安徽师范大学

2018 年硕士研究生招生考试初试试题

科目代码： 432

科目名称： 统计学

注意：可使用不带有存储功能的计算器。

一、单项选择题（本题包括 1—25 题共 25 个小题，每小题 2 分，共 50 分。在每小题给出的四个选项中，只有一个符合题目要求，把所选项前的字母填在答题纸相应的序号内）。

1. 为了对某产品的生产流程进行质量监控，每隔 5 分钟从生产线上抽取一件产品进行检测，这种抽样方法属于【　】。

   A. 简单随机抽样　　B. 等距抽样　　C. 分层抽样　　D. 整群抽样

2. 某大学研究生与本科生共有 20000 名，其中研究生占 40%。如果用分层比例抽样抽取 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 1/80

3. 以下关于数据类型的说法中错误的是【　】。

   A. 性别属于定类类型的数据　　B. 评级（如：优、良、差）属于定序类型的数据

   C. 绝对温度属于定距类型的数据　　D. 考试成绩（百分制）属于定比类型的数据

4. 下列关于定序数据（顺序数据）的描述，错误的是【　】。

   A．可进行＝或≠、＞或＜的运算

   B．在频数频率分布表中应按顺序列出其值，并给出累积频率

   C．适用于作柱状图、饼图和环形图

   D．适用于计算众数、中位数和四分位数

5. 下列统计量中，不能用于描述数据差异情况的统计量是【　】。

   A. 极差　　B. 标准差　　C. 方差　　D. 众数

6. 下列统计量中不易受到极端值影响的是【　】。

   A. 标准差　　B. 极差　　C. 中位数　　D. 均值

7. 对数据进行标准化变换是一种重要的数据预处理方法，其计算公式是用某一原始数据减去这组数据的【　】，再除以这组数据的标准差。

   A. 均值　　B. 中位数　　C. 众数　　D. 标准差

8. 在多元线性回归分析中，如果 F 检验表明线性关系显著，则意味着

   A. 在多个自变量中至少有一个自变量与因变量之间的线性关系显著

   B. 所有的自变量与因变量之间的线性关系都显著

   C. 在多个自变量中至少有一个自变量与因变量之间的线性关系不显著

   D. 所有的自变量与因变量之间的线性关系都不显著

9. 下列说法中错误的是【　】。

A. 样本方差反映了样本数据与样本平均值的偏离程度

B. 建立一个回归模型时若发现两个自变量不显著，应将二者同时剔除然后重新建立回归模型

C. 在回归直线方程 $y=0.1x+10$ 中，当解释变量 x 每增加一个单位时，预报变量 y 增加 0.1 个单位

D. 在回归分析模型中，残差平方和越小，说明模型的拟合效果越好

10. 在单因素方差分析中，要求各个水平具有等方差（方差齐性），假设进行方差齐性检验的 P 值是 0.838，则结论是【　】。

A. 各个水平具有等方差性　　B. 各个水平不具有等方差性

C. 还要结合实际方差大小才能确定　　D. 需要进一步做多重比较才能确定

11. 设 A_1, A_2, A_3 是三个事件，则事件 $A_1 \cup A_2 \cup A_3$ 表示【　】。

A. A_1, A_2, A_3 都发生　　B. A_1, A_2, A_3 都不发生

C. A_1, A_2, A_3 至少有一个发生　　D. A_1, A_2, A_3 不多于一个发生

12. 设 $f(x)$ 和 $g(x)$ 是两个概率密度，则下列函数中一定是概率密度的是【　】。

A. $f(x)\cdot g(x)$　　B. $f(x)+g(x)$　　C. $f(x)-g(x)$　　D. $0.5f(x)+0.5g(x)$

13. 设随机变量 X 的分布列（分布律）为：$P(X=k)=\dfrac{c}{2^k k!}$，$k=0,1,2,\cdots$，则常数 $c=$【　】。

A. e^2　　B. e^{-2}　　C. $e^{0.5}$　　D. $e^{-0.5}$

14. 设随机变量 X 和 Y 独立同分布，且 $P(X=1)=P(X=-1)=1/2$，则下列结论中正确的是【　】。

A. $P(X=Y)=1/2$　　B. $P(X=Y)=1$　　C. $P(XY=1)=1/4$　　D. $P(XY=1)=1$

15. 设某选手投篮的命中率为 0.9，他独立地投 10 次篮，则命中 9 次的概率为【　】。

A. $C_{10}^9\times 0.9\times 0.1^9$　　B. $C_{10}^1\times 0.9^9\times 0.1$　　C. $C_{10}^9\times 0.9^9$　　D. $C_{10}^1\times 0.1^9$

16. 设随机变量 $X\sim N(1,1)$，$Y\sim N(-1,6)$，且 X 和 Y 独立，则 $D(Y-2X)=$【　】。

A. 4　　B. 6　　C. 8　　D. 10

17. 设随机变量 X 和 Y 不相关，则下列式子中错误的是【　】。

A. $Cov(X,Y)=0$　　B. $E(XY)=E(X)\cdot E(Y)$

C. $D(X+Y)=D(X)+D(Y)$　　D. $D(XY)=D(X)\cdot D(Y)$

18. 设总体 $X\sim N(\mu,\sigma^2)$，其中 $\sigma^2>0$ 已知，μ 未知，$(X_1,X_2,\cdots,X_n)$ 是来自 X 的样本，则下列不是统计量的是【　】。

A. $\dfrac{1}{n}\sum_{i=1}^{n}X_i$　　B. $\sum_{i=1}^{n}X_i/\sigma^2$　　C. $\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$　　D. $\max\{X_1,X_2,\cdots,X_n\}$

19. 已知 $x_1,x_2,\cdots,x_n$ 的样本方差为 10；设 $y_i=2x_i-4,\ i=1,2,\cdots n$，则 $y_1,y_2,\cdots,y_n$ 的样本方差为【　】。

A. 10　　B. 20　　C. 40　　D. 80

20. 设 $X_1,X_2,\cdots,X_6$ 是来自于总体 $X\sim N(0,1)$ 的样本，则统计量 $\dfrac{X_1^2+X_3^2+X_5^2}{X_2^2+X_4^2+X_6^2}$ 服从【　】。

A. $\chi^2(6)$　　B. $t(3)$　　C. $N(0,6)$　　D. $F(3,3)$

21. 假设男子身高（单位：cm）服从正态分布。根据调查，2010 年上海成年男子身高的 68.26%的置信区间是[167.32, 175.02]。据此推算，该年上海成年男子身高的 99.74%的置信区间是【　】。

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

22. 设总体 $X\sim N(\mu,\sigma^2)$，其中 σ^2 已知，μ 未知，$x_1,x_2,\cdots,x_n$ 是来自总体 X 的样本观测值，已知 μ 的置信水平为 95%的置信区间为（4.35，5.54），则取显著性水平 $\alpha=0.05$ 时，检验假设

$H_0: \mu = 5.0$，$H_1: \mu \neq 5.0$的结果是【　　】。

A. 接受H_0　　B. 拒绝H_0　　C. 既不接受H_0，也不拒绝H_0　　D. 条件不足无法检验

23. 某校大二学生“高等数学”考试的平均成绩是75分，方差是144。从该校大二学生中随机抽取100名同学作为样本，则样本均值的期望和抽样分布的标准误差分别是【　　】。

A. 75, 12　　B. 75, 1.2　　C. 7.5, 12　　D. 7.5, 1.2

24. 已知总体$X \sim N(\mu, \sigma^2)$，设$(X_1, X_2, \cdots, X_n)$是来自X的样本，$n \geq 2$，$\overline{x}$和s^2分别是样本均值和样本方差，则下列说法中错误的是【　　】。

A. μ的矩估计和最大似然估计都是$\overline{x}$　　B. $\overline{x}$是μ的无偏估计

C. σ^2的矩估计和最大似然估计都是s^2　　B. s^2是σ^2的无偏估计

25. 在多元线性回归分析中，t检验是用来检验【　　】。

A. 总体线性关系的显著性　　B. 各回归系数的显著性

C. 样本线性关系的显著性　　D. 误差项的正态性

二、判断题（本题包括26—30题共5个小题，每小题2分，共10分。若论断正确，在题后的括号内打√；若论断不正确，在题后的括号内打×）。

26. 若事件A的概率为0，则A一定是不可能事件。【　　】

27. 设随机变量$T \sim t(n)$，则$T^2 \sim F(1, n)$。【　　】

28. 在假设检验中，若零假设正确，但检验的结果拒绝了零假设，这犯的是第二类错误。【　　】

29. 检验的P值反映的是数据与零假设的相容程度。P值越小，拒绝零假设的证据越强。【　　】

30. 相关系数的值越小，表示两变量之间的线性相关程度就越小。【　　】

三、简要回答下列问题（本题包括31—34题共4个小题，每小题10分，共40分）。

31. 什么是小概率事件原理？它是如何应用于假设检验问题的？

32. 重复测量某物体的重量，得到数据$x_1, x_2, \cdots, x_{20}$，试问采用$\frac{1}{10}\sum_{i=1}^{10} x_i$和$\frac{1}{20}\sum_{i=1}^{20} x_i$中的哪一个作为物体重量的估计比较好？为什么？

33. 区间估计中的置信度$1-\alpha$应如何理解？置信区间与假设检验中的哪个量有着密切的关系？

34. 在统计数据收集过程中，可能存在哪些误差，产生这些误差的原因是什么？

四、计算与分析题（本题包括35—38题共4个小题，第35小题和第36小题每题14分，第37小题10分，第38小题12分，共50分）

35. 玻璃环打包后成箱出售，每箱20只。假设每箱含0，1，2只残次品的概率分别为0.8，0.1和0.1。一个顾客欲买下一箱玻璃杯，在购买时，售货员随机地取出一箱，然后顾客开箱随机地查看其中的4只，若无残次品，则买下该箱玻璃杯；否则退回。试求：

(1) 顾客买下该箱玻璃杯的概率；

(2) 在顾客买下的一箱玻璃杯中，确实没有残次品的概率。

36. 设某次考试考生的成绩服从正态分布，现从中随机地抽取36名考生的成绩，算得平均成绩为66.5分，标准差为15分。

(1) 试问在显著性水平$\alpha = 0.05$下，是否可以认为这次考试考生的平均成绩为70分？

(2) 试求全体考生平均成绩的置信度为95%的置信区间。

（已知临界值：$t_{0.025}(35) = 2.0301$，$t_{0.025}(36) = 2.0281$，$t_{0.05}(35) = 1.6896$，$t_{0.05}(36) = 1.6883$）

37. 设总体$X \sim U[\alpha,\beta]$，其中α,β是未知参数，且$\alpha<\beta$，试求α和β的矩估计。

38. 货车的行驶时间Y（单位：小时）与行驶距离X_1（单位：百公里）及运送货物的次数X_2有关，下表给出的资料是从某运输队收集来的：

Y	7.3	4.8	9.5	5.2	7.2	6.2	7.4	6.6	7.6	6.3
X_1	1	0.6	1.5	0.7	0.5	0.8	0.75	0.75	0.9	0.9
X_2	4	2	4	2	4	2	3	4	4	2

根据统计软件输出如下的结果：

	Coefficients	标准误差	t Stat	P-value	Lower 95%	Upper 95%
Intercept	2.214	0.802	2.762	0.028	0.318	4.110
X Variable 1	2.688	0.792	A	0.012	0.816	4.560
X Variable2	B	0.219	3.449	0.011	0.237	1.271

方差分析

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

试根据上述结果，回答下列问题：

(1) 将表中A、B、C、D四处的数据补充完整（结果保留3位小数）；（4分）

(2) 在显著性水平$\alpha=0.05$下，你认为各个回归系数是不是显著不等于0？回归的整体显著性有没有通过检验？为什么？（4分）

(3) 写出Y与X_1及X_2之间的线性回归方程，并解释各个回归系数的意义。（4分）