

医学统计学

练习题一

一、单项选择题
1. 在根据两个独立样本资料比较总体均数时，进行 t 检验的前提条件是（ ）

- A. 两总体均数不等
- B. 两总体均数相等
- C. 两总体方差不等
- D. 两总体方差相等

E. 以上都不对

2. 欲描述 20 个血清滴度资料的集中趋势，最好采用（ ）

- A. 均数
- B. 众数
- C. 中位数
- D. 几何均数与中位数
- E. 几何均数

3. 已知正常成年男子的血红蛋白量服从正态分布 $N(\mu, \sigma^2)$ 。今测定某地 200 名正常成年男子的血红蛋白量，得到样本均数 \bar{X} 及其标准误 $S_{\bar{X}}$ ，则估计该地 95% 的正常男子血红蛋白平均数在哪个范围？

A. $\mu \pm 1.96\sigma_{\bar{X}}$

B. $\bar{X} \pm 1.96\sigma_{\bar{X}}$

C. $\bar{X} \pm 2.58S_{\bar{X}}$

D. $\bar{X} \pm 1.96S_{\bar{X}}$

E. $\mu \pm 2.58S_{\bar{X}}$

4. 两样本均数比较的 t 检验中，结果为 $P < 0.05$ ，有统计意义。 P 愈小则（ ）

- A. 说明两样本均数差别愈大
- B. 说明两总体均数差别愈大
- C. 说明样本均数与总体均数差别愈大
- D. 愈有理由认为两样本均数不同
- E. 愈有理由认为两总体均数不同

5. 以下关于正态分布特征的描述，不正确的是（ ）

- A. 正态分布为左右对称分布
- B. 正态分布的两个参数， σ 决定曲线向左或向右移动， μ 决定曲线是“胖”还是“瘦”
- C. 标准正态分布是正态分布确定两个参数后的一个特例
- D. 正态分布曲线均在横轴上方，在 μ 处最高

- E. 正态分布曲线下面积分布有一定规律
6. 欲比较 3 组均数是否相同，某人直接用两组独立样本 t 检验反复多次作两两比较，发现两组间在 $\alpha = 0.05$ 的检验水准下差异有统计学意义。以下说法正确的是（ ）
- A. 增大了犯 I 类错误的概率
 - B. 增大了犯 II 类错误的概率
 - C. 减小了犯 I 类错误的概率
 - D. 减小了犯 II 类错误的概率
 - E. 其做法正确
7. 方差分析中，组间变异的来源是哪个（些）方面（ ）
- A. 处理因素
 - B. 随机误差
 - C. 处理因素和随机误差
 - D. 测量误差
 - E. 处理因素和测量误差
8. 随机选取男性 200 人，女性 100 人为某寄生虫病研究的调查对象，测得其感染阳性率分别为 20% 和 15%，则合并阳性率为（ ）
- A. 35%
 - B. 16.7%
 - C. 18.3%
 - D. 无法计算
 - E. 30%
9. 比较两个率时，不需要进行率的标准化的情况有（ ）
- A. 内部构成不同
 - B. 内部构成相同
 - C. 未知各年龄别人口数
 - D. 缺乏各年龄别死亡人数
 - E. 不易寻找标准人口构成
10. 四格表资料 χ^2 检验中，出现下列哪种情况需要进行校正（ ）
- A. $T > 5$ 或 $n < 40$
 - B. $T < 1$ 且 $n > 40$
 - C. $T < 1$ 或 $n < 40$
 - D. $1 < T < 5$ 且 $n > 40$
 - E. $T > 5$ 且 $n < 40$
11. 从甲乙两文中，查到同类研究中的两个率比较的四格表资料，其 χ^2 检验结果：甲文 $\chi^2 > \chi^2_{0.01(1)}$ ，乙文 $\chi^2 > \chi^2_{0.05(1)}$ 。可认为（ ）
- A. 两文结果有矛盾
 - B. 两文结果基本一致
 - C. 甲文结果基本可信
 - D. 甲文中总体率的差值较大
 - E. 乙文的样本量需进一步增加
12. 配对四格表 χ^2 检验的资料，误作一般四格表 χ^2 检验，则（ ）
- A. 本来差别有统计学意义，可能判为差别无统计学意义
 - B. 本来差别无统计学意义，可能判为差别有统计学意义
 - C. 可能加大 I 型错误
 - D. I 型错误和 II 型错误不变
 - E. 可能同时增大 I 型错误和 II 型错误

13. 两小样本比较的假设检验，首先应考虑（ ）
A. t 检验或秩和检验均可
B. 秩和检验
C. t 检验
D. 符合 t 检验还是秩和检验条件
E. 方差分析
14. 设配对资料的变量为 X_1 和 X_2 ，则 Wilcoxon 符号秩和检验编秩为（ ）
A. 把 X_1 和 X_2 分别从小到大编秩
B. 把 X_1 和 X_2 综合从小到大编秩
C. 把 X_1 和 X_2 综合按绝对值从小到大编秩
D. 把 X_1 和 X_2 的差数从小到大编秩
E. 把 X_1 和 X_2 的差数按绝对值从小到大编秩
15. 在配对设计差值的符号秩和检验中，绝对值等于 3 的差值有 -3、-3、3、3。如果它们的位次为 4、5、6、7，则第 2 个 -3 的秩次应为（ ）
A. 5
B. -5
C. 7
D. 5.5
E. 4
16. 根据样本计算出两个随机变量 X 与 Y 的相关系数 r ，经 t 检验， $P < 0.01$ ，可认为（ ）
A. X 与 Y 间密切相关
B. 总体相关系数 $\rho = 0$
C. 总体相关系数 $\rho \neq 0$
D. 总体相关系数 $\rho > 0$
E. 总体相关系数 ρ 很大
17. 在简单线性回归分析中，回归系数计算为 -0.15，经检验有统计学意义，说明（ ）
A. X 对 Y 的影响占 Y 变异的 15%
B. X 每增加一个单位，Y 平均减少 0.15 个单位
C. X 每增加一个单位，Y 平均减少 15%
D. Y 每增加一个单位，X 平均减少 15%
E. Y 每增加一个单位，X 平均减少 0.15 个单位
18. 下列回归方程中，错误的是（ ）
A. $y = 2 + 5x, r = 0.86$
B. $y = -7 + 5x, r = 0.56$
C. $y = -8 - 6x, r = 0.73$
D. $y = 10 - 2x, r = -0.96$
E. $y = 15 + 9x, r = 0.67$
19. 实验设计的基本原则是（ ）
A. 随机、配对、盲法、对照
B. 随机、盲法、配对、均衡
C. 随机、重复、对照、均衡

D. 重复、随机、配对、均衡
E. 随机、盲法、齐同、均衡

20. 某校大二年级学生共 500 人参加考试，将 500 份试卷编好号码后，从中随机抽取 65 份计算平均成绩，此种抽样方法为（ ）
A. 简单随机抽样
B. 系统抽样
C. 分层抽样
D. 整群抽样
E. 便利抽样

二、简答题
21. 请结合标准正态分布，阐述 t 分布的特征

22. 请简述直线回归方程的应用

23. 比较某省城乡老年女性骨质疏松症的患病率，由于农村老龄人口的构成高于城市，进行了患病率的标准化。请问，用以反映城乡老年女性骨质疏松症的患病水平，应该用未标准化前的粗患病率还是标准化患病率，为什么？进行率的标准化，在报告结束时，需要报告哪些内容？

24. 同一资料，出于同一研究目的，当参数检验和非参数检验所得结果不一致时，应以哪个结果为准？请简述理由

25. 在其他条件一致的情况下，分别从一个变异大的总体和一个变异小的总体里进行抽样，哪个抽样所需要的样本量较大？除了这个因素以外，还有其他哪些因素会影响样本量的大小？

26. 请简述本课程所学范围内 χ^2 检验的用途

三、综合分析题

27. 30 名头痛患者按照病情分为 10 个区组，每一区组内 3 名患者随机接受标准方法或两种新方法治疗，止痛时间如下表所示。试选用合适的方法说明新方法与标准方法相比止痛时间是否不同。

表 1 三种方法治疗头痛患者的止痛时间 (min)

区组	治疗方法		
	标准方法	新方法一	新方法二
1	8.40	6.90	6.80
2	7.70	6.80	6.40
3	10.10	10.30	10.60
4	9.60	9.40	9.20
5	9.30	8.00	7.60
6	9.10	8.80	8.90
7	9.00	6.10	6.40
8	7.70	7.40	7.30
9	8.10	8.00	8.20
10	5.30	5.10	5.20

表 2 方差分析表

变异来源	SS	df	MS	F	P
处理	3.8527	()	()	()	0.001
区组	55.8137	()	()	()	<0.001
组内	()	()			
总	65.3537	()			
问题：					

- (1) 请完成方差分析计算表
(2) 根据结果比较在三种治疗方法下的止痛时间是否不同？10 个区组间的患者平均止痛时间是否不同？
请根据上述资料写出详细的统计分析步骤。

28. 研究小鼠实施胰腺切除术后是否对血钠离子造成影响，某课题组开展了一项随机对照实验。

(1) 以下是由电脑产生的 0~1 之间服从均匀分布的一批随机数字：

0.862, 0.852, 0.131, 0.238, 0.240, 0.211, 0.018, 0.208, 0.914, 0.932, 0.224, 0.074

请利用以上随机数字将 10 只小鼠随机分成两组。

(2) 两组小鼠中，其中一组小鼠完成打开腹腔、切除胰腺，然后缝合的手术。另一组小鼠完成打开腹腔，未切除胰腺并且缝合的手术。比较两组小鼠血钠离子浓度。问这种对照形式属于何种类型的对照？

(3) 比较两种手术对小鼠血钠离子水平的影响，应该如何收集资料？

(4) 假设检验的方法应该用 t 检验还是秩和检验？请简述理由。

练习题二

一、最佳选择题

1. 为了解某市 18 岁以上居民高血压患病率，某研究者从中随机抽取 2000 名 18 岁以上的居民，计算其高血压患病率为 7.2%，一下说法正确的是（ ）
- A. 该研究的总体是 2000 名 18 岁以上的居民
 - B. 该研究的总体是该市所有的 18 岁以上的居民
 - C. 该研究的样本是该市所有高血压患者
 - D. 该研究的样本是 2000 名 18 岁以上高血压患者
 - E. 计算出来的 7.2% 是一个参数
2. 比较身高和体重两组数据变异程度大小宜采用（ ）
- A. 变异系数
 - B. 方差
 - C. 极差
 - D. 标准差
 - E. 四分位数间距
3. 在描述定量资料的集中趋势时，一下论述中错误的是（ ）
- A. 均数适宜于对称分布的资料
 - B. 几何均数和中位数都适宜于正偏态分布
 - C. 同一正偏态分布资料，估计的均数值小于中位数值
 - D. 对称分布资料的均数值等于中位数值
 - E. 几何均数特别适宜于细菌学和免疫学检测指标的描述
4. 横轴上，标准正态曲线下从 0 到 1.96 的面积为（ ）
- A. 95%
 - B. 45%
 - C. 97.5%
 - D. 47.5%
 - E. 49.5%
5. 关于假设检验的一下论述中，错误的是（ ）
- A. 在已知 A 药降血压疗效只会比 B 药好或相等时，可选单侧检验
 - B. 检验水准 α 定得越小，犯 I 型错误的概率越小
 - C. 检验效能 $1-\beta$ 定得越小，犯 II 型错误的概率越小
 - D. P 值越小，越有理由拒绝 H_0
 - E. 在其他条件相同时，双侧检验的检验效能比单侧检验低
6. 对总体参数作估计时，关于 95% 可信区间 (CI)，正确的说法是（ ）
- A. 大约有 95% 样本的 95% CI 覆盖了总体参数
 - B. 对于每一个 95% CI 而言，总体参数约有 95% 的可能落在其内
 - C. 各个样本的 95% CI 是相同的
 - D. 对于每一个 95% CI 而言，有 95% 的可能性覆盖总体参数
 - E. 以上说法都不对
7. 在完全随机设计的方差分析中，组间变异主要反映（ ）
- A. 抽样误差的作用
 - B. 随机误差的影响
 - C. 系统误差的影响
 - D. 处理因素的作用
 - E. 全部观察值的离散程度
8. 随机区组设计中要求（ ）
- A. 区组内个体差异越小越好，区组间差异越大越好
 - B. 区组内及区组间的差异都是越小越好

- C. 区组内没有个体差异，区组间差异越大越好
 D. 区组内个体差异越大越好，区组间差异越小越好
 E. 区组内没有个体差异，区组间差异越小越好

9. 下列哪一项不是表示相对数的统计指标()
 A. 构成比 B. 率 C. 相对比 D. 标准差 E. 优势比

10. 以下说法中错误的是()
 A. 频率可以用来描述二分类变量的分布
 B. 构成比表示某一事物内部各组成部分的比重或分布
 C. 如果资料的总例数过少，计算的相对数偏差会过大
 D. 某社区中抽取 100 名高血压患者，其中有 50% 的患者为老年人，由此说明老年人最易患高血压
 E. 定基比是指报告期指标与基线指标之比

11. 四格表资料检验中，出现下列哪种情况需要进行校正()
 A. $T > 5$ 或 $n < 40$ B. $T < 1$ 且 $n \geq 40$ C. $1 \leq T < 5$ 或 $n < 40$
 D. $1 \leq T < 5$ 且 $n \geq 4$ E. $T > 5$ 且 $n < 40$

12. 为比较新疗法与常规方法治疗某病的效果，研究者将 100 名患者按性别、年龄等特征配成对子，分别接受两疗法的治疗。对两组疗效的观察，28 对患者同时有效，5 对患者同时无效，11 对患者新药有效，而常规治疗无效。欲比较两种疗法有效率是否相同，应选择的统计分析方法为()

- A. 完全随机设计的两组二分类资料比较 χ^2 检验
 B. 完全随机设计的两组二分类资料比较校正 χ^2 检验
 C. 配对设计的二分类资料比较 χ^2 检验
 D. 配对设计的二分类资料比较校正 χ^2 检验
 E. Fisher 确切概率法

13. 符合 t 检验条件的数值变量资料如果采用秩和检验，不拒绝 H_0 时()

组别	无效	有效	显效	治愈	合计	
左西替利嗪片组	2	5	8	49	64	A. I 类错误增大
西普利嗪片组	3	9	10	44	66	B. II 类错误增大 C. I 类错误减小 D. II 类错误减小 E. 两类错误都增大

14. 利用盐酸左西替利嗪片治疗慢性特发性荨麻疹临床试验，以西替利嗪片组作为对照组，治疗 28 天后结果如下表，现要比较两种药物的疗效，应选用的假设检验方法是：()

- A. 行 \times 列表资料的 χ^2 检验 B. 两样本率比较的 χ^2 检验 C. 两组均数比较的 t 检验
 D. 两样本比较的秩和检验 E. Spearman 等级相关分析

15. 两组资料进行相关性分析，一个 $t_{0.01, v1} > tr1 > t_{0.05, v1}$ ，另一个 $tr2 > t_{0.01, v2}$ ，下列正确的是()

- A. 第 1 组资料两变量关系更密切 B. 第 2 组资料两变量关系更密切
 C. 很难说哪一组变量关系更密切 D. $tr1 > tr2$ E. 以上选项都错

16. 下列有关线性相关和线性回归分析的说法不正确的是()

- A. 回归分析要求资料服从双变量正态分布
 B. r 和 b 的假设检验等价

- C. r 和 b 的正负号一致
D. 决定系数的大小决定了相关的密切程度
E. 双变量正态分布时可以用回归解释相关
17. 双变量正态分布时可以用回归解释相关，可用()
A. 直条图 B. 线图 C. 直方图 D. 圆形图 E. 百分条图
18. 比较甲、乙、丙三地区某年度某种疾病的发病率情况，可用()
A. 直条图 B. 线图 C. 直方图 D. 圆形图 E. 百分条图
19. 抽样调查中，样本含量的估计以下说法错误的是()
A. 其他条件不变时，容许误差越大，所需样本量越大
B. 其他条件不变时，调查总体的标准差越大，所需样本量越大
C. 其他条件不变时，I类错误概率越小，所需样本量越大
D. 其他条件不变时，双侧检验样本量大于单侧检验
E. 保证同样的精度情况下，整群抽样所需样本量大于单纯随机抽样
20. 多组样本比较的 Kruskal-Wallis 检验中，当比较组数超过 3 组时，H 值近似服从什么分布()
A. t 分布 B. Z 分布 C. F 分布 D. χ^2 分布 E. H 分布
21. 为研究新药“匹伐他汀”疗效，在某医院选择 50 例高血脂病人，随机分成试验组和对照组，试验组用匹伐他汀治疗，对照组用公认有效的常规药治疗，这种对照属于()
A. 实验对照 B. 空白对照 C. 相互对照 D. 标准对照 E. 历史对照

二、简答题

- 医学统计学中变量的类型有哪两种？请举例说明这两种类型的变量
- 简述三组样本均数比的检验方法和检验步骤
- 请说明应用直接标准化法的目的、前提条件和步骤
- 简述线性回归分析的应用条件
- 举例说明调查研究中的系统抽样是如何实现的？

三、综合分析题

- 某肿瘤防治中心用甲乙两种血清学方法检查鼻咽癌患者血清 412 份，结果两法都是阳性的 261 份，都是阴性的 31 份；甲法阳性但乙法阴性的 110 份，甲法阴性但乙法阳性的 10 份。请比较两种方法检出率。

参考公式：

$$\chi^2 = \frac{(ad-bc)^2 * n}{(a+b)(c+d)(a+c)(b+d)} \quad \chi^2 = \frac{(b-c)^2}{b+c}$$

参考界值： $\chi^2_{0.05(1)}=3.84$, $\chi^2_{0.01(1)}=6.63$, $\chi^2_{0.05(4)}=9.49$, $\chi^2_{0.01(4)}=13.28$

- 为研究维生素 A 强化食用油对少年儿童的免疫功能有无改善作用，拟在食用油中强化维生素 A 后，对维生素 A 缺乏的小学生进行干预实验。问：

- 可采用何种设计方案？
- 如何设置对照？
- 选择观察指标时应注意些什么？
- 该试验可能有哪些混杂因素？应该如何控制？

练习题三

一、最佳选择题（15题，每题2分，共30分）（1~15题）

- 某研究用 x 表示儿童在一年中患感冒的次数，共收集了 1000 人，请问：儿童在一年中患感冒次数的资料属于：
A. 连续型资料 B. 离散型资料 C. 有序分类资料 D. 无序分类资料 E. 以上均不对
- 对于尿铅，适宜使用什么指标进行统计描述：
A. 算术均数与极差

B. 中位数与变异系数

C. 众数与标准差

D. 算数均数与标准差

E. 中位数与四分位数间距

3. 某研究者欲通过抽样调查获得某地区成年男性血压的平均水平，在样本量计算中，以下哪项不是要考虑的因素？

A. 估计的置信度

B. 该人群血压的变异情况

C. 可接受的估计精度

D. 抽样方式

E. 第二类错误

4. 秩和检验和 t 检验相比，其优点是

A. 计算简便，不受分布限制

B. 公式更为合理

C. 检验效能高

D. 抽样误差小

E. 需要样本量小

5. 均数的标准误反映了，

A. 个体变异程度的大小

B. 个体集中趋势的位置

C. 指标的分布特征

D. 频数的分布特征

E. 样本均数与总体均数的差异

6. 95% 参考值范围的正确理解是：

A. 在此范围之外者为异常

B. 此范围外者 95% 不正常

C. 尚有 5% 的正常人在此范围之外

D. 在此范围内者一定正常

E. 所有正常人在此范围内

7. 若分析肺活量和身高之间的数量关系，拟用身高值预测肺活量值，则采用：

A. 线性相关分析

B. 秩相关分析

C. 线性回归分析

D. 多元回归分析

E. 以上都不是

8. 计算标准化死亡率的目的是：

A. 减少死亡率估计的偏倚

B. 减少死亡率估计的抽样误差

C. 减少死亡率估计的随机误差

D. 减少死亡率估计的系统误差

E. 消除两组个体内部构成不同的影响

9. 两样本率比较的 χ^2 检验，差别有统计学意义时，P 值越小

A. 两样本率差别越大

B. 两总体率差别越大

- C. 越有理由认为两总体率不同
 D. 越有理由认为两样本率不同

E. 以上都不对

10. 如果线性相关系数 $r=1$, 则一定有:

- A. $SS_{\text{总}} = SS_{\text{差}}$

- B. $SS_{\text{固}} = SS_{\text{差}}$

- C. $SS_{\text{固}} = SS_{\text{总}}$

- D. $SS_{\text{总}} > SS_{\text{差}}$

E. 以上都不正确

11. 下列关于标准差与标准误的表述错误的是:

- A. 样本统计量所对应的标准差称为标准误

- B. 标准误反映抽样误差的大小, 其产生原因为总体中存在个体差异

- C. 抽样误差反映的是样本统计量与总体参数之间存在的差异

- D. 样本量越大, 标准差越小

- E. 样本量越大, 标准误越小

12. 关于标准正态分布曲线下的面积, 下列说法中错误的是:

- A. 大于 1.96 的曲线下面积占总面积的 2.5%

- B. 大于 1.645 的曲线下面积占总面积的 2.5%

- C. 从 1.96 到 2.58 间曲线下面积占总面积的 2%

- D. 从 -1.96 到 1.96 间曲线下面积是占总面积的 95%

- E. 从 -1.96 到 -1.645 间曲线下面积占总面积的 2.5%

13. 在简单线性回归分析(X 为自变量, Y 为因变量)中, 经检验模型有统计学意义。得到决定系数为 0.55, 其含意是:

- A. X 增加一个单位, Y 平均增加 55%

- B. X 增加一个单位, Y 平均增加 0.55 个单位

- C. Y 增加一个单位, X 平均增加 0.55 个单位

- D. Y 增加一个单位, X 平均增加 55%

- E. X 能够解释 Y 总变异的 55%

14. 某放射科医生收集脑外伤患者 60 例, 观察脑出血病灶的直径和病人昏迷程度(轻度、中度、重度), 欲分析昏迷程度与病灶大小的关系, 可进行

- A. Pearson 相关分析

- B. Spearman 等级相关分析

- C. 两样本比较的 t 检验

- D. 方差分析

- E. χ^2 检验

15. 估计样本含量的容许误差是指

- A. 系统误差

- B. 测量误差

- C. 样本统计量值之差

- D. 总体参数值之差

- E. 样本统计量值和所估计的总体参数值之差

二 是非题(每题 3 分, 共 30 分)(16~25 题)

16. 比较两种传染病随时间变化的速度在绘制图形的时候应该选择线图.

- A. 正确 B. 错误

17. 某次调查发现, 某市 40~49 岁年龄组高血压患者在该市所有高血压患者中所占的构成比最大, 所以 40~49 岁的人群最易患高血压.

A. 正确 B. 错误

18. 应做配对卡方检验却误用了两独立样本四格表资料的卡方检验会降低检验效能.

A. 正确 B. 错误

19. 某项研究选择抽样调查还是普查，是根据研究的经费是否充裕来决定的.

A. 正确 B. 错误

20. 进行统计分析时，如果资料满足参数检验的条件，这时对其使用非参数检验方法也可，二者的结果等价。

A. 正确 B. 错误

21. 如果样本不是通过随机抽样得来的，假设检验就不能推断总体.

A. 正确 B. 错误

22. 能进行线性回归的资料一定能进行线性相关分析.

A. 正确 B. 错

23. 回归系数没有度量衡单位.

A. 正确 B. 错误

24. 同质个体之间的变异称为个体变异.

A. 正确 B. 错误

25. 欲比较身高与体重变异程度的大小，可以采用变异系数.

A. 正确 B. 错误

三、分析题：

1. 为了比较两种治疗“焦虑症”精神药物的疗效，将 100 位符合纳入和排出标准的患者随机分为两组，分别服用两种药物。患者的焦虑水平用专门的量表进行测评。请问：服用两种药物，患者焦虑水平的具体内容分别是什么？(6 分)

1) 在该临床试验中，“实验设计三要素的具体内容分别是什么？(6 分)

2) 当两组个体的焦虑分值的改变量满足怎样的条件时，可以使用 t 检验进行疗效的比较？列出具体检验步

骤。此时，关于疗效差异较恰当的统计描述指标是什么？(7 分)

3) 如果疗效的记录结果不是焦虑分值的改变量，而是症状“加重”，“基本未变”，“有所缓解”、“明显改善”，则可以用什么样的方法比较两组疗效？列出具体检验步骤。此时，关于疗效差异较恰当的统计描述指标是什么？(7 分)

2. 抗癌新药的毒理研究中，欲分析不同剂量时病理效果，将 80 只大鼠按性别、窝别、体重、年龄等因素配成 40 对，每对大鼠随机分配分别接受甲剂量和乙剂量注射，结果两种剂量都死亡的有 30%，甲剂量的死亡率为 50%，乙剂量的死亡率为 70%。

1) 这是何种设计类型的资料？(4 分)

2) 两种剂量均生存的大鼠数是多少？请列出整理表。(8 分)

3) 请对这份资料作统计分析。(8 分)

练习题四

一、选择题

1. 比较两地乙肝疫苗接种后的血清学检查的阳转率有无区别应采用什么方法：

A. 用 t 检验

B. 用秩和检验

C. t 检验与秩和检验均可

D. 卡方检验

E. 看资料符合 t 检验还是秩和检验

2. 以下关于 χ^2 检验的自由度的说法正确的是：

B、对一个 3×4 表进行检验时， $v=11$

C、对四格表检验时， $v=4$

D、若 $\chi^2_{0.05,v} > \chi^2_{0.05,\eta}$ ，则 $v > \eta$

E、样本含量减一

3、 χ^2 的取值范围是：

A、 $-\infty < \chi^2 < +\infty$

B、 $-\infty < \chi^2 < 0$

C、 $0 < \chi^2 < +\infty$

D、 $-1 < \chi^2 < 1$

E、 $\chi^2 \leq 1$

4、以下哪个不是定性变量：

A、某药物的治疗效果（有效、无效）

B、接受免疫接种后抗体滴度水平（1:4, 1:8, 1:16……）

C、最高学历（小学、初中、高中、大学、大学以上）

D、某地区 120 名调查者是否患高血压

E、病人的病情分段

5、A 城市的高血压粗率为 15%，B 城市的高血压粗率为 10%，经过年龄别标准化后，A 城市的高血压直接标化率为 11%，B 城市的高血压直接标化率为 14%，请问以下说法正确的是：

A、能说明两城市实际高血压患病水平是高血压的直接标化率

B、直接标化率比粗率更准确

C、粗率和标化率的差别是因为 A 城市中大龄者比例低于 B 城市

D、粗率和标化率的差别是因为 A 城市中大龄者比例高于 B 城市

E、以上说法均不对

6、在线性相关分析中，已知相关系数 $r=1$ ，则一定有：

A、回归系数 $b=1$

B、回归系数 $b=0$

C、截距 $a=1$

D、 $SS_{\text{剩余}}=0$

E、 $SS_{\text{剩余}}=SS_{\text{总}}$

7、各观察值均加（或减）同一数后：

A、均值不变，标准差改变

B、均数改变，标准差不变

- C、两者均不变
- D、两者均改变有关

8、有关四分位数间距描述中不正确的是

- A、四分位数间距 $= P_{75} - P_{25}$
- B、四分位数间距比极差稳定
- C、四分位数间距即中间 50% 观测值的极差
- D、主要用于描述正态分布资料的变异度
- E、四分位数间距越大，表示变异越大

9、对含有两个随机变量的同一批资料，既作直线回归分析，又作直线相关分析。令对相关系数检验的 t 值为 t_1 ，对回归系数检验的 t 值为 t_2 ，二者之间有什么关系：

- A、 $t_1 > t_2$
- B、 $t_1 < t_2$
- C、 $t_1 = t_2$
- D、 $t_1 \neq t_2$
- E、二者大小关系不能肯定

10、以下哪种实验研究设计需要洗脱期：

- A、析因设计
- B、重复测量设计
- C、完全随机设计
- D、随机区组设计
- E、交叉设计

11、在一项有关诱发电位的研究中，甲组受试者被要求借助计算器完成百位数内乘法计算，乙组受试者被要求用心算的方法完成相同题目，对比两组快波（ β 波）占比。同卵双生的姐妹被随机分配至甲组、乙组，这种研究设计属于：

- A. 完全随机设计
- B. 配对设计
- C. 盲法设计
- D. 重复测量设计
- E. 随机区组设计

12、药物临床试验数据统计分析中，PPS 指的是：

- A. 符合方案分析集
- B. 意向性分析集
- C. 全分析集
- D. 安全性数据集

E. 空集

13、下列关于 I 类错误 α 和 II 类错误 β 的说法错误的是：

- A、当样本含量确定时， α 越小， β 越大
- B、当样本含量确定时， α 越大， β 越小
- C、要减小犯 I 类错误的概率，可取较小 α
- D、要减小犯 II 类错误的概率，可取较大 α
- E、若样本含量足够大，可同时避免犯这两类错误

14、进行单因素方差分析时要求

- A. 各个总体均数相等
- B. 各个样本均数相等
- C. 各个样本来自同一总体
- D. 各个总体方差相等
- E. 两样本方差相等

15、设配对设计资料的变量值为 X_1 , X_2 , 则配对资料的秩和检验

- A、把 X_1 和 X_2 的差值的绝对值从小到大编秩
- B、把 X_1 和 X_2 综合从小到大编秩
- C、把 X_1 和 X_2 综合按绝对值从小到大编秩
- D、把 X_1 和 X_2 的差值从小到大编秩
- E、分别按 X_1 和 X_2 从小到大编秩

二、是非题

16、统计学的主要任务是：在变异的背景上描述同一总体的同质性，揭示不同总体的异质性

- A. 正确
- B. 错误

17. 比较 3 组均数是否相同，某人直接用两组独立样本的 t 检验反复多次作两两比较，发现各组间在 $\alpha=0.05$ 的检验水准下差异具有统计学意义。这样做会增大了 I 类错误的概率

- A. 正确
- B. 错误

18. 当样本含量足够大时，t 分布就是正态分布

- A. 正确
- B. 错误

20. 实验设计遵循的“随机原则”是指各个实验组的样本含量相等。

- A. 正确
- B. 错误

21. 总体均数 95% 置信区间的含义是：总体均数落在该区间的可能性为 95%。

- A. 正确
- B. 错误

22. 发展速度属于相对数中率的统计指标。

A. 正确 B. 错误

23. 配对资料的秩和检验，若检验假设 H_0 成立，则正秩和和负秩和相差不会很大。

A. 正确 B. 错误

24. 在两组资料中，回归系数 b 大的那一组两变量相关比较密切。

A. 正确 B. 错误

25. 满足配对 t 检验条件的定量资料，也可以使用随机区组设计的方差分析，并且存在关系 $F = t^2$

三、分析题（2题，每题 20 分，共 40 分）

1、为研究某市男中学生胸围与肺活量之间的关系，研究者随机抽取了该市 200 名男中学生并收集了他们的胸围和肺活量数据

(1) 若要了解胸围与肺活量是否有关，应该如何分析？请阐述具体步骤（不需要列出公式）。（6分）

(2) 若想建立肺活量 Y 与胸围 X 的回归方程，请阐述具体步骤（不需要列出公式）。（6分）

(3) 简述相关与回归的区别与联系（8分）

2、在抗癌药物筛选试验中，拟用 N 只小鼠体重相近归为一个区组，一共 b 个区组，每个区组各有 K 个小鼠。分别观察 K 种药物对小鼠移植性肉瘤 S 180 的抑瘤效果，结局指标为肿瘤的重量 (mg)，各区组内例数相同，未出现缺失值。

考虑区组因素、处理因素做方差分析，得如下方差分析表：

变异来源	SS	V	MS	F
处理组间	60	2		
区组间	180		20	
误差		18	5	
总变异	330	29		

参考界值： $F_{0.05(2,18)} = 3.55$, $F_{0.05(9,18)} = 2.46$, $F_{0.05(2,27)} = 3.35$, $F_{0.05(2,29)} = 3.33$

基于上述部分方差分析表结果及题目中的信息请回答

(1) 把上述方差分析表补充完整，并写出针对处理因素区组因素所做的假设检验的步骤以及得到的统计推断结论。（8分）

(2) 如果在统计分析阶段误将资料按照完全随机方案处理则会有怎样的假设检验过程，会得出怎样的统计推断结论？（8分）

(3) 试比较考虑区组因素与不考虑区组因素，对统计推断结论的影响（4分）

练习题五

一、最佳选择题

1. 在统计学上的参数是指：

A. 总体均数

C. 描述样本特征的指标

B. 描述总体特征的指标

D. 描述人群特征的指标

E. 描述变量特征的指标

2. 抽样误差是指:

A. 个体值和总体参数值之差

D. 不同总体参数之差

B. 个体值和样本统计量之差

E. 以上都不是

C. 样本统计量和总体参数值之差

3. 假设将一个正态分布的资料所有原始数据都加上一个常数, 下列说法正确的是:

A. 均数不变化

D. 标准差不发生改变

B. 变异系数不发生变化

E. 中位数不发生改变

C. 均属是否变化不一定

4. t 分布与标准正态分布有一定的关系, 下述错误的是:

A. 参数数目不同

C. 均为单峰分布

B. t 分布中的自由度趋于无限大时, 曲线逼近标准

D. 对称轴位置在 0

正态分布

E. 曲线下面积的分布规律相同

5. 在抽样研究中, 当样本例数逐渐增多时

A. 标准差逐渐加大

D. 标准误逐渐减小

B. 标准差逐渐加大

E. 标准差趋于 0

C. 标准差逐渐减小

6. 随机区组设计的方差分析中, 将 30 名患者按病情严重程度分为 10 个区组, 每个区组的三个患者随机分到 3 个处理组, 则随机误差项对应的自由度是:

A. 29

D. 19

B. 9

E. 2

C. 18

7. 若完全随机设计的方差分析中检验统计量 $F \approx 1$, 则可推测:

A. $SS_{\text{组内}}$ 不包含处理因素的影响

B. $SS_{\text{组内}}$ 包含处理因素的影响

C. SS_{组间}包含处理因素的影响

D. 方差分析结果拒绝零假设

E. 方差分析结果应不拒绝零假设

8. 从调查数据得知甲市阿尔兹海默症患病率高于乙市，经标准化后甲市阿尔兹海默症患病率低于乙市，其原因最有可能的是：

A. 甲市的诊断水平比乙市高

B. 甲市的阿尔兹海默症健康教育防止工作比乙市好

C. 甲市的老年人口在总人口中所占比例比乙市好小

D. 甲市的老年人口在总人口中所占比例比乙市好大

E. 两市的统计资料存在偏倚

9. 为研究性别和新冠感染者症状的关系，某科研人员收集到新冠感染者 35 例，其中 ICU 住院者男性 12 例，女性 8 例；非 ICU 住院者男性 8 例，女性 7 例。欲比较男性和女性感染者中 ICU 住院率是否相同，应选择的统计分析方法是：

A. 独立的两组二分类资料的 χ^2 检验

E. Fisher 确切概率法

B. 独立的两组二分类资料的矫正 χ^2 检验

D. 配对设计的二分类资料的矫正 χ^2 检验

C. 配对设计的二分类资料的 χ^2 检验

10. 利用 χ^2 检验公式不适合解决的实际问题是：

A. 比较两种药物的有效率

B. 拟合优度检验

C. 老年人与中年人新冠感染症状“无症状、轻、中、重”的严重程度比较

D. 接种流感疫苗的意愿与职业种类的关联性检验

E. 比较多个人群职业种类的频率分布

11. 在统计分析中，是否选用非参数统计方法：

A. 根据研究目的和数据特征作出决定

D. 要看哪个 P 值更小

B. 可在算出几个统计量和得出初步结论后进行选择

E. 根据研究者的意愿

C. 要看哪个统计结论复核预期结果

12. 适合分析糖尿病人的血糖水平与胰岛素水平之间关系的方法是：

- A. 配对比较的 t 检验
- B. 成组比较的 t 检验
- C. 秩和检验
- D. χ^2 检验
- E. 相关分析或回归分析

13. 以下类型的统计图中，纵坐标须从 0 开始的图形为：

- A. 线图
- B. 直条图
- C. 散点图
- D. 箱式图
- E. 饼图

14. 为了解某市儿童心理健康状况，分别从不同年龄段儿童中随机抽取部分进行抽样调查，这种抽样方法

属于：

- A. 单纯随机抽样
- B. 系统抽样
- C. 整群抽样
- D. 分层抽样

15. E. 多阶段抽样

16. 关于完全随机设计两组样本均数比较，样本量的估计要素不包括以下哪项：

- A. 容量误差
- B. 第一类错误
- C. 第二类错误
- D. 总体标准差
- E. 总体标准误

二、是非题

三、

17. 统计研究的过程包括：研究设计、收集资料、整理资料、分析资料等环节，其中分析资料是研究成功的关键环节。

A. 正确 B. 错误

18. 均数表示一组同质的计量资料的平均水平，不同资料的均数之间均可以直接相互比较。

A. 正确 B. 错误

19. 完全随机设计的方差分析在检验多组间差异是否有统计学意义时，应该查 F 统计量的双侧界值表。

A. 正确 B. 错误

20. 进行两均数差别的假设检验时，当 $P \leq 0.05$ 时，则拒绝 H_0 ；当 $P > 0.05$ 时，则接受 H_0 ，认为两总体均数无差别。

A. 正确 B. 错误

21. 标准化法的目的是消除内部构成的不同对粗略比较的影响，标准化后的率即可以反映该资料当时当地的实际水平，也可以反映相互比较的资料间的相对水平。

A. 正确 B. 错误

22. 自由度为 1 的 χ^2 分布是对称分布。

A. 正确 B. 错误

23. 在进行两组等级资料比较两组的中心位置差异时，宜用秩和检验。

A. 正确 B. 错误

24. 线性相关系数假设检验时，其自由度是 $n-2$ 。

A. 正确 B. 错误

25. 抽样误差最大的抽样方法是单纯随机抽样。

A. 正确 B. 错误

26. 实验设计设立对照的首要目的是通过对比发现差异，消除会减少实验误差，显示处理因素的效应。

- A. 正确 B. 错误

三、分析题（每题 20 分）

27. 中山大学附属第一医院神经内科对两组癫痫患者的发作次数进行研究，甲组患者为父母一方有癫痫病史，乙组患者为父母双方均无癫痫病史。所有纳入研究的患者本人均需有三级以上医院对其癫痫疾病的确诊，父母的癫痫病史也需有医疗机构的诊断。课题组队两组患者在确诊后近 3 年的癫痫年均发作次数进行比较。

(1) 本研究属于观察性研究，还是干预期研究。为什么？(4 分)

(2) 比较两组年均发作次数时需进行假设检验，该用 *t* 检验还是秩和检验？

(3) 基于前述第(2)问确定的统计学分析方法，请写出假设检验的步骤（不用具体计算）。

(4) 在以上两组患者的确诊后 3 年内癫痫年均发作次数比较的研究中，有人建议每组样本量

$n_1 = n_2 = 50$ ，另有人建议 $n_1 = n_2 = 100$ 。你的观点是怎样的？

28. 随机抽取 10 名 20 岁男性青年的身高和前臂长, 数据记录如下:

表 1 10 名 20 岁男性青年的身高与前臂长

编号	1	2	3	4	5	6	7	8	9	10
身高 (cm)	170	173	155	173	188	178	183	180	165	160
前臂长 (cm)	45	42	41	47	50	47	46	49	43	44

(1) 若欲了解身高 Y 与前臂长 X 是否有关, 该如何分析? 请阐述具体步骤 (不用列出公式)。

(2) 若想建立身高 Y 与前臂长 X 的回归方程, 请阐述具体步骤 (不用列出公式, 不用列出具体的线性回归方程)。

(3) 对拟建立的回归方程进行假设检验。请补充下面回归分析结果表, 并作出统计推断 (其中 $F_{0.05(v_1, v_2)} = 5.32$)

变异来源	V	SS	MS	F
回归				
剩余		311.020		
总变异		962.500		

练习题六

分别对 8 名未患妊娠合并症得孕妇和 9 名患有妊娠合并症得孕妇进行空腹葡萄糖测试, 结果见表 1. 问两类孕妇得葡萄糖耐受能力是否不同? 如何对该资料做统计分析?

要求写出完整的假设检验步骤, 须计算检验统计量和 P 值

表 1 两组孕妇葡萄糖耐受水平得测试结果 (mmol/L)

未患妊娠合 并症组	2.50	1.60	1.70	3.00	1.40	4.50	4.60	2.50
患有妊娠合 并症组	3.70	5.50	5.00	5.20	1.80	1.20	1.60	3.40

练习题七

一、最佳选择题

1. 临床试验中某种药物的疗效(无效、好转、痊愈)资料属于
A. 计量资料 B. 二分类资料 C. 无序分类资料 D. 等级资料 E. 以上都不是
2. 某定量资料的分布性质尚不明确,要描述其离散趋势,宜选择:
A. 中位数
B. 方差
C. 算术均数
D. 四分位数间距
E. 标准差
3. 横轴上、标准正态分布曲线下从 0 到 2.58 的面积为:
A. 47.5%
B. 45%
C. 97.5%
D. 95%
E. 49.5%
4. 关于 t 分布的特点,以下说法错误的是:
A. t 分布曲线以 0 为中心
B. t 分布曲线形状相比正态分布曲线更为高瘦
C. 分布曲线为单峰分布
D. t 分布曲线形状与自由度 v 有关
E. 分布的自由度取值范围是 1 到正无穷
5. 下列关于置信区间与假设检验的说法,错误的是:
A. 置信区间不能提示差别是否具有实际意义
B. 置信区间可以提供假设检验所不能提供的信息
C. 假设检验及置信区间均属于统计推断
D. 假设检验可用于推断总体参数之间是否存在差异
E. 一般情况下,置信区间估计所采用的统计量与假设检验的相同
6. 与完全随机设计方差分析相比,随机区组设计方差分析的特点是:
A. 从原来的组内变异中进一步分离出区组间变异,因此随机区组设计的检验效率更高
B. 以处理组间的均方与区组间均方的比值,计算处理组对应的检验统计量 F
C. 随机误差固定时,处理组间的均方越大,区组间的均方越小,因为它们的总和是一定的
D. 在总自由度的计算原则上,与完全随机设计不同
E. 在总变异的计算原则上,与完全随机设计不同
7. 从 3 个总体中各随机抽取 4 个样本,用完全随机设计的方差分析比较 3 个总体均数是否相同(假设资料满足方差分析的前提条件),计算得到离均差平方和 SS 组间=536、SS 组内=828,则相应的均方分别为:
A. 268、92

- B. 134、103.5
- C. 179、92
- D. 179、82.8
- E. 134、82.8

8. 动态数列指标中, 发展速度和增长速度的关系是:

- A. 发展速度=增长速度-1
- B. 增长速度=发展速度-1
- C. 发展速度=增长速度-100
- D. 增长速度=发展速度-100
- E. 增长速度=(发展速度-1)100

9. 下列哪种资料不适合用 χ^2 检验分析:

- A. 分析是否饮酒对成年人血液中甘油三酯水平(正常、不正常)的
- B. 比较不同性别人群血红蛋白含量的平均水平有无差异
- C. 比较两种药物治疗高血压的疗效(有效、无效)
- D. 比较配对设计中两种诊断方法的阳性率是否相同
- E. 比较急性白血病和慢性白血病患者血型总体构成有无差别

10. 关于四格表资料的 χ^2 检验, 以下说法不正确的是(T 为理论频数, n 为样本量)

- A. 所有格子 $T \geq 5$, 且 $n \geq 40$ 时, 直接采用 χ^2 检验的公式计算检验统计量
- B. 有一个格子 $T < 1$ 或 $n < 40$, 不宜用 χ^2 检验, 使用 Fisher 确切概率法进行分析
- C. 当 $n \geq 240$ 、但有一个格子 $1 \leq T < 5$, 可用 χ^2 检验的连续性校正公式计算检验统计量
- D. 检验统计量 χ^2 反映了理论频数和实际频数的差距
- E. 当四格表的行合计数和列合计数不变时, 如果格子的实际频数变化, 则对应的理论频数也会发生变化

11. 多组计量资料的比较, 当分布类型未知时, 应选择的统计方法是

- A. 方差分析
- B. Wilcoxon 秩和检验
- C. Kruskal-Wallis H 检验
- D. 列联表 χ^2 检验
- E. t 检验

12. 配对设计资料的秩和检验, 对子之间差值分别 -0.5、1.5、3.0、5.0、0.0、-7.0、1.0、0.5、0.0、-0.1, 用查表法判断 P 值, 则参与编秩的差值个数 N 为:

- A. 8
- B. 10
- C. 9
- D. 2
- E. 无法判断

13. 以下关于线性相关系数(r)的说法中, 错误的是:

- A. r 的计算不受变量所取量纲的影响
- B. r 的取值范围是 [-1, +1]
- C. r 为正, 则表明两变量变化方向相同

D. r 的取值越接近 -1 或 1, 说明两变量相关程度越高
E. r 可衡量两个定量变量之间的非线性关系

14. 用最小二乘法确定线性回归方程的原则是

- A. 各实测点距回归直线的纵向距离之和最小
- B. 各实测点距回归直线的纵向距离的平方和最小
- C. 各实测点距回归直线的垂直距离之和最小
- D. 各实测点距回归直线的垂直距离的平方和最小
- E. 各实测点距回归直线的水平距离的平方和最小

15. 描述血清甘油三酯水平与胰岛素水平的相关关系, 以下统计图最合适的是

- A. 直方图
- B. 箱式图
- C. 线图
- D. 散点图
- E. 条图

16. 某医院精神科分别采用两种心理疗法对抑郁症患者进行治疗, 测量治疗后 0 周、2 周、4 周、6 周、8 周、10 周、12 周两组受试者的 SDS 抑郁量表得分情况, 为了直观显示两组平均得分的变动情况, 最合适的统计图是:

- A. 饼图
- B. 线图
- C. 箱式图
- D. 直方图
- E. 百分条图

17. 对广州市儿童膳食营养情况进行抽样调查, 先按城乡分层, 然后从各层中随机抽取若干社区行政村, 再从社区/行政村中随机抽取若干学校, 对抽中学校的全部儿童进行调查, 这种抽样方法是:

- A. 整群抽样
- B. 分层抽样
- C. 滚雪球抽样
- D. 多阶段抽样
- E. 单纯随机抽样

18. 按调查对象的范围, 常用调查方法包括

- A. 普查、病例-对照调查
- B. 病例-对照调查、队列调查
- C. 非全面调查、横断面调查
- D. 普查、抽样调查
- E. 抽样调查、队列调查

19. 实验研究随机化分组的主要目的是:

- A. 减少抽样误差
- B. 减少实验例数
- C. 保证结果稳定性
- D. 提高检验准确度
- E. 使各组非处理因素的分布尽量保持均衡一致

20. 某医师观察某种新药治疗急性气管炎的疗效, 用该疾病的的标准用药氨苄西林作为对照, 对 140 例患者进行临床试验。患者入院时, 按户口所在地进行分组, 城市户口分入治疗组, 农村户口分入对照组, 统计结果显示新药疗效优于氨苄西林, 关于该临床试验的相关评论, 不正确的是:

- A. 该临床试验未遵循随机化原则
- B. 该临床试验中试验组和对照组的选择不合理
- C. 该临床试验的结论不可靠, 很可能存在偏倚
- D. 两组患者的非处理因素可能不均衡
- E. 该临床试验未遵循重复原则

二、是非题

21. 变量的转换只能从定量变量—有序分类变量→无序分类变量, 不能做相反方向的转换

- A. 正确
- B. 错误

22. 比较 20 头大象体重和 20 只大白兔体重的变异程度, 宜采用四份位数间距

A. 正确 B. 错误

23. 假设检验的检验水准 α 是人为设定的。

A. 正确 B. 错误

24. 两组完全随机设计计量资料均数比较的假设检验, t 检验和方差分析是等价的。

A. 正确 B. 错误

25. 对某疾病患者年龄别构成比数据进行分析, 结果显示, 30~40 岁组构成比为 17.1%, 60 岁以上组构成比为 32.9%, 可认为 60 岁以上人群比 30~40 岁人群更易患有该疾病

A. 正确 B. 错误

26. 非参数检验适用于总体分布不确定的资料

A. 正确 B. 错误

27. 多组率比较的 χ^2 检验, 得到 $P < 0.05$, 可认为多组总体率之间两两全不相等。

A. 正确 B. 错误

28. 双变量正态分布的资料, 线性回归分析得到样本回归系数 $b < 0$, 假设检验 $P < 0.05$ 。则线性相关分析时, 可得到两变量间呈负相关的结论

A. 正确 B. 错误

29. 抽样调查方法中, 整群抽样主要适用于群间差异较大的情形。

正确 B. 错误

30. 实验设计中, 计算样本量时, 容许误差 δ 越大, 研究所需样本含量越大。

A. 正确 B. 错误

三、分析题

1. 为了解男性服用肠溶醋酸棉酚片是否会导致精子浓度下降, 研究者随机选取 8 名健康男子, 给予相同剂量的肠溶醋酸棉酚片口服 3 个月, 对每名男子服药前后进行精液检查, 结果见表 1

表 1 8 名男子服用肠溶醋酸棉酚片前后精子浓度

编号	1	2	3	4	5	6	7	8
服药前 (万/ml)	6000	22000	5900	4400	6000	6500	26000	5800
服药后 (万/ml)	660	5600	3700	4200	6300	1200	1800	2200

请分析并回答以下问题

(1) 上述研究用了什么试验设计方法? 资料为何种类型资料?(6 分)

(2) 若该资料不服从正态分布, 用什么统计方法进行假设检验? 简述理由。(4 分)

(3) 欲了解服药前后精液中精子浓度有无区别, 应如何进行假设检验? 请阐述具体步骤。(10 分)

临界值信息: 当样本量为 8 时 ($N=8$) 的 Wilcoxon 符号秩和检验 T 临界值见表 2; 当最小样本量组样本量为 8 时 ($n_1=8$) Wilcoxon 秩和检验 T 临界值见表 3。不同自由度 t 分布的临界值如下:

$t_{0.05/2, 7}=2.36$

$t_{0.05/2, 8}=2.30$,

$t_{0.05, 7}=1.89$,

$t_{0.05, 8}=1.86$

$t_{0.05/2, 13}=2.16$,

$t_{0.05/2, 14}=2.14$,

$t_{0.05, 13}=1.77$,

$t_{0.05, 14}=1.76$ 。

表 2 N=8 时 Wilcoxon 符号秩和检验 T 临界值

N 8	单侧: 0.05	0.025	0.010	0.005
	双侧: 0.10	0.050	0.020	0.010
	5-31	3-33	1-35	0-36

表 3 m₁=8 时 Wilcoxon 秩和检验 T 临界值

m_2-m_1 0	单侧: 0.05	0.025	0.010	0.005
	双侧: 0.10	0.050	0.020	0.010
1	51-58	49-87	45-91	43-93
2	54-90	51-93	47-97	45-99
3	56-96	53-99	49-103	47-105
	59-101	55-105	51-109	49-111

注: m_1 为样本量较小组样本量, m_2 为样本量较大组样本量。

2. 从中山大学随机抽取 20 名大三学生, 调查了他们的抑郁症状及压力的情况, 其中抑郁症状及压力均采用标准化量表进行测量, 结果见表 4。

表 4 20 名大三学生抑郁症状及压力评分

编号	1	2	3	4	5	6	7	8	9	10
抑郁症状(CESD-20 评分)	30	32	28	27	22	37	35	15	16	17
压力(SSCS 评分)	43	49	40	40	31	50	50	25	22	25

表 4 20 名大三学生抑郁症状及压力评分(续表)

编号	11	12	13	14	15	16	17	18	19	20
抑郁症状(CESD-20 评分)	40	45	50	53	10	36	33	32	20	55
压力(SSCS 评分)	60	60	70	82	17	57	50	45	35	80

- (1) 欲了解大学生抑郁症状评分 Y 与压力评分 X 是否有关, 该如何分析? 请阐述具体步骤(不需要列出公式)。
 (2) 若想分析抑郁症状评分 Y 与压力评分 X 间的数量依从关系, 该如何分析? 请阐述具体步骤(不需要列出公式)。
 (3) 对拟建立的国归方程进行假设检验。请补充表 5 方差分析结果, 并作出统计推断(其中 F0.05(v1, v2)=4.41)。

表 5 方差分析表

变异来源	V	SS	MS	F
回归		3041.350		4.289
剩余				
总变异				

练习题八

1. 某研究记录了 50 名病人体重的测定结果: 小于 50kg 的 13 人, 介于 50kg 和 70kg 间的 20 人, 大于 70kg 的 17 人, 此种资料属于:
A. 计量资料 B. 计数资料 C. 无序分类资料 D. 等级资料 E. 以上都不是
2. 与四分位数间距相比, 极差:
A. 更接近标准差
B. 更小
C. 更接近中位数
D. 稳健性更好
E. 稳健性更差
3. 在一项调查空腹血糖的研究中, 研究者从广州市成年男性中, 随机抽取了 1000 名男性样本, 并测量了他们的空腹血糖。如果空腹血糖服从正态分布, 用以下哪个指标描述平均水平最好?
A. 算术均数 B. 中位数 C. 频数分布 D. 几何均数 E. 标准差
4. 两组数据来源于不同的正态分布总体, 以下哪个指标更小时, 样本均数的可靠性更好?
A. 样本标准差
B. 变异系数
C. 样本标准误
D. 样本均数
E. 样本中位数
5. 根据样本资料算得健康成年男子总胆固醇含量均数的 95% 置信区间为 180.46mg/dL
183.69 mgdL, 其含义是:
A. 有 95% 的置信度认为该区间包含了总体均数
B. 总体中有 95% 的个体值落于此区间
C. 样本中有 95% 的观察值在此范围内
D. 该区间包含总体片数的可能性为 95%
E. 样本均数落在此区间的可能性为 95%
6. 若进行两独立样本检验, 检验结果 $P > 0.05$ 说明:
A. 两总体均数的差别较小
B. 两总体均数的差别较大
C. 尚不能认为两总体均数有差别
D. 支持两总体均数有差别的结论
E. 可以确认两总体均数无差别
7. 以下哪一项不属于随机区组设计方差分析变异分解时所涉及的变异?
A. 总变异
B. 处理组间变异
C. 个体内变异
D. 误差变异
E. 区组间变异
8. 关于动态数列指标的描述中不正确的:
A. 增长速度 = 发展速度 - 1
B. 发展速度和增长速度都是相对比
C. 平均发展速度是发展速度的几何平均数
D. 环比发展速度反映事物逐期的波动或发展速度, 以基线期指标作基数

- E. 发展速度表示报告期指标的水平相对于基线期(或前一期)指标的比例
10. 多组定量资料的比较, 当分布类型未知时, 应选择的统计方法是!
- A. Kruskal-Wallis H 检验
- B. Wilcoxon 秩和检验
- C. 列联表 χ^2 检验
- D. Z 检验
- E. 方差分析
11. 进行 Wilcoxon 符号秩和检验, H_0 应该为:
- A. 两总体分布一致
- B. 差值的总体均数为 0
- C. 两总体均数相同
- D. 差值的总体中位数为 0
- E. 两总体的中位数相同
12. 在直线回归分析中, 剩余平方和(SS_{res})反映的是:
- A. 因变量的变异程度
- B. 自变量入可解释的变异
- C. 扣除 X 影响后 Y 的变异程度
- D. 扣除入影响后因变量预测值 P 的变异程度
- E. 回归系数 b 的变异程度
13. 某医生收集患者 60 例, 观察血小板数和患者出血症状(阴性、出现、中度、重度), 欲分析出血症状程度与血小板数量的关系, 可进行:
- A. Pearson 相关分析
- B. Spearman 等级相关分析
- C. 两独立样本比较的 t 检验
- D. 单因素方差分析
- E. 两独立样本比较的 χ^2 检验
15. 两个定量变量间的直线相关分析, 应先绘制:
- A. 直方图 B. 半对数线图 C. 条图 D. 散点图 E. 饼图
16. 统计表的制表原则是:
- A. 尽量借助软件, 避免手工误差
- B. 精美、精致、精确、精当
- C. 数值信息丰富, 线条简洁明快
- D. 充分借助线条、逻辑呈现完整
- E. 重点突出、简单明了; 主谓分明、层次清楚
17. 为了解某人群血压的平均水平, 进行调查研究, 样本量估算时, 无需考虑的要素是:
- A. 第 I 类错误
- B. 第 II 类错误
- C. 标准差
- D. 容许误差
- E. 资料类型
18. 以下抽样方法中, 来自同一个总体且样本量相同的情形下, 哪一种方法的抽样误差最小?
- A. 简单随机抽样
- B. 系统抽样
- C. 分层抽样
- D. 整群抽样

E. 偶遇抽样

19. 实验设计应遵循随机、均衡、对照和重复的基本原则，其主要目的是：

A. 使统计分析更简单

B. 严格控制或消除随机误差 C. 使实验操作更简单

D. 能更快得到实验结果

E. 避免或尽量减少非处理因素对结果的影响

20. 某药物的效果观察欲用“双盲”试验，所谓“双盲”，指：

A. 试验组接受药物，对照组接受安慰剂

B. 仅试验对象不知道自己是试验组还是对照组

C. 观察者和试验对象都不知道安慰剂的性质

D. 观察者和试验对象都不知道谁接受药物谁接受安慰剂

E. 试验组和对照组都不知道谁是观察者

二、是非题

21. 比较 3 岁女童体重与 50 岁妇女体重的变异程度，宜采用的指标是标准差

A. 正确 B. 错误

22. 若某单峰分布的中位数小于算术均数，则该分布是正偏峰分布。

A. 正确 B. 错误

23. t 分布的自由度越大，相同尾部面积所对应的 t 界值的绝对值越大

A. 正确 B. 错误

24. 增加样本含量可以减少抽样误差，所以样本含量越大越好。

A. 正确 B. 错误

25. 随机区组设计的方差分析中，各区组内的样本数量可以少于处理因素水平的个数。

A. 正确 B. 错误

26. 5 岁以下儿童死亡率(计算方法：同年 5 岁以下儿童死亡数/同年活产儿总数 ×1000‰)是相对比指标。

A. 正确 B. 错误

27. 当四格表的周边合计数不变时，如果某个格子的实际频数变化，则其理论频数随该格子实际频数的增减而增减。

A. 正确 B. 错误

28. 符合/检验的数值变量资料，如果采用秩和检验进行分析，可能会增加犯第 I 错误的概率。

A. 正确 B. 错误

29. 符号秩和检验出现两个差值为零，编秩时，它们不参与编秩。

A. 正确 B. 错误

30. 两个变量之间的线性相关系数等于 0，说明二者间没有任何形式的关联关系。

A. 正确 B. 错误

31. 直线相关分析中经假设检验得出总体相关系数不为 0，且 P 值很小，尚不可以认为变量间关系很密切。

A. 正确 B. 错误

32. 统计图的下方应该有标题，概略地介绍统计图拟说明的问题。

A. 正确 B. 错误

33. 研究设计包括如何收集资料、整理资料，不包括如何分析资料

A. 正确 B. 错误

34. 实验设计中重复的原则是指对同一实验单位的重复观察。

A. 正确 B. 错误

35. 观察某疫苗预防儿童某种传染病的效果，试验组儿童接种该疫苗，对照组儿童不接种这种疫苗，这种对照称为实验对照。A. 正确 B. 错误

练习题九

教材：（北大出版社第1版）

一、选择题

1. 某人记录了30位患者的空腹血糖测量结果：小于6.0的13人，6.0~7.2的4人，大于7.2的3人，此处空腹血糖属于什么类型的变量？

- A. 离散型定量变量 B. 连续型定量变量 C. 无序多分类变量

- D. 有序多分类变量 E. 二分类变量

2. 关于变量的性质，下列描述中错误的是：

- A. 定量变量可以转化为定性变量，但会损失一定信息

- B. 血型(A、B、AB、O型)属于无序多分类变量

- C. 连续型变量和离散型变量是定性变量的两种类型

- D. 无序多分类变量的变量类别间不具有顺序大小的关系

- E. 定性变量无法转化为定量变量

3. 以下哪个指标对离群值最不敏感？

- A. 中位数 B. 算数均数 C. 极差 D. 标准差 E. 几何均数

4. 已知某中学初一男生身高均数为140.1cm，标准差为7.5cm；体重均数为35.6kg，标准差为5.1kg，则：

- A. 该中学初一男生体重的变异程度大于身高的变异程度

- B. 该中学初一男生体重的变异程度等于身高的变异程度

- C. 该中学初一男生体重的变异程度小于身高的变异程度

- D. 因变量不同，无法比较变异程度

- E. 因变量的测量单位不同，无法比较变异程度

5. 某市对管辖范围内的城市和农村居民糖尿病的患病率进行抽样调查，得到城市和农村居民患病率之差的95%置信区间为(0.3%，2.8%)。如果进行城市和农村居民患病率是否有差异的假设检验，检验水准 $\alpha=0.05$ ，可以得到：

- A. 差异有统计学意义 B. 差异无统计学意义 C. P值小于0.025

- D. P值小于0.028 E. 无法从以上信息判断差异有无统计学意义

6. 从某地随机抽取400名健康成年女性，进行血清总蛋白含量检测，计算得到血清总蛋白含量均数为74g/L，标准差为4g/L，则其总体均数95%置信区间为：

- A. $74 \pm 1.96 \times 2/20$ B. $74 \pm 1.96 \times 4/20$ C. $74 \pm 1.96 \times 4^2/20$

- D. $74 \pm 1.96 \times 4$ E. $74 \pm 1.96 \times 4/400$

7. 下列哪个说法是错误的？

- A. H_0 成立时拒绝 H_0 则犯了II类错误 B. H_0 成立时拒绝 H_0 称为假阳性

- C. H_1 成立时不拒绝 H_0 称为假阴性 D. 检验功效为 $1-\beta$ ，其中 β 为犯II类错误的概率

- E. H_0 成立时不拒绝 H_0 的概率为 $1-\alpha$ ，其中 α 为检验水准

8. 两组完全随机设计定量资料的均数比较，假设资料满足参数检验的条件，应用方差分析与t检验进行假设检验所得结果：

- A. 方差分析结果更准确 B. t检验结果更准确 C. 完全等价且 $F=t$

- D. 完全等价且 $F=\sqrt{t}$ E. 完全等价且 $t=\sqrt{F}$

9. 完全随机设计的方差分析中，共设3组，每组10例。在回答组间差异是否有统计学意义时，设检验水准 $\alpha=0.05$ ，计算得到的统计量F应该与以下哪个界值进行比较？

- A. $F_{0.05(2,27)}$ B. $F_{0.05(3,30)}$ C. $F_{0.025(1,1)}$ D. $F_{0.025(2,27)}$ E. $F_{0.025(3,30)}$

10. 随机区组设计资料的方差分析中，将总变异分解为：

- A. 组间变异、组内变异两部分 B. 标准差、标准误两部分

- C. 处理、区组和误差三部分 D. 抽样、系统和随机测量三部分
 E. 以上说法均不正确
11. 某地区某年的医疗人均支出记为 a_0 , 以该年份为基线期, 后续年份医疗人均支出记为 a_1, a_2, \dots, a_n , 则该地区医疗人均支出的年平均增长速度为:
- A. $\frac{a_0 + a_1 + \dots + a_n}{n}$ B. $\sqrt[n+1]{a_0 a_1 \dots a_n}$ C. $\sqrt{\frac{a_n}{a_0}}$ D. $\sqrt{\frac{a_n}{a_0}} - 1$ E. $\sqrt{n-1} \sqrt{\frac{a_n}{a_0}}$

12. 某医院用两种不同方法对 97 例缺铁性贫血患者进行诊断, 收集到结果如下表 1, 若想知道两种方法的检出率有无差别, 应使用的统计学方法为:

表 1 两种方法的诊断结果

甲方法	乙方法		合计
	+	-	
+	74	8	82
-	14	1	15
合计	88	9	97

- A. 两独立样本资料的 χ^2 检验 B. 两独立样本资料连续性校正 χ^2 检验 C. Fisher 确切概率法
 D. 配对设计资料的 χ^2 检验 E. 拟合优度 χ^2 检验
13. 多个独立样本列联表资料的 χ^2 检验中, 如果 $\chi^2 = 0$, 下列说法正确的是:
- A. 不可能出现 $\chi^2 = 0$ 的情形 B. 与 $\chi^2 = 1$ 完全等价
 C. 这种情形下 P 值一定小于 0.05 D. 此时实际频数是理论频数的 2 倍 E. 此时理论频数和实际频数完全相等
14. 配对符号秩和检验中 (原始观测值的对子数 n=10), 正秩和 T 的值不可能等于:
- A. 0 B. 5 C. 10 D. 45 E. 64

15. 对于同一组资料, 若同时进行了线性相关分析与线性回归分析, 则相关系数 $r(r \neq 0)$ 与回归系数 $b(b \neq 0)$ 之间的关系正确的是:

- A. $r=b$ B. $r > b$ C. $r < b$ D. $r \times b < 0$ E. $r \times b > 0$

16. 下列哪项不是箱式图中的元素:

- A. 最大值 B. 异常值 C. 均数 D. 四分位数间距 E. 最小值

17. 某团队研制出一种预防丙型病毒性肝炎的疫苗, 要评价该疫苗的效果, 以下人群哪个最适合作为研究对象?

- A. 丙型病毒性肝炎感染的高危人群
 B. 丙型病毒性肝炎发病率低的人群
 C. 曾感染过丙型病毒性肝炎的人群, 体内有无抗体均可
 D. 曾感染过丙型病毒性肝炎, 但体内没有抗体的人群
 E. 曾感染过丙型病毒性肝炎, 且体内有抗体的人群

18. 某研究者欲通过抽样调查获得某地区成年女性血红蛋白的平均水平, 在样本量计算中, 以下哪项不是需要考虑的因素?

- A. 估计的置信度 B. 该人群血红蛋白的变异情况
 C. 可接受的估计精度 D. 抽样方式 E. 第 II 类错误

19. 在以下哪种形式的实验研究设计, 患者在整个研究过程当中需要先后接受 1 种以上的干预措施?

- A. 完全随机设计 B. 交叉设计 C. 配对设计 D. 随机区组设计 E. 盲法设计

20. 相较于效力试验, 以下哪一项是临床研究中效果试验的特点?

- B. 患者服药的依从性更高
- C. 研究耗时短
- A. 患者的同质性更差
D. 疗效用单个结局指标来反映 E. 严格执行既定的纳入/排除标准

二、是非题

在 A 和 B 中选出最佳答案，将答案对应的字母在选择题答题卡上涂黑（填在其他地方无效）。

21. 描述总体特征的指标是参数，而描述样本特征的指标是统计量。
 A. 正确 B. 错误
22. 在负偏态分布数据中，通常均数小于中位数。
 A. 正确 B. 错误
23. 从一个正态分布的总体中随机抽取样本，总体方差已知，如果样本量增加，则总体均数的 95% 置信区间将更宽。
 A. 正确 B. 错误
24. 进行配对检验时，研究人员 A 使用干预后的值减去干预前的值来计算差值，研究人员 B 使用干预前的值减去干预后的值来计算差值，则他们计算得到的值方向相反，但得到的统计结论相同。
 A. 正确 B. 错误
25. 完全随机设计方差分析中，组间均方仅表示抽样误差的大小。
 A. 正确 B. 错误
26. 与完全随机设计方差分析相比，随机区组设计方差分析从原来的组间变异中进一步分离出了区组间变异，因此无论何种资料，随机区组设计方差分析的检验效能要高于完全随机设计方差分析的检验效能。
 A. 正确 B. 错误
27. 计算标准化死亡率的目的是减少死亡率估计的抽样误差。
 A. 正确 B. 错误
28. χ^2 检验可用于探究饮酒行为对成年人血甘油三酯水平（正常、稍高、极高）的影响。
 A. 正确 B. 错误
29. 两独立样本的四格表资料，在所有格子的理论频数大于 1、总样本量大于 40 时，做 χ^2 检验时不必校正。
 A. 正确 B. 错误
30. 对于定量资料的比较，在满足参数检验方法条件下、采用非参数检验方法分析，可能会增加 I 类错误。
 A. 正确 B. 错误
31. 最小二乘原则是指所有原始数据点距离回归直线的纵向距离平方和最小。
 A. 正确 B. 错误
32. 半对数线图可用于描述某指标随时间变化的速度。
 A. 正确 B. 错误
33. 调查研究按抽样方式是否遵循等概率原则，可分为概率抽样调查和全面调查。
 A. 正确 B. 错误
34. 某调查研究，将总体分为若干个群，当各群间差异较大时，选择整群抽样比较恰当。
 A. 正确 B. 错误
35. 安慰剂对照与空白对照是相同的概念，因为分在这样组别的研究对象，均不会被观测到任何实验效应。
 A. 正确 B. 错误

三、分析题

1. 测得铅作业工人与非铅作业工人的血铅值 ($\mu\text{g}/100\text{g}$) 见表 2，设资料不满足参数检验的条件。

表 2 铅作业与非铅作业工人血铅值 ($\mu\text{g}/100\text{g}$) 的检测结果

非铅作业组	5	5	6	7	9	12	13	15	18	21
铅作业组	17	18	20	25	34	41	44			

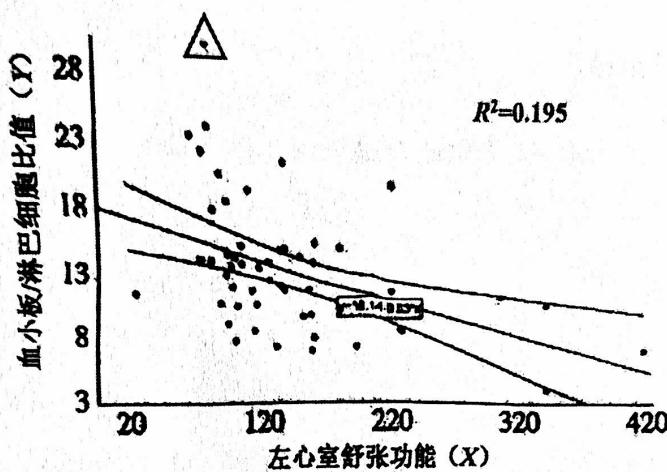
- (1) 用什么指标描述平均水平？请具体算出来。
 (2) 用什么指标描述变异度？（只需列出采用的指标，不必具体算出）
 (3) 请给出统计学的假设检验完整过程，并回答，哪一组患者的血铅平均水平高？用到的统计表见表 3。

表3 T检验界值表 (双侧 $P=0.05$)

n_1 (较小的 n)	$n_2 - n_1$				
	1	2	3	4	5
5	18.42	20.45	21.49	22.53	23.57
6	27.57	29.61	31.65	32.70	34.74
7	38.74	40.79	42.84	44.89	46.94

2. 某研究开展老年心力衰竭病人血小板/淋巴细胞比值(Y)与左心室舒张功能(X)的关联研究。左心室舒张功能通过二尖瓣E峰流速与二尖瓣环舒张早期峰值速度比值评判。数据满足线性相关分析和线性回归分析的应用条件。基于53例老年人的数据，该研究发现 Y 与 X 呈负相关 ($r = -0.46, P < 0.01$)，回归方程 $P = 18.15 - 0.03 \times X$ ，绘制出相应的回归直线图如图1所示。

- (1) 请写出上述研究中对回归系数进行假设检验的具体步骤，并给出假设检验的结论（不需要列出统计量的具体计算过程）。
- (2) 请分别解释回归方程 ($P = 18.15 - 0.03 \times X$) 中回归系数、及图1中 $R^2 = 0.195$ 的含义。
- (3) 若研究者在完成上述分析后，发现图1中△圈出的点是离群值，请你简述此时的统计分析策略（不需要进行具体计算）。

图1 左心室舒张功能(X)与血小板/淋巴细胞比值(Y)的简单线性回归分析结果

练习题十

一、选择题

1. 现有一研究想要调查广州市越秀区常住居民高血压情况（轻度、中度、重度）的构成比，调查地选择应当在，原因是？
A. 医院，更易获得数据
B. 医院，数据更具代表性
C. 医院，诊断标准更明确
D. 社区，患有轻度高血压的病人也能被调查出来
E. 社区，居民配合度更高
2. 变异是指？
A. 用不同的降压药治疗病情、年龄相同的原发性高血压病人，其血压值降低的有多有少
B. 用相同的降压药治疗病情、年龄不同的原发性高血压病人，其血压值降低的有多有少
C. 用相同的降压药治疗病情、年龄相同的原发性高血压病人，其血压值降低的有多有少
D. 用相同的降压药治疗病情、年龄相同的原发性高血压病人，其血压值降低的相同
3. 以下哪项是描述定量资料平均水平的统计指标？
A. 四分位数间距
B. 极差
C. 中位数
D. 方差
E. 标准差
4. 以下哪项是正偏态分布资料的特点？
A. 均数值大于中位数值
B. 均数值等于中位数值
C. 均数值小于中位数值
D. 无法判断
5. 分别调查 A, B 两地居民的肝吸虫病感染情况，选取两地各 1000 名居民进行调查，并得出以下统计量：
A 地 99% 居民患肝吸虫病的 95% 置信区间为 (4.1%, 7.9%); B 地 99% 居民患肝吸虫病的 95% 置信区间为 (9.3%, 14.7%)，若按照同一检验水准 $\alpha = 0.05$ ，则下列说法正确的是：
A. 两地居民肝吸虫病感染情况差异有统计学意义
B. 两地居民肝吸虫病感染情况差异无统计学意义
C. P 值 < 0.05
D. P 值 < 0.10
E. 无法从以上信息判断差异是否有统计学意义
6. 关于红细胞总体均数的抽样分布（样本量为 5），下列说法正确的是？
A. 抽样分布的均数一定等于红细胞总体均数
B. 抽样分布的标准差就是红细胞总体的标准差
C. 抽样分布一定是正态分布
D. 抽样分布的形状与样本量无关

D

7. 下列不适合运用方差分析方法的资料的分布类型是?

- A. 近似正态分布, 且方差齐性的计量资料
- B. 严重偏态分布的计量资料
- C. 满足独立性、正态性和方差齐性的多组计量资料
- D. 经数据变换后满足正态性和方差齐性的计量资料

8. 在调查某疾病时, 调查了 200 个患者(男 90 个, 女 110 个), 能推断出的信息是?

- A. 男女患病率有差别
- B. 男女患病率没差别
- C. 无法判断男女患病率是否有差别
- D. 标准化后, 可判断男女患病率是否有差别

9. 下列哪一项不是表示相对数的统计指标?

- A. 构成比
- B. 率
- C. 相对比
- D. 标准差
- E. 优势比

10. 在研究血压(轻度, 中度, 重度)与血型(O,A,B,AB 型)的关系时, 若采用卡方检验来分析两者之间的关联性, 其自由度的计算结果为?

- A. 4
- B. 6
- C. 8
- D. 12

10. 比较同一学校内六年级小学生男女平均视力水平是否有差异, 可采用的统计方法是?

- A. 配对样本资料 t 检验
- B. 两独立样本资料 t 检验
- C. 秩和检验
- D. 卡方检验

11. 下列关于调查设计说法错误的是?

- A. 连续性变量必须采用开放式问题
- B. 调查设计中应明确调查目的、对象和范围, 否则可能导致调查结果不准确或无意义
- C. 抽样调查时, 样本量越大, 抽样误差不一定越小, 还需考虑抽样方法等因素
- D. 问卷设计时, 问题的排列顺序应先易后难, 先一般性问题后敏感性问题
- E. 预调查可以帮助发现调查问卷中存在的问题, 以便进行修改和完善

12. 在简单线性回归方程中, 每个 X 增大 10 倍, 回归系数会如何变化?

- A. 增大 10 倍
- B. 减小为原来的 1/10
- C. 不变
- D. 无法判断

13. 线性相关中, 当 $r=-1$ 时, 下列哪项说法正确

- A. 两变量呈完全正相关关系
- B. 两变量呈完全负相关关系
- C. 两变量无线性相关关系
- D. 两变量的关系无法确定

14. 若所有 Y 均可用 X 解释, 下列说法正确的是

- A. $SS_{\text{回归}}=SS_{\text{总}}$
- B. $SS_{\text{残差}}=SS_{\text{总}}$
- C. $SS_{\text{回归}}=SS_{\text{残差}}$
- D. $SS_{\text{回归}}=0$

15. 最适合描述某医院 2024 年心内科就诊患者的年龄分布的统计图是

- A. 直方图
- B. 饼图
- C. 百分条图
- D. 半对数线图
- E. 散点图

16. 比较两种药物对某一种疾病的治疗效果, 已知治疗效果用患者治疗后的某项生理指标值来衡量, 且该指标值服从正态分布, 两组方差齐, 可采用的检验方法是?

- A. 两独立样本 t 检验
- B. 卡方检验
- C. 秩和检验
- D. 回归分析

17. 为了探究三种不同药物对某疾病模型小鼠的治疗效果, 在实验安排上, 三种药物, 各设置 10 个重复组, 目的是()

- A. 使样本更具有代表性
- B. 便于统计计算
- C. 减小随机误差
- D. 以上都是
- E. 以上都不是

二. 判断题

1. 在定量资料统计中, 频率可以用来估计概率, 且样本量越大, 误差越大。
2. 统计中的定量变量, 分类变量, 等级变量之间均可以相互转换。
3. 定量资料的分析中, 描述分布类型不明确的资料的离散程度可以使用四分位数间距。
4. 某一地区成人体内总胆固醇水平的 95% 置信区间是 (180.46, 186.39), 则说明该地区胆固醇水平正常的人中有 95% 的人胆固醇水平位于此区间。
5. 小样本的两样本 t 检验需要满足两个总体正态分布, 方差齐。
6. 在样本含量不增加的情况下, 降低发生 I 类错误的概率, 则会提高发生 II 类错误的概率。
7. 标准化率可以表示某地区发病率的实际水平。
8. 对于患病率和某因素进行卡方检验, 结果表示有关联 ($P=0.001$), 则说明该因素一定是该疾病的致病因素。

9. 秩和检验和卡方检验都是非参数检验，均不需要满足正态分布。
10. 使用 Wilcoxon 符号秩和检验，若拒绝 H_0 ，则表示正，负秩和的差值具有统计学意义。
11. 回归系数假设检验中， $P < 0.001$ ，则说明两变量之间有较强的线性关系。
12. 简单线性回归中， $SS_{\text{回归}}=854$, $SS_{\text{残差}}=146$ ，则说明因变量的总变异中有 85.4% 可以由自变量解释。
13. 半对数线图纵坐标的起点不是“0”。
14. 用患者进行临床药物试验时，常选用空白对照。
15. 对于群体内个体的差异大，群体与群体之间的差异小的总体，选择整群抽样的误差一定小于简单随机抽样，且便于组织，节省人力，物力，财力。

三、分析题

1. 为研究三种抗癌药物对于小鼠肉瘤的抑癌效果，研究人员将 15 只小鼠分成 5 个区组，每个区组中的三只小鼠分别用三种不同的抗癌药物处理，观察一段时间后小鼠肉瘤的重量变化。

- (1) 该实验的实验设计类型是？请写出本研究设计的三要素。
- (2) 若实验数据满足参数检验的条件，请将下方的表格补充完整（不需要计算出具体的 P 值）
- (3) 请写出完整的假设检验的步骤，并说明在假设检验后是否还需要进行处理组之间的两两检验。

变异类型	自由度	SS	MS	F	P
总变异					
处理组间		0.169			
区组间			0.060		
误差		0.070			

2. 研究人员分别用 A, B 两种仪器对 40 位患者检测是否患有甲状腺结节，实验结果如下表所示。
- (1) 本研究的实验设计类型是什么？需要选用哪种统计方法进行统计检验？
 - (2) 请写出详细的检验过程。

		B 仪器		合计
A 仪器	有结节	有结节	无结节	
		8	3	11
有结节			24	29
无结节	5		27	40
合计	13			