

# 中国科学技术大学

## 2012 年硕士学位研究生入学考试试题

(统计学)

所有试题答案写在答题纸上, 答案写在试卷上无效

□需使用计算器

### 一、单项选择题 (本题包括 1-10 题, 每小题 3 分, 共 30 分)

1. 假设一个袋子中有黑色、白色和红色三种颜色的球, 它们的比例为 3:4:3, 现每次有放回地从袋子随机摸出一个球, 记下被摸出球的颜色, 如此反复, 则白球比黑球先被摸出的概率为 ①.  
(A)  $3/7$  (B)  $4/7$  (C)  $4/10$  (D)  $3/10$
2. 设  $A, B$  表示两个随机事件, 若  $P(AB)=0$ , 则事件  $A, B$  ②.  
(A) 互斥 (B) 不同时发生 (C) 相互独立 (D) 以上都不对
3. 投掷一枚硬币  $n$  次, 若记其中正面和反面出现的次数分别为  $X$  和  $Y$ , 则  $X$  和  $Y$  的相关系数为 ③.  
(A) 0 (B) 0.5 (C) -1 (D) 1
4. 假设  $X_1, \dots, X_n$  为从正态总体  $N(\mu, \sigma^2)$  中抽取的简单随机样本, 其中  $\mu, \sigma^2$  均未知, 记样本均值和样本方差分别为  $\bar{X}$  和  $S^2$ , 则  $\frac{n(\bar{X}-\mu)^2}{S^2}$  的分布为 ④.  
(A) 自由度是  $n-1$  的  $t$  分布 (B) 自由度是  $n$  的  $t$  分布  
(C) 自由度是  $1, n$  的  $F$  分布 (D) 自由度是  $1, n-1$  的  $F$  分布
5. 假设  $X_1, \dots, X_n$  为从正态总体  $N(\mu, 1)$  中抽取的简单随机样本, 其中  $\mu$  未知, 如果要求  $\mu$  的 95% 置信区间的长度不超过 0.6, 则样本量  $n$  至少需要等于 ⑤.  
(A) 42 (B) 43 (C) 44 (D) 45
6. 若一个参数的估计量值为 2.4, 该估计量的标准差值为 0.2, 则该参数的一个约 95% 置信区间为 ⑥.  
(A)  $[2.008, 2.792]$  (B)  $[2.0, 2.8]$   
(C)  $[2.2, 2.6]$  (D)  $[2.071, 2.729]$

7. 某保险公司多年的统计资料表明, 在家庭财产保险中被盗索赔占 20%, 以  $X$  表示随机抽查 100 个家庭财产保险中因被盗向保险公司索赔的户数, 则可以以 95% 的把握得出其中至少有 ⑦ 户会因被盗而向银行索赔.  
 (A) 13      (B) 12      (C) 15      (D) 14
8. 若  $X_1, \dots, X_n$  为从总体  $X$  中抽取的简单随机样本, 则经验分布函数  $\hat{F}(x) = \{\text{在 } X_1, \dots, X_n \text{ 中不超过 } x \text{ 的个数}\} / n$  为总体分布  $F(x)$  的一个估计. 则对固定的  $x$ , 估计量  $\hat{F}(x)$  的方差为 ⑧.  
 (A)  $nF(x)(1-F(x))$       (B)  $F(x)(1-F(x))/n$   
 (C)  $F^2(x)$       (D)  $(1-F(x))^2$
9. 设  $X$  为连续型随机变量,  $g$  为一连续函数, 则  $g(X)$  也为连续型随机变量. 该说法 ⑨.  
 (A) 正确      (B) 错误
10. 设  $X, Y$  和  $Z$  都是均值为 0, 方差为 1 的随机变量, 若  $X$  和  $Y$  的相关系数为 0,  $Y$  和  $Z$  的相关系数为 1, 则  $X$  和  $Z$  的相关系数为 ⑩.  
 (A) 0      (B) 1      (C) 0.5      (D) 无法得到

## 二、简答题 (每题 8 分, 共 40 分)

1. 试写出概率的公理化定义.
2. 试解释随机变量的联合分布和边际分布之间的关系.
3. “正态随机变量的线性组合仍服从正态分布”这种说法是否正确? 为什么?
4. 试指出置信区间和假设检验之间的等价性.
5. 试比较矩估计方法和极大似然估计方法的特点.

## 三、计算分析题 (共 80 分)

1. 某地方银行对其信用卡制度进行审核并考虑收回部分信用卡. 过去, 大约有 5% 的信用卡持有者不履行债务并造成银行无法收回的坏账. 因此, 管理层认为某特定的信用卡持有者不履行债务的先验概率为 0.05. 该银行进一步发



现,最终履行债务的信用卡持有者有 0.20 的概率会拖欠一个或几个月后支付。当然,对不履行债务的信用卡持有者,拖欠一个或几个月支付的概率为 1. 则

(1) (8 分) 如果某个信用卡持有者已经拖欠了一个月的支付, 计算他将不履行债务的概率.

(2) (7 分) 如果某个信用卡持有者不履行债务的概率超过了 0.20, 银行就会收回他的信用卡。如果某信用卡持有者已经拖欠了一个月的支付, 则银行是否会收回他的信用卡? 为什么?

2. 假设从  $X$  和  $Y$  各有分布

$X$	-1	0	1
P	0.25	0.5	0.25

$Y$	0	1
P	0.5	0.5

且  $X$  和  $Y$  不相关. 试

(1) (7 分) 求  $X$  和  $Y$  的联合分布.

(2) (8 分) 求  $Var[X|Y=0]$  和  $E[X|Y=0]$ .

3. StreetInsider.com 报道了 2002 年一些著名公司的每股收益的数据. 在 2002 年之前, 财务分析家就预测了这些公司 2002 年的每股收益. 利用下面数据评论实际的和预测的每股收益的差异.

公司	实际每股 收益	预测每股 收益	公司	实际每股 收益	预测每股 收益
美国电报 电话公司	1.29	0.38	埃克森-美 孚	2.72	2.19
美国运通	2.01	2.31	通用电气	1.51	1.71
花旗银行	2.59	3.43	强生	2.28	2.18
可口可乐	1.60	1.78	麦当劳	0.77	1.55
杜邦	1.84	2.18	沃尔玛	1.81	1.74

试

(1) (7 分) 在显著性水平 0.05 下, 检验实际的和预测的每股平均收益之间是否存在差异, 你的结论是什么?

(2) (4 分) 两均值之差的点估计是多少? 分析家是低估了还是高估了每股的收益?

(3) (4 分) 给出两均值之差的 95% 置信区间, 并据此对 (1) 的检验问题作出结

论并解释。

4. 美国睡眠基金会用一个抽样调查来研究每晚睡眠时间是否与年龄有关 (Newsweek, 2004. 1. 19). 下面是一个 49 岁及以下的样本和一个 50 岁及以上的样本提供的周日夜晚的睡眠时间数据.

年龄	睡眠时间				总计
	小于 6 小时	6-6.9	7-7.9	8 及以上	
49 岁及以下	38	60	77	65	240
50 岁及以上	36	57	75	92	260

试

(1) (10 分) 使用独立性检验以确定周日夜晚的睡眠时间和年龄是否有关, 并给出原假设. 在显著性水平为 0.05 下, 能得到什么结论?

(2) (5 分) 根据(1)的结论分析周日夜晚睡眠时间小于 6 小时、6-6.9 小时、7-7.9 小时和 8 小时及以上的人们的比例的差异, 能得到什么结论? 给出一个可能的合理解释.

5. 假设  $X_1, \dots, X_n$  和  $Y_1, \dots, Y_m$  分别是来自正态总体  $N(a, \sigma^2)$  和  $N(b, k\sigma^2)$  的两组独立的简单样本, 其中  $k$  为一已知的正数,  $a, b$  和  $\sigma^2$  均为未知的参数. 试

(1) (10分) 求出  $a, b$  和  $\sigma^2$  的极大似然估计;

(2) (10分) 根据 (1) 构造假设  $a-b$  的一个置信水平为  $1-\alpha$  ( $0 < \alpha < 1$ ) 的置信区间.

附: 可能用到的上分位数

$$u_{0.05} = 1.645, \quad u_{0.025} = 1.96, \quad t_{0.05}(9) = 1.83, \quad t_{0.025}(9) = 2.26, \quad \chi_{0.05}^2(3) = 7.815,$$

$$\chi_{0.025}^2(3) = 9.348$$