

# 中山大学

## 二〇一一年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

考试时间: 1 月 16 日 下午

### 考生须知

全部答案一律写在答题纸上, 答在试题纸上的不计分! 请用蓝、黑色墨水笔或圆珠笔作答。答题要写清题号, 不必抄题。

一、单项选择题 (30 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, 请将所选正确答案对应的字母写在答题纸上, 并标明题号)

1. 为调查在中国的省会城市和 4 个直辖市的居民年收入, 需要从这些城市的居民中抽取一个样本, 你认为以下四种抽样方式哪一种会得到更有代表性的样本? ( )

A. 简单随机抽样  
B. 整群抽样  
C. 系统随机抽样  
D. 分层随机抽样

2. 根据有关专家的建议, 51 岁以下成年女性每日的铁摄入量应为 16 毫克; 为了解这一人群铁摄入量的情况, 有人抽取一个容量为 45 的样本并取得她们在 24 小时内的铁摄入量, 所得数据列在下表:

15.0	18.1	14.4	14.6	10.9	18.1	18.2	18.3	15.0
16.0	12.6	16.6	20.7	19.8	11.6	12.8	15.6	11.0
15.3	9.4	19.5	18.3	14.5	16.6	11.5	16.4	12.5
14.6	11.9	12.5	18.6	13.1	12.1	10.7	17.3	12.4
17.0	6.3	16.8	12.5	16.3	14.7	12.7	16.3	11.5

以下图表, 哪一种可以让我们迅速得到以上样本数据的中位数与两个四分位数? ( )

A. 直方图  
B. 饼图  
C. 条形图  
D. 茎叶图

3. 在第 3 小题中, 经过计算可知, 样本平均值为 14.680, 样本标准差为 3.083; 假定每日的铁摄入量近似服从正态分布, 根据以上数据, 在 5% 水平下可以认为 51 岁以下成年女性的铁摄入量 ( )

A. 跟建议值没有显著差别  
B. 显著偏低  
C. 显著偏高  
D. 无法判断

4. 美国汽车制造商协会想了解消费者购车时的颜色偏好趋势, 抽取新近售出的 40 辆车并记录其颜色种类 (黑、白、红、绿、棕) 和深浅类型 (亮色、偏淡、中等、偏浓); 你认为以下展示数据的图表中, 哪一种不适合用来处理这一样本数据? ( )

A. 散点图  
B. 饼图  
C. 条形图  
D. 频数图

5. 现有一份样本, 为 100 名中学生的 IQ 分数, 由此计算得到以下统计量: 样本平均(mean) = 95, 中位数(median) = 100, 下四分位数(lower quartile) = 70, 上四分位数(upper quartile) = 120, 众数(mode) = 75, 标准差(standard deviation) = 30, 则关于这 100 名中学生, 下面哪一项陈述正确? ( )

A. 有一半学生分数小于 95  
B. 有 25% 的学生分数小于 70  
C. 中间一半学生分数介于 100 到 120 之间  
D. 出现频次最高的分数是 95

6. 若总体服从均值为  $\mu$ , 标准差为  $\sigma$  的正态分布; 从中抽出一个容量为 10 简单随机样本, 则样本平均的抽样分布为 ( )

A.  $N(\mu, \sigma^2/10)$   
B.  $N(\mu/10, \sigma^2)$   
C.  $N(\mu/10, \sigma^2/100)$   
D.  $N(\mu, \sigma^2/100)$

7. 关于方差分析, 下列说法正确的是 ( )

A. 方差分析的目的是分析各组总体方差是否相同  
B. 方差分析的组间均方仅仅衡量了随机误差的变异大小  
C. 各组数据呈严重偏态时, 也可以作方差分析  
D. 方差分析的目的是分析各组总体的均值是否相同

8. 以下统计方法中, 哪一种不能用来研究变量之间的关系? ( )

A. 样本比例估计  
B. 列联表分析  
C. 一元线性回归  
D. 多元线性回归

9. 如果把  $\alpha$  从 5% 降到 2.5%, 则置信程度为  $1-\alpha$  的样本平均的置信区间的宽度将 ( )

A. 增加  
B. 不变  
C. 降低  
D. 可能增加, 也可能降低

10. 在假设检验中, 如果检验结果是拒绝零假设; 那么, 以下哪一种情形的检验结果更显著? ( )

A. 样本平均值更小  
B. P-值更小  
C. 样本平均值更大  
D. P-值更大

11. 1998 年的一次网络民意调查中, 共 7553 人接受调查, 其中 4381 人认为: “白水事件”调查执行官 Kenneth Starr 应该询问所有可以找到的证人, 包括时任总统克林顿的助手; 由此, 你能得出以下哪一个结论? ( )
- A. 以上数据提供了充分证据, 说明多数人认为 Kenneth Starr 应该询问所有可以找到的证人, 包括时任总统克林顿的助手
- B. 以上数据未提供充分证据, 说明多数人认为 Kenneth Starr 应该询问所有可以找到的证人, 包括时任总统克林顿的助手
- C. 以上数据提供了充分证据, 说明多数人认为 Kenneth Starr 不应询问所有可以找到的证人, 包括时任总统克林顿的助手
- D. 以上数据未提供充分证据, 说明多数人认为 Kenneth Starr 不应询问所有可以找到的证人, 包括时任总统克林顿的助手
12. 一元线性回归中, 以下哪一种残差图特点可以说明回归模型的运用是不合理的? ( )
- A. 残差落在一直水平直线附近
- B. 残差落在一直倾斜直线附近
- C. 残差的正态概率图大致落在一直直线附近
- D. 残差关于一直水平直线大致对称
13. 考虑总体均值的 95.44% 置信区间, 已知总体服从正态分布且标准差为 10; 要使得到的置信区间的半径不超过 1, 需要的最小样本容量为 ( )
- A. 100
- B. 400
- C. 900
- D. 1600
14. 官方数据显示, 2008 年, 北京地区移动电话用户的月均电话费为 50 元; 从 2009 年 3 月份的用户费用清单中, 随机抽取 40 人, 发现他们的月均话费为 60 元; 移动电话公司想通过假设检验来判断 “2009 年的月均话费是否显著提高”, 设定零假设为 “ $H_0$ : 2009 年和 2008 年的月均话费没有显著差别”, 那么, 备择假设应该选取为 ( )
- A.  $H_1$ : 跟 2008 年相比, 2009 年的月均话费有显著减少
- B.  $H_1$ : 跟 2008 年相比, 2009 年的月均话费没有显著减少
- C.  $H_1$ : 跟 2008 年相比, 2009 年的月均话费有显著增加
- D.  $H_1$ : 跟 2008 年相比, 2009 年的月均话费没有显著增加
15.  $X_1, X_2, \dots, X_n$  为独立同分布的随机样本, 设统计量  $T(X_1, X_2, \dots, X_n)$  为  $\mu = E(X)$  的无偏估计量, 下面哪项指标小, 表示用该统计量估计均值  $\mu$  的可靠性好? ( )
- A.  $\text{var}[T(X_1, X_2, \dots, X_n)]$
- B.  $E[T(X_1, X_2, \dots, X_n)]$
- C.  $\text{var}(X)$
- D.  $\max\{X_1, X_2, \dots, X_n\} - \min\{X_1, X_2, \dots, X_n\}$
16. 考虑关于两个正态总体的均值  $\mu_1$  与  $\mu_2$  的假设检验, 经常用到的零假设为 ( )
- A.  $H_0: \mu_1 + \mu_2 = 0$
- B.  $H_0: \mu_1 - \mu_2 = 0$
- C.  $H_0: \mu_1 \times \mu_2 = 0$
- D.  $H_0: \mu_1 + \mu_2 = 0$

17. 设  $\{X_t\}$  是平稳时间序列, 则下面陈述不正确的是 ( )
- A.  $t$  时刻的均值  $E(X_t)$  不依赖  $t$
- B.  $t$  时刻的方差  $\text{var}(X_t)$  不依赖  $t$
- C.  $t$  时刻与  $s$  ( $s \neq t$ ) 时刻的协方差  $\text{cov}(X_t, X_s)$  不依赖  $t$ , 也不依赖  $s$
- D.  $t$  时刻与  $s$  时刻的协方差与  $t+1, s+1$  时刻的协方差相等, 即  $\text{cov}(X_t, X_s) = \text{cov}(X_{t+1}, X_{s+1})$
18. 假设检验中, 若零假设是简单的, 则显著性水平  $\alpha$  指 ( )
- A. 犯第一类错误的概率
- B. 犯第二类错误的概率
- C. 置信水平
- D. P-值
19. 在回归变量  $Y$  关于预测变量  $X$  的回归分析中, 若以  $x$  为横坐标,  $y$  为纵坐标, 绘散点图, 最小二乘原则是指 ( )
- A. 各点到直线的垂直距离的和最小
- B. 各点到  $x$  轴的纵向距离的平方和最小
- C. 各点到直线的垂直距离的平方和最小
- D. 各点到直线的纵向距离的平方和最小
20. 欲调查两变量 ( $X$  和  $Y$ ) 的相互关系, 收集一份数据作线性相关分析, 经计算得到样本相关系数  $r=0.38$ , 可以说 ( )
- A.  $X$  和  $Y$  无关, 因  $r$  值较小
- B. 不能确定  $X$  和  $Y$  是否相关以及相关密切程度, 因不知  $n$  的大小
- C. 虽然  $X$  和  $Y$  相关, 但不能认为  $X$  和  $Y$  有因果关系
- D. 因  $r>0$ , 可以认为  $X$  和  $Y$  存在线性相关关系
21. 假设检验难免犯错误, 以下论断中不正确的是 ( )
- A. 第一类错误是指零假设成立时, 拒绝零假设
- B. 第二类错误是指零假设不成立时, 接受零假设
- C. 其他条件不变, 显著性水平  $\alpha$  定得越高, 犯第一类错误的机会越小
- D. 其他条件不变, 显著性水平  $\alpha$  定得越高, 犯第二类错误的机会越小
22. 以回归方程  $Y=a+bX$  作相关分析与回归分析中, 关于样本相关系数  $r$  与回归系数  $b$ , 下列各项中哪项是正确的? ( )
- A.  $r>0$  时,  $b<0$
- B.  $r>0$  时,  $b>0$
- C.  $|r|=1$  时,  $|b|=1$
- D.  $|r|=1$  时,  $|b|=0$
23.  $2 \times 2$  的析因试验的方差分析中, 若关于 A、B 两个因素的主效应分别做假设检验, 结论均为拒绝零假设; 并且, 关于 A、B 两个因素的交互效应的假设检验结论为接受零假设, 则可认为 ( )
- A. 因素 A 的效应依赖于因素 B 的水平
- B. 因素 A 的效应不依赖于因素 B 的水平
- C. 因素 A 的主效应与因素 B 的主效应相等
- D. 两个因素可合二为一, 采用单因素的方差分析

24. 根据某地 6 至 16 岁学生近视情况的调查资料, 反映患者的年龄分布可用 ( )

- A. 线图
- B. 散点图
- C. 直方图
- D. 条形图

25. 设  $X_1, X_2, \dots, X_n$  是来自正态总体  $N(\mu, \sigma^2)$  的一个随机样本,  $\mu$  已知, 定义三个统计量如下:

$$S_1^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2, S_2^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, S_3^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$$

那么, 下面论断中, 哪一项成立? ( )

- A.  $S_1^2$  是  $\sigma^2$  的无偏估计
- B.  $S_2$  是  $\sigma$  的无偏估计
- C.  $S_3$  是  $\sigma$  的无偏估计
- D.  $S_3^2$  是  $\sigma^2$  的无偏估计

26. 随机事件  $A, B, C$  中恰有两个事件发生的复合事件为 ( ) (这里,  $A \cap B$  表示事件  $A$  与  $B$  都发生)

- A.  $(A \cap B) \cup (A \cap C) \cup (B \cap C)$
- B.  $\overline{A \cup B \cup C}$
- C.  $(A \cap B \cap \bar{C}) \cup (A \cap \bar{B} \cap C) \cup (\bar{A} \cap B \cap C)$
- D.  $(A \cap \bar{B} \cap \bar{C}) \cup (\bar{A} \cap \bar{B} \cap C) \cup (\bar{A} \cap B \cap \bar{C})$

27. 已知  $P(A) = 0.4, P(B) = 0.25, P(A - B) = 0.25$ , 则  $P(A \cup B) = ( )$ .

- A. 0.4
- B. 0.5
- C. 0.6
- D. 0.65

28. 某种动物活到 25 岁以上的概率为 0.8, 活到 30 岁的概率为 0.4, 则现年 25 岁的这种动物活到 30 岁以上的条件概率是 ( )。

- A. 0.76
- B. 0.5
- C. 0.4
- D. 0.32

29. 盒中有 5 个球, 其中 3 个红球, 2 个黑球。从中不放回的取, 每次一个, 则取两次, 每次都取到红球的概率为 ( )。

- A.  $\frac{3}{10}$
- B.  $\frac{6}{25}$
- C.  $\frac{2}{5}$
- D.  $\frac{1}{5}$

30. 投资某项目的收益率  $R$  是随机变量, 其分布见下表: 某位投资者在该项目上投资 1000 元, 他的预期收入和收入的方差分别为

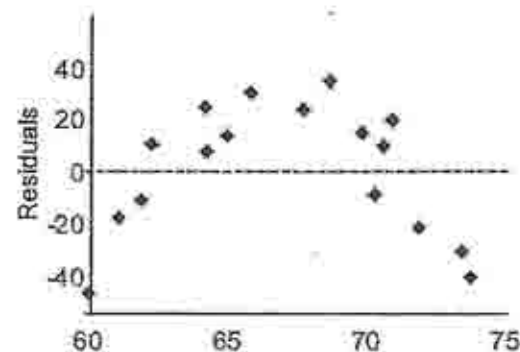
( ) 元和 ( ) (元)<sup>2</sup>

$R$ (取值)	4%	5%	6%	7%
$p_i$ (取值概率)	0.3	0.5	0.1	0.1

- A. 50, 10
- B. 1050, 10
- C. 1050, 80
- D. 50, 80

二、简要回答下列问题 (本题包括 1—4 题共 4 个小题, 每题 10 分, 共 40 分)

1. 下面两个统计图分别是对某数据集中  $y$  关于  $x$  的线性回归分析后的残差 (Residual)。请指出这个回归分析所存在的问题, 并提出解决方案



- 2. 考虑总体参数  $\theta$  的估计量, 简述无偏估计量与最小方差无偏估计量的定义
- 3. 欲调查广州市初中生的身高情况, 随机抽取 100 名广州市初中学生, 测量了身高。
  - 1) 用此例说明这几个统计概念, 总体 (population), 样本 (sample), 参数 (parameter), 统计量 (statistics)。
  - 2) 请说明如何对这 100 例身高数据进行描述性统计分析。
- 4. 在投掷一枚均匀硬币进行打赌时, 出现正面时投掷者赢 5 元, 出现反面时输 3 元, 记投掷者赢钱数为  $X$ , 试写出此问题的样本空间  $\Omega$ , 以及随机变量  $X$  的定义和概率分布。

三、计算与分析题 (本题包括 1—3 题共 3 个小题, 前两题每题 20 分, 第三小题 10 分, 共 50 分)

1. 一油漆制造商宣称, 他们生产的一种新型乳胶漆的平均干燥时间为 120 分钟; 为检验

这一数值是否属实，从该种乳胶漆中随机抽出 20 罐做试验，发现它们的干燥时间（分钟）为：123, 109, 115, 121, 130, 127, 106, 120, 116, 136, 131, 128, 139, 110, 133, 122, 133, 119, 135, 109；

- 1) 假定干燥时间近似服从正态分布，在 5% 的显著性水平下，检验以上数据是否提供充分证据说明这种乳胶漆的平均干燥时间大于制造商宣称的 120 分钟，要求给出零假设、备择假设、检验统计量、检验结果。（经计算可知，样本平均值为 123.1，样本标准差为 10.0，自由度为 19 的 t 分布的 0.05 上侧分位数为 1.729）
- 2) 给出平均干燥时间  $\mu$  的 10% 置信区间（要求给出枢轴量、置信区间的最后结果）
- 3) 简述上述假设检验问题和置信区间问题的主要联系？

2. 欲调查噪声强度对学生完成作业的准确度的影响，随机抽取了 15 名学生，分配到低噪声组，中噪声组，高噪声组中，得到准确度均数与方差如下表：

	低噪声组	中噪声组	高噪声组
样本量 (n)	4	5	6
均数 ( $\bar{x}$ )	60	66	45
方差 ( $s^2$ )	50	48	45

$$\text{样本平均 } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \text{ 样本方差 } s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left( \sum_{i=1}^n x_i^2 - n\bar{x}^2 \right)$$

附：分子自由度为 df1，分母自由度为 df2 的  $\alpha=0.05$  对应的 F 界值表

df1 \ df2	11	12	13	14
1	4.84	4.75	4.67	4.60
2	3.98	3.88	3.81	3.74
3	3.59	3.49	3.41	3.34

- 1) 计算完成方差分析表，并说明噪声强度能解释作业准确度的变异的比列。
  - 2) 用  $\alpha=0.05$  的显著性水平，分析噪声强度对作业准确度是否有影响。
  - 3) 要具体确定低噪声组与中噪声组间的准确度是否有差异，怎样分析(不需计算)
3. 某公司的一种设备从甲、乙、丙三个厂进货。甲、乙、丙三厂的进货比例分别占 30%，20% 和 50%。甲、乙、丙三厂生产的设备的次品率分别为 3%，3% 和 2%。
- 1) 现从这批进货中任取一件，求取到的设备是次品的概率；
  - 2) 经检验发现取到的设备是次品，求该设备是丙厂生产的概率。