST / f_k$       C.  $(SST / f_r) / (SST / f_g)$       D.  $SST / f_g$
- 13.因子A有r个水平,每个水平各做m次重复试验,则在单因素方差分析中误差的偏差平

方和 (SSE) 的自由度为 ( )

- A.  $r-1$       B.  $mr-1$       C.  $r(m-1)$       D.  $m(r-1)$

14. 回归分析中, 残差检验的功能不包含以下哪一项? ( )

- A. 回归模型的显著性检验      B. 方差齐性检验  
C. 独立性检验      D. 正态性检验

15. 对一个正态总体  $N(\mu, \sigma_0^2)$  ( $\sigma_0^2 (>0)$  未知) 的均值  $\mu$  进行假设检验。设

$H_0: \mu = \mu_0$   $H_1: \mu \neq \mu_0$ 。抽取容量为  $n (>1)$  的样本, 得到样本均值和样本标准差分别为  $\bar{x}$  和

2, 则显著性水平  $\alpha$  的拒绝域应为 ( )。

- A.  $\{|\bar{x} - \mu_0| \geq t_{1-\alpha/2}(n-1) \cdot s / \sqrt{n}\}$   
B.  $\{|\bar{x} - \mu_0| \geq u_{1-\alpha/2} \cdot \sigma / \sqrt{n}\}$   
C.  $\{|\bar{x} - \mu_0| \geq t_{1-\alpha}(n-1) \cdot s / \sqrt{n}\}$   
D.  $\{|\bar{x} - \mu_0| \geq u_{1-\alpha} \cdot \sigma / \sqrt{n}\}$

16. 已知  $x_1, \dots$  的样本方差为  $s^2$ , 设  $y_i = 2x_i - 4, i=1, \dots$ 。则  $y_1, \dots$  的样本方差为 ( )

- A.  $s^2$       B.  $2s^2 - 4$       C.  $2s^2$       D.  $4s^2$

17. 对同一组数据得到几个一元回归模型, 则下列说法错误的是 ( )

- A. 调整的  $R^2$  越大的模型效果越好  
B. SSE 越小的模型效果越好  
C. SSR/SST 越小的模型效果越好  
D. MSE 越小的模型效果越好

18. 在作假设检验时, 若接受原假设可能 ( )。

- A. 犯第一类错误  
B. 犯第二类错误  
C. 既犯第一类错误, 又犯第二类错误  
D. 不犯任一类错误

19. 有一个时间序列预测模型为  $X_t = X_{t-1} - 0.1X_{t-2}$ , 若已知  $X_0 = 0, X_1 = -1$ , 则  $\hat{X}_3 = ( )$ 。

- A. -1.1      B. -1.0      C. -0.9      D. 0.9

20. 以下是回归分析的一些内容: (1) 回归模型的显著性检验; (2) 回归系数的显著性检验; (3) 残差检验; (4) 回归模型的效果评价。那么, 进行回归分析正确的顺序应是 ( )。

- A. (2) - (1) - (3) - (4)  
B. (1) - (2) - (3) - (4)

C. (2) - (1) - (4) - (3)

D. (1) - (2) - (4) - (3)

21. 对一个正态总体  $N(\mu, \sigma_0^2)$  ( $\sigma_0^2 (> 0)$  已知) 的均值  $\mu$  进行估计, 希望所得的置信度为 95% 的区间估计的长度不超过 20, 则所需要的样本量至少为 ( )。

A.  $(u_{0.975}\sigma_0 / 10)^2$

B.  $(u_{0.975}\sigma_0 / 20)^2$

C.  $(u_{0.95}\sigma_0 / 10)^2$

D.  $(u_{0.95}\sigma_0 / 20)^2$

22. 有 4 位同学的某一门课程成绩分别是 71, 82, 87, 90, 则他们成绩的中位数是 ( )。

A. 81

B. 82.5

C. 84.5

D. 87

23. 设回归模型为  $y = \beta_0 + \beta_1 x + \varepsilon$ , 则下列说法错误的是 ( )

A.  $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

B.  $\hat{\beta}_1 \cdot r \geq 0$ , 其中  $r$  是  $x$  与  $y$  的相关系数

C. 若取  $z = (x - \bar{x}) / (s_x)$  ( $s_x (> 0$  且  $s_x \neq 1)$  为  $x$  的样本标准差), 建立回归模型  $y = r_0 + r_1 z + \varepsilon$ ,

则  $\hat{r}_0 = \bar{y}, \hat{r}_1 = \hat{\beta}_1$

D.  $y_1, \dots$  不一定独立同分布

24. 关于两个正态总体 (方差均未知) 的检验  $H_0: \mu_x = \mu_y + 2, H_1: \mu_x > \mu_y + 2$ , 以下说法正确的是 ( )

A. 若  $x$  与  $y$  是非成对数据, 则进行方差相等下的均值检验

B. 若  $x$  与  $y$  是非成对数据, 则应先进行方差相等的检验, 然后根据方差检验的结果进行均值检验

C. 若  $(x, y)$  是成对数据, 只需令  $z = x - y$ , 然后检验  $H_0: \mu_z = 0, H_1: \mu_z > 0$

D. 若  $(x, y)$  是成对数据, 将其看做非成对数据而采用非成对数据的方法进行检验也是恰当的

25. 关于单因子方差分析的说法中, 错误的是 ( )。

A. 总的偏差平方和 (SST) 可分解为因子的偏差平方和 (SSA) 和误差的偏差平方和 (SSE)

B. 因子的偏差平方和反应了因子水平之间的差异所导致的偏差

C. 误差的偏差平方和反应了随机因素所导致的偏差

D. 若因子的偏差平方和大于误差的偏差平方和, 则说明因子的影超越了随机因素的影响水平, 即说明该因子显著

26 已知  $A \subset B, P(A|B) = 0.8, P(B) = 0.3$ , 则  $P(B|A) = ( )$

A.0.24                      B.0.30                      C.1.00                      D.无法确定

27.从 1000 米长的一条布匹中随机抽取 5 米进行检验,若 5 米中无缺陷才可以接受这条布匹。假设送检布匹平均每米有 1 个缺陷,则这条布匹被拒收的概率为 ( )。

A.0.10                      B.0.63                      C.0.86                      D.0.99

28.设  $u_a$  是标准正态分布的  $a$  分位数,即设  $X$  服从标准正态分布,  $P(X \leq u_a) = a$ , 则有 ( )

A.  $u_{0.6} < 0$                       B.  $u_a + u_{1-a} = 0$                       C.  $u_a - u_{1-a} = 0$                       D.  $x - u_a + u_{1-a} = 1$

29.某产品的寿命服从指数分布  $\text{Exp}(3)$ , 平均寿命为 3 小时, 则该产品寿命超过 1 小时的概率为 ( )。

A.0.05                      B.0.28                      C.0.72                      D.0.95

30.X 与 Y 为两个独立的随机变量, X 的方差为 1, Y 的方差为 2, 则  $3Y-2X$  的方差为 ( )。

A.4                      B.8                      C.14                      D.2

**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.假设检验中需要使用 P 值, 请说明 P 值含义, 并说明如何使用 P 值?
- 2.给出样本的经验分布函数的定义, 并计算其期望与方差。
- 3.简述充分统计量的定义, 样本均值是不是正态总体均值的充分统计量? 为什么。

**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.随机抽查了 50 名小学生的身高数据如下(单位: cm) 117 122 124 129 139 106 117 130 122 125 108 131 125 117 122 133 126 122 118 108 110 118 123 126 133 134 127 123 118 112 112 134 127 123 119 113 120 123 127 135 137 114 120 128 124 115 139 128 124 121。请先适当分组, 再作出频率直方图, 最后根据图形, 判断这组数据的分布形状。

2.货车的行驶时间与行驶距离的远近及运送货物的次数有关, 下表给出的资料是从某运输队收集来的。

Y: 小时	7.3	4.8	9.5	5.2	7.2	6.2	7.4	6.6	7.6	6.3
X <sub>1</sub> : 100 公里	1	0.6	1.5	0.7	0.5	0.8	0.75	0.75	0.9	0.9
X <sub>2</sub> : 次数	4	2	4	2	4	2	3	4	4	2

根据 Excel 输出结果, 回答后面的几个问题。

- (1) 将上述表中的数据 A, B, C, D, E 求出;
- (2) 您认为用线性回归模型来拟合原始数据是否合适? 为什么?
- (3) 写出  $y$  与  $X_1$  及  $X_2$  之间线性回归方程, 并解释各个回归系数的意义;
- (4) 您认为各个回归系数是不是显著不等于 0? 为什么? (取显著性水平为 0.05)
- (5) 找出各个回归参数的 95% 区间估计。

回归统计	
Multiple R	0.913

R Square	0.833
Adjusted R Square	A
标准误差	0.616
观测值	10

方差分析

	df	SS	MS	F	P-value
回归分析	2	13.25	6.63	C	0.002
残差	B	2.66	0.38		
总计	9	15.91			

	Coefficients	标准误差	t Stat	P-value	Lower95%	Upper95%
Intercept	2.214	0.802	2.762	0.028	0.318	4.110
X Variable 1	2.688	0.792	D	0.012	0.816	4.560
X Variable 2	E	0.219	3.449	0.011	0.237	1.271

3.在寿命数据分析中,记  $X$  为寿命,假设  $X$  的分布函数为  $F(x)$ , 密度函数为  $f(x)$ , 在研究中通常还引入生存函数  $S(x) = P(X \geq x)$ , 及失效率函数  $\lambda(x) = \frac{f(x)}{1-F(x)}$ 。请导出用失效率函数  $\lambda(x)$  表示  $F(x)$ 、 $S(x)$ 、 $f(x)$  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**