

2014 年华东师范大学 432 应用统计考研

真题



启航教育

2014 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.假设男子身高服从正态分布，根据调查，2009 年上海成年男子身高 68%的区间估计为 [167.32, 175.02]，据此推算，99.7%的上海成年男子的身高的区间估计是（ ）。

- A.[161.32, 177.02]
- B.[159.62, 182.72]
- C.[163.47, 178.87]
- D.[155.77, 186.57]

2.为探索收入与时间关系的趋势，下列图形中，最适合的是（ ）。

- A.雷达图
- B.相关图
- C.直方图
- D.茎叶图

3.BMI 指数（英文为 Body Mass Index，简称 BMI），是用体重公斤数除以身高米数平方得出的数字，是国际上常用的衡量人体胖瘦程度以及是否健康的一个标准。当我们需要比较及分析一个人的体重对于不同高度的人所带来的健康影响时，BMI 值是一个中立而可靠的指标。WTO 的正常值范围是 18.5-24.9，如果这些范围是 95%的区间估计，那么 BMI 的点估计是：

- A.18.5
- B.21.7
- C.24.9
- D.22.8

4.某银行在分析其不良贷款率（Y）时，分析出 4 个可能影响不良贷款因素，它们之间的相关系数如表 1 所示，如果建立回归方程，那么最合适的是（ ）。

- A.一元回归
- B.二元回归
- C.三元回归
- D.四元回归

表 1 不良贷款与 4 个因素之间的相关系数

	x1	x2	x3	x4	y
x1	1				
x2	0.678	1.000			
x3	0.848	0.586	1.000		
x4	0.780	0.472	0.746	1.000	

y	0.829	0.742	0.685	0.500	1.000
---	-------	-------	-------	-------	-------

5.交警部门发布报告称:被怀疑酒驾司机中,72%的司机被要求采用呼吸仪测量,36%的司机被要求采血液仪测量,18%的司机被要求既采用呼吸仪测量又采用血液仪测量,那么,一个被怀疑酒驾的司机,不用这两种仪器测量的比例是()

- A.0.5
- B.0.25
- C.0.2
- D.0.1

6.某大学研究生与本科生共有 20000 名,其中研究生占 40%,如果用分层抽样抽 100 名学生的随机样本,那么()

- A.每个研究生被抽到的概率大于每个本科生被抽到的概率
- B.每个研究生被抽到的概率小于每个本科生被抽到的概率
- C.每个研究生被抽到的概率等于每个本科生被抽到的概率
- D.每个研究生被抽到的概率是八十分之一

7.莎士比亚在戏剧里的用字长度的分布是右偏的,说明莎翁()

- A.更习惯用长的字
- B.更习惯用短的字
- C.用字的长度无规律
- D.不能说明任何问题

8.以下关于极差离散系数的说法错误的是()。

- A.极差离散系数等于极差除以均值
- B.极差离散系数越大的数据,方差也越大
- C.极差离散系数不是稳健的统计量
- D.极差离散系数未必等于标准差离散系数

9.已知一总体服从指数分布,其均值为 μ ,取样本 $x_1, \dots, x_n (n>1)$,得到样本均值 \bar{x} 。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 的充分统计量
- C.在 μ 所有估计中, \bar{x} 的均方误差最小
- D.当样本量趋于无穷时, \bar{x} 的极限是 μ

10.正态性检验有很多方法,其中有一种 W 检验(它是国家标准 GB4882-85 推荐使用的犯第二类错误最小的检验),这种检验适用的样本量最小值和最大值分别为()。

- A.35, 50
- B.50, 不限
- C.8, 50
- D.8, 35

11.一种药冷藏 2 年后的有效率为 25%,通过改良,希望有效率提高到原来的一倍,为了检验改良效果,随机确定 20 名自愿者进行试验,如果 20 人中至少有 9 人以上注射后有效果,

则认为改良是成功的, 若低于 9 人, 则认为改良是不成功的, 这个检验的两类错误之和 ()

- A. 小于 0.5
- B. 在 0.5 与 1 之间
- C. 等于 1
- D. 大于 1

12. 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 。当真实的均值为 $\bar{\mu} (\bar{\mu} < \mu_0)$ 时, 检验的势为 ()。

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

13. 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, 则根据以下关系式不成立的是 ()。

- A. $\frac{\hat{\beta}_1^2 l_{xx}}{\hat{\sigma}^2} \sim F(1, n-2)$
- B. $\frac{(n-2)r^2}{1-r^2} \sim F(1, n-2)$
- C. $S_R = r l_{xy} \sqrt{\frac{l_{yy}}{l_{xx}}}$
- D. $S_R = \frac{l_{xy}^2}{l_{yy}}$

14. 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 5 次试验, 现已求得每个水平下试验结果的样本标准差分别为 1、2、2、3, 若 $ST=180$, 则 F 值为 ()。

- A. 5.33
- B. 4
- C. 8
- D. 16

15. 一个打篮球的男生跟您说, 他投篮命中率为 80%, 如果您请他投 20 个球, 结果中了 8 个, 那么, 根据假设检验原理, 原假设 $H_0: p=0.8$, 该检验的 P 值为 ()。

- A. $C_{20}^8 (0.8)^8 (0.2)^{12}$
- B. $C_{20}^8 (0.4)^8 (0.6)^{12}$
- C. $\sum_{k=0}^8 C_{20}^k (0.8)^k (0.2)^{20-k}$
- D. $\sum_{k=0}^8 C_{20}^k (0.4)^k (0.6)^{20-k}$

16. 单因子方差分析要求各个水平具有等方差, 利用 SPSS 可以检验这个假设。下表是某个方差分析问题的部分输出结果:

Test of Homogeneity of Variances
销售数据

Levence Statisitc	df1	df2	Sig
.282	3	16	.838

根据这些输出, 您的结论是 ()

- A.各个水平具有等方差
- B.各个水平不具有等方差
- C.还要结合实际方差大小才能确定
- D.需要做多重比较才能确定

17.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, 通常会考虑著名的 Box-Cox 变换, 这种数据是 ()

- A.对数变换
- B.幂变换
- C.线性变换
- D.三角函数变换

18.箱线图 (box plot) 不仅可以反映出一组数据的分布特征, 还可以进行多组数据分布特征的比较, 其绘制过程中除了需要最大、最小、两个四分位数外, 还需要 ()

- A.平均数
- B.众数
- C.中位数
- D.方差

19.卡方检验应用非常广泛, 特别是处理分类数据检验问题。这一方法是英国统计学家 () 于 1900 年提出的。

- A.K.Pearson
- B.R.A.Fisher
- C.M.G.KEndall
- D.F.GAlton

20.多重比较中常使用的检验方法有两种 S 法和 T 法, 下列说法正确的是 ()

- A.样本重复数不等用 S 法
- B.样本重复数不等用 T 法
- C.样本重复数不等不能用 S 法和 T 法
- D.S 法和 T 法在各种场合都可以使用

21.检验一个总体是否服从 Poisson 分布, 可通过以下哪种检验方法实现? ()

- A.卡方检验
- B.正态概率纸
- C.列联表检验
- D.方差齐性检验

22.设 X_1, X_2, \dots, X_n 是来自正态总体 $N(\mu, \sigma^2)$ 的一个样本, 下列统计量中, 均方误差最小的是 ()

A. $\frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2$ B. $\frac{1}{n} \sum_{k=1}^n (X_k - \bar{X})^2$

C. $\frac{1}{n+1} \sum_{k=1}^n (X_k - \bar{X})^2$

D. $\frac{1}{n+2} \sum_{k=1}^n (X_k - \bar{X})^2$

23. 设 X_1, X_2, \dots, X_n 是来自正态总体 $N(\mu, \sigma^2)$ 的一个简单随机样本, σ^2 的极大似然估计

计为 $\delta^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X})^2$ 则 σ^2 的渐近分布是 ()

- A. $N(\sigma^2, 2\sigma^4)$ B. $N(\sigma^2, \sigma^4)$ C. $N\left(\sigma^2, \frac{2\sigma^4}{n}\right)$ D. $N\left(\sigma^2, \frac{\sigma^4}{n}\right)$

24. 时间序列长期趋势的测定的主要方法有线性模型法、移动平均法和 ()

- A. 趋势剔除法 B. 季节平均法 C. 循环波动法 D. 指数平滑法

25. 以下关于抽样分布的说法错误的是 ()。

- A. 抽样分布一般与样本量有关
 B. 常用的 t 分布、 χ^2 分布、F 分布都是基于正态总体得来的
 C. 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
 D. 可以通过随机模拟的方法获得抽样分布

26. 对两个正态总体方差的检 $H_0: \sigma_1^2 \leq \sigma_2^2$ 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, 您的结论是 ()

- A. 只有 (2) 正确
 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

27. 在 $\triangle ABC$ 的两边 AB、AC 上各任取一点 P、Q, 则四边形 PBCQ 的面积数学期望等于 $\triangle ABC$ 面积的 ()

- A. 2/3 B. 3/4 C. 4/5 D. 1/2

28. 均值 μ 的 95% 的置信区间是 (-0.3, 10.2), 而利用同样的样本计算得均值 μ 的 90% 的置信区间是 (0.5, 9.4), 则对假设检验问题 $H_0: \mu=0$ vs $H_A: \mu \neq 0$, 下列选项中, () 最有可能是该检验 P 值。

- A. 0.0648 B. 0.1296 C. 0.0162 D. 0.0324

29. CPI 是居民消费价格指数 (consumer price index) 的简称。居民消费价格指数, 是一个反映居民家庭一般所购买的消费商品和服务价格水平变动情况的宏观经济指标, 目前我国居民消费价格指数涵盖全国城乡居民生活消费的食品、烟酒及用品、衣着、家庭设备用品及维修服务、医疗保健和个人用品、交通和通信、娱乐教育文化用品及服务、居住等八大类, 包括

() 个基本分类的商品与服务价格.

- A.155 B.199 C.228 D.262

30.华东师范大学图书馆一个电梯上的标签标出: 限乘 16 人, 限重 2500 磅, 假设学生和教职员工的体重近似服从均值为 150 磅、方差为 1600 的正态分布, 那么, 在随机 16 人乘电梯的过程中, 超重的概率是 ()

- A.0.16 B.0.26 C.0.36 D.0.46

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 4 分)

- 1.简述线性回归模型的基本假设及其检验方法,
- 2.判断一组数据异常值有哪些方法?
- 3.简述试验设计的三个基本原则。
- 4.给出 t 分布的定义, 计算 t 的期望与方差, 并回答当自由度趋向无穷时极限分布是什么。

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

1.已知某种电视机的寿命服从指数分布密度: $f(t) = \frac{1}{\lambda} e^{-\frac{t}{\lambda}}$, 其中 $\lambda > 0$ 为电视机的平均寿命, 现在随机抽取 n 台进行寿命测试, 试验到第 r 台失效为止, 并且这些失效时间为 $t_1 \leq t_2 \leq \dots \leq t_r$, 其余 n-r 台直到试验结束还未失效, 请回答下列问题:

(1) 求电视机平均寿命 λ 的极大似然估计;

(2) 若 λ 的先验分布为: $\pi(\lambda) = \frac{b^a}{\Gamma(a)} \lambda^{-(a+1)} e^{-\frac{b}{\lambda}}$, 其中 $\lambda > 0, a > 0, b > 0$, 求电视机平均

寿命 λ 的贝叶斯估计。

2.某饮料生产企业研制出一种新型饮料。饮料的颜色共有四种, 分别为橘黄色、粉色、绿色和无色透明这四种饮料的营养含量、味道、价格、包装等可能影响销售量的因素全部相同。现从地理位置相似、经营规模相仿的五家超级市场上收集了前一时期该饮料的销售情况, 见表 2。

表 2 该饮料在五家超市的销售情况

超市	无色	粉色	橘黄色	绿色
1	26.5	31.2	27.9	30.8
2	28.7	28.3	25.1	29.6
3	25.1	30.8	28.5	32.4
4	29.1	27.9	24.2	31.7
5	27.2	29.6	26.5	32.8

为了分析材料的颜色是否对销售量产生影响, 利用 SPSS 中的分析功能, 运行后得到如

下结果:

Descriptives

销售量

N	Mean	Std.Deviation	Std.Error	95% Confidence Interval for Mean		Minimum	Maximum	N
				Lower Bound	Upper Bound			
1	5	28.3200	.90664	.40546	27.1943	29.4457	27.20	29.10
2	5	29.5600	1.46390	.65468	27.7423	31.3777	27.90	31.20
3	5	26.6400	2.41413	1.07963	23.6425	29.6375	24.20	30.50
4	5	31.4600	1.28763	.57585	29.8612	33.0588	29.60	32.80
Total	20	28.9950	2.33722	.52262	27.9011	30.0889	24.20	32.80

Test of Homogeneity of Variances

销售量

Levene Statistic	df1	df2	Sig.
818	3	16	.503

ANOVA

销售量

	Sum of Squares	df	Mean Squares	F	Sig.
Between Groups	61.986	3	20.662	7.908	.002
Within Groups	41.804	16	2.613		
Total	103.790	19			

Multiple Comparisons

		Mean Difference	Std.Error	Sig.	95% Confidence interval	
					Lower Bound	Upper Bound
1	2	-1.24000	1.02230	.243	-3.4072	.9272
	3	1.68000	1.02230	.120	-4.872	3.8472
	4	-3.14000	1.02230	.007	-5.3072	.9728
2	1	1.24000	1.02230	.243	.9272	3.4072
	3	2.92000	1.02230	.011	.7528	5.0872

	4	-1.90000	1.02230	.082	-4.0672	.2672
3	1	-1.68000	1.02230	.120	-3.8472	.4872
	2	-2.82000	1.02230	.011	-5.0872	.7528
	4	-4.82000	1.02230	.000	-6.9872	-2.6528
4	1	3.14000	1.02230	.007	.9278	5.3072
	2	1.90000	1.02230	.082	.2672	4.0672
	3	4.82000	1.02230	.000	2.6528	6.9872

The mean difference is significant at the 0.05 level.

请回答下列问题:

- (1) 上述输出结果是利用了 SPSS 中的什么分析功能?
- (2) 从上述输出结果看, 使用这种方法的各个条件是否满足? 为什么?
- (3) 饮料的颜色是否对销售量产生影响? 说说您的结论

3. 若随机变量 X 的密度函数: $f(x) = \frac{1}{C\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$, 其中 $x \geq 0$, $C > 0$, $\sigma > 0$, 则称 X 服从

截尾正态分布, 记标准正态密度函数 $\phi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}$ 、分布函数为 $\Phi(t)$, 请回答下列问题:

- (1) 求常数 C 的值 (用 μ 、 ϕ 、 Φ 表示);
- (2) 求随机变量 X 的期望 (用 μ 、 σ 、 ϕ 、 Φ 表示);
- (3) 求随机变量 X 的方差 (用 μ 、 σ 、 ϕ 、 Φ 表示)。

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