

2015 年华东师范大学 432 应用统计考研

真题



启航龍圖
SAILING EDUCATION GROUP

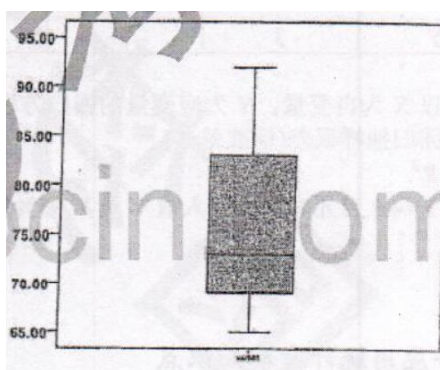
2015 年华东师范大学 432 应用统计考研真题

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。）

1. 下列关于调查中非抽样误差的说法，错误的是：（ ）

- A. 非抽样误差是指除抽样误差以外的、由于各种原因而引起的误差
- B. 问卷设计不合理所导致的误差属于非抽样误差
- C. 非抽样误差理论上可以消除，但在实际调查的各个阶段都有可能产生、很难消除
- D. 非抽样误差可以计算和控制

2. 右图显示了一个班级学生某课程成绩的箱线图（box plot），据此判断其对称情况是：（ ）



- A. 对称
- B. 左偏
- C. 右偏
- D. 无法判断

3. 现了解到 2014 年世界杯 6 位巴西队参赛球员的年龄：21、22、25、28、30、34，其下四分位数是：（ ）

- A. 21.25
- B. 21.50
- C. 21.75
- D. 31.00

4. 对数据 x_1 作线性变换，令 $y_1 = 2x_1 - 3$ 。则下列由 y_1 计算所得的统计量中，哪一个会与由 x_1 计算所得的统计量相同？（ ）

- A. 均值
- B. 方差
- C. 变异系数
- D. 偏态系数

5. 下列关于数据标准化的功能描述中，哪一项不成立？（ ）

- A. 数据标准化可将各变量数据转化为相同的分布
- B. 标准化后的数据消除了量纲和单位
- C. 各变量数据经标准化后就具有可比性了
- D. 各评价指标变量经标准化后就可进行加减等运算、从而汇总出综合性评价指标

6. 以下适用于多个总体比较的图形是：（ ）

- A. 环形图
- B. 箱线图
- C. 雷达图
- D. 上述三个都适用

7. 下列描述分布离散程度的统计量中，哪一个具有稳健性？

- A. 标准差
- B. 四分位差
- C. 极差
- D. 变异系数

8. 下列关于定序数据（顺序数据）的描述，错误的是：（ ）

- A. 可进行=或 \neq 、 $>$ 或 $<$ 的运算

- B.在频数频率分布表中应按顺序列出其取值，并给出累积频率
C.适用于作柱状图、饼图、环形图和直方图
D.适用于计算众数、中位数和四分位数
- 9.统计学中的许多“系数”是不带量纲可直接进行比较的，下列哪个“系数”不具有这种特征：
()
A.变异系数 B.偏态系数 C.一阶自相关系数 D.回归系数
- 10.设总体 X 服从 $[a, b]$ 上的均匀分布， a 和 b 未知。抽取样本 x_1, \dots, x_n ，则 b 的极大似然估计量为：()
A. $\min\{x_1, \dots, x_n\}$ B. \bar{x}
C. $\max\{x_1, \dots, x_n\}$ D. $\min\{x_1, \dots, x_n\} + \max\{x_1, \dots, x_n\}$
- 11.检验所建立的回归模型是否满足方差齐性，可以采用下列哪种检验方法？()
A. Spearman 等级相关系数检验
B. Durbin-Watson 检验
C. Shapiro-Wilk 检验
D. Kolmogorov-Smirnov 检验
- 12.分赌本问题是历史上非常著名的概率问题：甲和乙两人各出本金 1000 元，约定谁先胜三局谁就可拿走全部本金。假定每局中两人获胜的概率相等。现在进行了三局，胜了两局，乙胜了一局，但因故需停止赌博，问题就在于这本金 2000 元应如何分配。一种分配方案是：如果赌博继续进行下去，按照各自获胜的概率分配本金。请问：按此方案，甲可分得的本金比例是：()
A. 0.25 B. 0.50 C. 0.65 D. 0.75
- 13.以下关于大数定律的说法，错误的是：()
A. 大数定律使得大量重复相同的试验所得的平均值会趋近于期望值
B. 在偶然性和必然性之间搭起桥梁的是大数定律
C. 大数定律不能保证频率会趋近于概率
D. 赌场和保险公司都需要依赖大数定律实现期望收益
- 14.以下关于假设检验的说法，错误的是()
A. 在给定的样本量下，假设检验中的第 I 类错误和第 II 类错误是此消彼长的关系
B. 显著性检验中，显著性水平 α 是“有多少证据才足够”的标准，即规定了多大的概率才算是“小概率”
C. 检验得出某个假设“显著”，并不代表该假设“重要”，而只代表“光靠机遇不容易发生”
D. 显著性水平 α 的值越大，则对原假设的保护越强，即越不容易拒绝原假设
- 15.对总体均值进行估计，其估计误差与下列哪个量无关？()

- A.总体方差 B.置信水平 C.样本量 D.总体均值的真实值
- 16.下列关于分布关系的描述, 错误的是: ()
- A.服从超几何分布的随机变量在一定条件下会近似服从二项分布
B.服从泊松分布的随机变量在一定条件下会近似服从二项分布
C.由服从正态分布的随机变量可构造出卡方分布、t 分布和 F 分布
D.若一段时间内受冲击的次数服从泊松分布, 则两次冲击的间隔时间服从指数分布
- 17.设 Z 服从标准正态分布, 记 u_α 为标准正态分布的分位数, 即 $P(Z \leq u_\alpha) = \alpha$, 则下列式子中不成立的是: ()
- A. $u_\alpha + u_{1-\alpha} = 0$ B. $u_\alpha > u_\beta$ ($\alpha < \beta$)
C. $u_{0.5} = 0$ D. $P(Z > u_{1-\alpha/2}) + P(Z < u_{\alpha/2}) = \alpha$
- 18.股票 X 的方差为 1, 股票 Y 的方差为 2, 二者协方差为-2。若设计一个 X 和 Y 的投资组合方案使得组合资产的方差达到最小, 在该投资组合中股票 X 应占多少比重? ()
- A.37% B.43% C.57% D.77%
- 19.假设有两只灯泡的寿命都服从同一个指数分布, 现在让灯泡 1 工作了 100 个小时且仍可继续工作, 灯泡 2 未被使用过。记 P 为灯泡 1 以后能工作超过 100 个小时的概率, Q 为灯泡 2 以后能工作超过 100 个小时的概率, 那么有: ()
- A.P=Q B.P<Q C.P>Q D.P 与 Q 的关系不能确定
- 20.若 X 和 Y 的 Pearson 线性相关系数为 0, 则下列说法错误的是: ()
- A.X 和 Y 可能存在一定的相关关系 B.X 和 Y 不相关
C.X 和 Y 不存在线性相关关系 D.对 X 和 Y 建立线性回归模型没有意义
- 21.设 X_1, \dots, X_n 独立同分布, 其均值为 \bar{X} 。则由中心极限定理, 当 n 充分大时, 下列哪一项不成立? ()
- A.不论 X_1, \dots, X_n 服从什么分布, \bar{X} 近似服从正态分布
B.若 X_1, \dots, X_n 服从参数为 n 和 p ($0 < p < 1$) 的二项分布, 则 \bar{X} 近似服从正态分布
C. \bar{X} 的分布不一定以 $E(X_1)$ 为中心
D. \bar{X} 会趋近于一个常数
- 22.为了保证未来水域的扇贝供应, 美国渔业和野生动植物服务署 (USFWS) 要求任何商业捕鱼船在捕获时, 每个扇贝肉的平均重量至少是 0.5 盎司, 假设 USFWS 检查员登上了一条捕捞了 1000 万个扇贝的渔船, 并且 UBFN 检查员采集了 1000 个扇贝的随机样本, 得扇贝肉重量均值是 $\bar{x}=0.48$ 盎司, 样本标准差是 $s=0.3$ 盎司。设 w 表示这艘船上的 1000 万个扇贝肉的平均重量, 那么 w 小于或等于 0.50 盎司的可信度是多少? ()
- A.低于 90% B.[90%, 95%] C.(95%, 97.5%] D.高于 97.5%

23. 章鱼保罗曾因准确预测了一些世界杯比赛而声名鹊起。对下列假设进行检验:

H_0 : 保罗不具有预测能力 (纯属乱猜), H_1 : 保罗具有预测能力

假设保罗在 8 场比赛中预测对了 7 场, 那么检验的 P 值应为: ()

- A. 0.0039 B. 0.0078 C. 0.0313 D. 0.0352

24. 已知随机变量 X 服从正态分布 $N(\mu, \sigma^2)$, 设随机变量 Y 为 X 的指数函数, 即 $Y = \exp\{X\}$, 则 Y 的期望值为: ()

- A. $\mu + \sigma^2$ B. $\exp\left\{\mu + \frac{\sigma^2}{2}\right\}$ C. $\exp\left\{\frac{\mu}{2\sigma^2} + 1\right\}$ D. $\exp\{\mu\}$

25. 为评估因子 A (3 个水平) 和因子 B (3 个水平) 对某个生产指标 y 是否有影响, 进行了 9 次试验, 结果如下:

	B1	B2	B3	行之和
A1	1	2	2	5
A2	3	5	1	9
A3	4	4	5	13
列之和	8	11	8	27

请计算因子 A 的离差平方和: ()

- A. 2.00 B. 3.56 C. 10.67 D. 20.00

26. 常用样本方差来估计总体方差。样本方差不满足以下哪个性质? ()

- A. 相合性 B. 有效性 C. 无偏性 D. 渐近正态性

27. 在回归分析中, 已知所得的合 $\delta > 0$ 。记 L 表示对 y_0 的个别值进行预测所得的区间长度, 记 K 表示对 y_0 的平均值进行预测所得的区间长度, 则 L 和 K 的关系为: ()

- A. $L > K$ B. $L = K$
C. $L < K$ D. 上述三种情况都有可能

28. 关于回归分析中决定系数 (R Square) 的描述, 错误的是: ()

- A. 决定系数的功能是评价模型拟合效果, 是一个绝对的拟合效果的度量
B. 决定系数表示由自变量的线性回归模型解释的因变量的变化占总变化的比例
C. 决定系数表示引入自变量后模型的残差平方和减小的比例
D. 如果有 2 个及以上的自变量, 应采用调整的复决定系数 (Adjusted R Square) 来评判模型拟合效果

29. 已知总体 X 服从 $N(\mu_1, \sigma^2)$, 总体 Y 服从 $N(\mu_2, \sigma^2)$, 现从总体 X 中抽取 n_1 个观测值, 其样本方差记为 S_1^2 ; 从总体 Y 中抽取 n_2 个观测值, 其样本方差记为 S_2^2 。已知两组样本相互独立, σ^2 未知, 那么 $\mu_1 - 2\mu_2$ 的置信度为 $1-\alpha$ 的区间估计的长度应为: ()

- A. $2t_{1-\alpha/2}(n_1+n_2-2)S_W\sqrt{1/n_1+1/n_2}$, 其中 $S_W^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$
- B. $2t_{1-\alpha/2}(n_1+n_2-1)S_W\sqrt{1/n_1+1/n_2}$, 其中 $S_W^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$
- C. $2t_{1-\alpha/2}(n_1+n_2-2)S_W\sqrt{1/n_1+1/n_2}$, 其中 $S_W^2 = \frac{(n_1-1)S_1^2 + 4(n_2-1)S_2^2}{n_1+n_2-2}$
- D. $2t_{1-\alpha/2}(n_1+n_2-1)S_W\sqrt{1/n_1+1/n_2}$, 其中 $S_W^2 = \frac{(n_1-1)S_1^2 + 4(n_2-1)S_2^2}{n_1+n_2-2}$

30. 设总体中含有 N 个单元, 各单元取值为 y_1, \dots, y_N , 总体均值为 $\bar{Y} = \sum_{i=1}^N y_i / N$, 总体方差为 $S^2 = \sum_{i=1}^N (y_i - \bar{Y})^2 / (N-1)$ 。采用简单随机抽样从中抽取容量为 n 的样本 x_1, \dots, x_n , 记样本均值为 $\bar{x} = \sum_{i=1}^n x_i / n$, 样本方差为 $s^2 = \sum_{i=1}^n (x_i - \bar{x})^2 / (n-1)$, 那么下列说法错误的是: ()

- A. $E(\bar{x}) = \bar{Y}$
- B. $D(\bar{x}) = \frac{(N-n)S^2}{nN}$
- C. 由于是从有限总体中抽样, 故一般 x_1, \dots, x_n 并不独立
- D. 若要求 $P(|\bar{x} - \bar{Y}| \leq d) = 0.95$, 则样本量应取为 $n \geq 1.96^2 S^2 / d^2$

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. 显著性检验中, 原假设和备择假设的地位有何差异? 设置原假设应该遵循哪些原则?
2. 对于二维随机变量 (X, Y) , X 与 Y 相互独立是如何定义的? 当 (X, Y) 为连续型随机变量时, 独立性定义有何等价形式?
3. 区间估计中的置信度 $1-\alpha$ 应如何理解? 这个置信度 $1-\alpha$ 与假设检验中的什么量会有一些的关系?
4. 简述极大似然估计的基本思想及其性质。在现实问题中采用极大似然估计方法进行估计, 面临的最大问题是什么?

三、计算与分析题 (本题包括 1-3 题共 3 个小题, 第 1 小题 10 分, 第 2 和第 3 小题每题 20 分, 共 50 分)。

1. 抽样调查中有著名的沃纳随机化回答模型。假定对学生作弊问题进行调查, 设置两种问题:

问题 A: 您曾经作过弊, 是吗?

问题 B: 您从未作过弊, 是吗?

调查者事先在一个袋子里放置红色和白色两种颜色的球, 球的数目与被调查者人数一致, 其中比例 p ($0.5 < p < 1$) 的球为红色, 所有被调查者一起去抽取球, 每人抽一个, 抽好后把球藏起来入座 (不让别人知道他们所抽球的颜色)。现在要求抽中红色球的被调查者回答问题 A, 抽中白色球的被调查者回答问题 B。假定通过调查了解到全体被调查者回答“是”

的概率为 w ，请推导这些学生曾作过弊的概率 π 的表达式（用 p 和 w 表示，请具体写出推导过程）。

2. 调查了游泳、打篮球、骑自行车等 3 种不同的运动在 30 分钟内消耗的热量（单位：卡路里）。假设这些运动 30 分钟内消耗的热量满足方差分析的 3 个假定，采用 EXCEL 进行方差分析，得到如下结果：

方差分析：单因素方差分析

SUMMARY

组	观测数	求和	平均	方差		
游泳	5	1530	306.0	210.5		
打篮球	5	1726	345.2	1318.7		
骑自行车	5	1201	240.2	2707.7		
方差分析						
差异源	SS	df	MS	F	P-value	F crit
组间	28152.13	2	14076.07	9.966768	0.002816	3.885294
组内	16947.60	12	1412.30			
总计	45099.73	14				

请回答下面几个问题（取显著性水平为 0.05）：

- （1）根据方差分析表，你所得的结论是什么？
- （2）系统误差包含于哪个平方和之中？它是否显著存在（请说明原因）？
- （3）随机误差的估计值是多少？
- （4）请检验游泳与打篮球在 30 分钟内消耗的热量是否有显著差异（请写出计算过程）。

（可能用到的分位数： $t_{0.975}(8) = 2.3060$ ， $t_{0.975}(9) = 2.2622$ ， $t_{0.975}(10) = 2.2281$ ，

$t_{0.975}(11) = 2.2010$ ， $t_{0.975}(12) = 2.1788$ ）

3. 对一组 (x_1, y) 建立回归模型，由 EXCEL 运行得如下结果（记为“模型 1”）：

回归统计	
Multiple R	0.711675
R Square	0.506481
Adjusted R Square	0.457129

标准误差 2.853606

观测值 12

方差分析

	df	SS	MS	F	Significance F
回归分析	1	83.56934	83.56934	10.26264	0.009435
残差	10	81.43066	8.143066		
总计	11	165			

现加入一个自变量 x_2 , 所得的结果如下 (记为“模型 2”):

回归统计

Multiple R 0.919621

R Square 0.845703

Adjusted R Square 0.811415

标准误差 1.681897

观测值 12

方差分析

	df	SS	MS	F	Significance F
回归分析	2	139.541	69.7705	24.66455	0.000223
残差	9	25.45899	2.828777		
总计	11	165			

	Coefficients	标准误差	t Stat	P-value
Intercept	5.345394	1.660538	3.237139	0.010205
X Variable 1	-1.28549	0.486287	-2.64348	0.026761
X Variable 2	3.01183	0.677089	4.448205	0.001604

请回答下面几个问题 (取显著性水平为 0.05)

- (1) 请求出 x_1 与 y 的 Pearson 线性相关系数。
- (2) 请分别写出回归模型“模型 1”和“模型 2”。它们是否显著 (请说明原因)?
- (3) 两个模型中, x_1 的系数估计值各为多少? 应分别怎样解释?

(4) 请对两个模型的效果进行比较（请说明原因）。

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