

---

---

# 中国人民大学 2016 年攻读硕士学位研究生入学考试试题

考试科目：卫生统计 科目代码：613 考试时间： 月 日

(注：特别提醒所有答案一律写在答题纸上，直接写在试题或草稿纸上的无效!)

---

## 一、名词解释：

- |          |          |           |
|----------|----------|-----------|
| 1. 同质与变异 | 2. 总体和样本 | 3. 参数和统计量 |
| 4. 抽样误差  | 5. 概率    | 6. 计量资料   |
| 7. 计数资料  | 8. 等级资料  |           |

## 二、是非题：

1. 用定性的方法得到的资料称作数值变量资料，亦称作计数资料。 ( )
2. 观察某人群的血型，以人为观察单位，结果分为 A 型、B 型、AB 型和 O 型，是有序分类资料。 ( )
3. 分类变量或称定量变量，其变量值是定量的，表现为数值大小，一般有度量衡单位，亦称计量资料。 ( )
4. 调查某地 10 岁女孩的身体发育状况，以人为观察单位，测得每个人的身高 (cm)、体重 (kg)、血压 (mmHg)，此资料称为多项分类变量资料。 ( )
5. 用定量的方法测定观察单位某个量的大小的资料，称数值变量资料。 ( )
7. 等级分组资料兼有计数与计量资料的性质。 ( )
8. 将观察单位按某一属性的不同程度分组计数，所得各组的观察单位称为计数资料。 ( )

## 三、单选题：

1. 计量资料、计数资料和等级分组资料的关系有 \_\_\_\_\_  
A. 计量资料兼有计数资料和等级分组资料的一些性质  
B. 计数资料兼有计量资料和等级分组资料的一些性质  
C. 等级分组资料兼有计量资料和计数资料的一些性质  
D. 计数资料有计量资料的一些性质  
E. 等级分组资料又可叫半计数资料
2. 为了由样本推断总体，样本应该是 \_\_\_\_\_  
A. 总体中任意一部分  
B. 总体中的典型部分  
C. 总体中有意义的一部分  
D. 总体中价值的一部分  
E. 总体中有代表性的一部分
3. 统计学上所说的系统误差、测量误差和抽样误差三种误差，在实际工作中有 \_\_\_\_\_  
A. 三种误差都不可避免  
B. 系统误差和测量误差不可避免  
C. 系统误差和抽样误差不可避免  
D. 测量误差和抽样误差不可避免  
E. 三种误差都可避免
4. 抽样误差指的是 \_\_\_\_\_  
A. 个体值和总体参数值之差  
B. 个体值和样本统计量值之差

- 
- C. 样本统计量值和总体参数值之差    D. 总体参数值和总体参数值之差
5. 医学统计工作的基本步骤是\_\_\_\_\_
- A. 调查资料、校对资料、整理资料                      B. 调查资料、归纳资料、整理资料  
C. 收集资料、校对资料、整理资料                      D. 收集资料、整理资料、分析资料  
E. 收集资料、校对资料、归纳资料
6. 统计学中所说的总体是指\_\_\_\_\_
- A. 任意想象的研究对象的全体                      B. 根据研究目的确定的研究对象的全体  
C. 根据地区划分的研究对象的全体                      D. 根据时间划分的研究对象的全体  
E. 根据人群划分的研究对象的全体

## 答案

### 名词解释：

- (1) 同质与变异：同质指被研究指标的影响因素相同，变异指在同质的基础上各观察单位（或个体）之间的差异。
- (2) 总体和样本：总体是根据研究目的确定的同质观察单位的全体。样本是从总体中随机抽取的部分观察单位。
- (3) 参数和统计量：根据总体个体值统计计算出来的描述总体的特征量，称为总体参数，根据样本个体值统计计算出来的描述样本的特征量称为样本统计量。
- (4) 抽样误差：由抽样造成的样本统计量和总体参数的差别称为抽样误差。
- (5) 概率：是描述随机事件发生的可能性大小的数值，用  $p$  表示
- (6) 计量资料：由一群个体的变量值构成的资料称为计量资料。
- (7) 计数资料：由一群个体按定性因数或类别清点每类有多少个个体，称为计数资料。。
- (8) 等级资料：由一群个体按等级因数的级别清点每类有多少个体，称为等级资料。

### 是非题：

1. ×    2. ×    3. ×    4. ×    5. √    6. √    7. ×

### 单选题：

1. C    2. E    3. D    4. C    5. D    6. B



1. 平均数是一类用于推断数值变量及资料平均水平（或集中趋势）的指标。 ( )
2. 反映频数分布的两个重要特征是集中趋势与散离趋势。 ( )
3. 标准差是最常用的变异指标，它既可以用于正态资料亦可用于非正态资料。 ( )
4. 计算中位数时要求组距相等。 ( )
5. 计量单位相同，均数相差不大时，可使用变异系数反应两组变量值的离散程度。( )
6. 变量值之间呈倍数或等比关系的数据，宜用几何均数表示其平均水平。 ( )
7. 百分位数应用中提到，分布中部的百分位数相当稳定具有较好的代表性，但靠近两端的百分位数只在样本例数足够多时才比较稳定。 ( )
8. 为了解数值变量分布规律，可将观察值编制频数表，绘制频数分布图，用于描述资料的分布特征以及分布类型。 ( )
9. 如果少数几个数据比大部分数据大几百倍一般就不宜计算均数 ( )
10. 原始数据有零，就不能直接计算几何均数 ( )
11. 正态分布是以均数为中心的钟型分布 ( )
12. 高峰位于中央，两侧逐渐下降并完全对称的频数分布即为正态分布 ( )
13. 理论上，对于正态分布资料的  $P_5-P_{95}$  和  $\bar{X} \pm 1.96S$  范围内都包含有 95% 的变量值。 ( )
14. 制定正常值范围应选足够数量正常人作为调查对象，所谓正常人就是排影响被研究指标的各种疾病的人 ( )
15. 描述频数分布离散程度的最常用的指标是变异系数和标准差 ( )
16. 正态分布用  $N(0, 1)$  表示，为了应用方便，常对变量  $X$  作  $u = (x - \mu) / \sigma$  变换，使  $\mu = 0$   $\sigma = 1$  则将正态分布转换为标准正态分布用  $N(\mu, \sigma^2)$  表示 ( )
17. 只要单位相同，用标准差和用变异系数来比较两组变量值的离散度结论是完全一致的 ( )
18. 四分位数间距  $Q = P_{75} \sim P_{25}$ ，常用于描述近似正态分布资料的离散程度 ( )
19. 频数表和频数分布图用以推断变量值的分布特征和揭示变量值的分布规律 ( )
20. 制定正常值范围，如取 95% 界限，是指 95% 的正常人本项指标在此范围 ( )
21. 指标无论过高或过低均属异常，可用  $\bar{X} \pm us$  来计算正常值范围 ( )

#### 四、 单选题：

1. 以年龄(岁)为例，最常用的组段表示法是\_\_\_\_\_
  - A. 0—5, 5—10, 10—15, 15—20……;
  - B. 0— , 5— , 10— , 15— ……;
  - C. 0—4, 5—9, 10—14, 15—19……;
  - D. —5, —10, —15, —20……;
  - E. 以上都不是;
2. 以下指标中\_\_\_可用来描述计量资料离散程度。
 

A. 算术平均数	B. 几何均数	C. 中位数
D. 标准差	E. 第 50 百分位数	
3. 偏态分布资料宜用\_\_\_描述其分布的集中趋势。
 

A. 算术平均数	B. 标准差	C. 中位数
D. 四分位数间距	E. 方差	

- 
4. 用均数和标准差可全面描述\_\_\_资料的分布特征。  
 A. 正态分布                      B. 正偏态分布                      C. 负偏态分布  
 D. 对称分布                      E. 任何计量资料分布
5. \_\_\_可用于比较身高与体重的变异度  
 A. 方差                      B. 标准差                      C. 变异系数  
 D. 全距                      E. 四分位数间距
6. 各观察值均加(或减)同一个数后,\_\_\_\_\_。  
 A. 均数不变,标准差不一定变      B. 均数不变,标准差变  
 C. 均数不变,标准差也不变      D. 均数变,标准差不变  
 E. 均数变,标准差也变
7. 各观察值同乘以一个不等于0的常数后,\_\_\_\_\_不变。  
 A. 均数                      B. 标准差                      C. 几何均数  
 D. 中位数                      E. 变异系数
8. \_\_\_的资料,均数等于中位数。  
 A. 对称                      B. 正偏态  
 C. 负偏态                      D. 对数正态
9. 最小组段无下限或最大组段无上限的频数分布表资料,可用\_\_\_描述其集中趋势。  
 A. 均数                      B. 标准差                      C. 中位数  
 D. 四分位数间距                      E. 几何均数
10. 描述一组偏态分布资料的变异度,以\_\_\_\_\_指标较好。  
 A. 全距(R)                      B. 标准差(s)  
 C. 变异系数(CV)                      D. 四分位数间距 (Qu-QL)
11. 计算某抗体滴度的平均水平,一般选择\_\_\_\_\_。  
 A. 算术均数                      B. 几何均数  
 C. 中位数                      D. 标准差                      E. 标准误
12. 计算某病的平均潜伏期,一般选择 \_\_\_\_\_。  
 A. 算术均数                      B. 几何均数  
 C. 中位数                      D. 标准差                      E. 变异系数
13. 表示正态分布资料个体变量值的变异程度的常用指标是 \_\_\_\_\_。  
 A. 均数                      B. 全距  
 C. 标准差                      D. 标准误                      E. 变异系数
14. 平均数是表示\_\_\_\_\_。  
 A. 性质相同的变量值的相对水平      B. 性质相同的变量值的实际水平  
 C. 性质相同的变量值的平均水平      D. 性质不同的变量值的平均水平  
 E. 性质相同的变量值的变异程度
15. 用变异系数比较变异程度,适于\_\_\_\_\_。  
 A 两组观察值单位不同,或两均数相差较大  
 B 两组观察值单位相同,标准误相差较大  
 C 两均数相差较大,标准误相差较大      D 以上都不是
16. 正偏态资料计算平均水平,首选\_\_\_\_\_。  
 A. 算术均数                      B. 几何均数                      C. 中位数

D. 加权均数                      E. 百分位数

17. 均数与标准差的关系是\_\_\_\_\_

- A. 均数越大, 标准差越大                      B. 均数越大, 标准差越小  
C. 标准差越大, 均数代表性越好              D. 标准差越小, 均数代表性越差  
E. 标准差越小, 均数代表性越好

18. 有 8 名某传染病患者, 潜伏期分别为: 2, 1, 21, 7, 12, 1, 4, 13 天。其平均潜伏期为\_\_\_\_天。

- A. 4                      B. 5.5                      C. 7                      D. 12                      E. 9.5

19. 五小鼠出生体重分别为 4, 5, 6, 7, 8 (g); 染毒后存活日数分别为 2, 5, 6, 7, 1 (天), 问以何种指标比较两组数据变异大小为宜\_\_\_\_\_

- A. S                      B.  $S_{\bar{x}}$                       C. 全距                      D. CV                      E. 自由度

20. 调查 50 例链球菌咽峡炎患者潜伏期如下, 为计算均数平均数, 应首选\_\_\_\_\_

潜伏期	12—	24—	36—	48—	60—	72—	84—	96—	108—120	合计
病例数	1	7	11	11	7	5	4	2	2	50

- A. 算术均数                      B. 几何均数                      C. 中位数  
D. 百分位数                      E. 以上均可以

21. 102 名健康人钩端螺旋体血液抗体滴度分布如下, 欲表示其平均水平, 宜用\_\_\_\_\_

抗体滴度	1:100	1:200	1:400	1:800	1:1600	合计
人数	7	19	34	29	13	102

- A. 算术均数                      B. 几何均数                      C. 中位数  
D. 百分位数                      E. 以上均可以

22. 对于均数  $\mu$ 、标准差为  $\sigma$  的正态分布, 95% 的变量值分布范围为\_\_\_\_\_

- A.  $\mu - \sigma \sim \mu + \sigma$                       B.  $\mu - 1.96\sigma \sim \mu + 1.96\sigma$   
C.  $0 \sim \mu + 1.96\sigma$                       D.  $-\infty \sim \mu + 1.96\sigma$   
E.  $\mu - 2.58\sigma \sim \mu + 2.58\sigma$

23. 若  $u$  服从均数为 0, 标准差为 1 的正态分布, 则\_\_\_\_\_

- A.  $u \geq 2.58$  的  $P=0.01$                       B.  $u \geq 2.58$  的  $P=0.005$   
C.  $-2.58 < u < 2.58$  的  $P=0.01$               D.  $u \geq 2.58$  的  $P=0.05$   
E.  $u \geq 2.58$  的  $P=0.025$

24. 正态分布有两个参数  $\mu$  与  $\sigma$ , \_\_\_\_\_ 曲线的形状越扁平。

- A.  $\mu$  越大                      B.  $\mu$  越小                      C.  $\sigma$  越大  
D.  $\sigma$  越小                      E.  $\mu$  与  $\sigma$  越接近 0

25. 对数正态分布是一种\_\_\_\_\_分布

- A. 正态                      B. 近似正态                      C. 左偏态  
D. 右偏态                      E. 对称

26. 正态分布曲线下, 横轴上, 从均数  $\mu$  到  $+\infty$  的面积占总面积的比例为\_\_\_\_\_

- A. 97.5%                      B. 95%                      C. 50%  
D. 5%                      E. 不能确定 (与标准差的大小有关)

27. 标准正态分布的均数与标准差分别为\_\_\_\_\_

- A. 0 与 1                      B. 1 与 0                      C. 0 与 0  
D. 1 与 1                      E. 1.96 与 2.58

28. 若  $X$  服从以  $\mu, \sigma^2$  为均数和方差的正态分布, 则  $X$  的第 95 百分位数即\_\_\_\_\_
- A.  $\mu - 1.64\sigma$                       B.  $\mu - 1.96\sigma$                       C.  $\mu + \sigma$   
D.  $\mu + 1.64\sigma$                       E.  $\mu + 1.96\sigma$
29. 若正常成人的血铅含量  $X$  服从近似对数正态分布, 则可用公式\_\_\_\_\_制定 95% 正常值范围。(其中:  $Y = \log X$ )
- A.  $< \bar{X} + 1.96S$                       B.  $< \bar{X} + 1.64S$                       C.  $< \bar{Y} + 1.64S_Y$   
D.  $< \log^{-1}(\bar{Y} + 1.64S_Y)$                       E.  $< \log^{-1}(\bar{Y} + 1.96S_Y)$
30. 正态分布曲线下, 横轴上, 从均数  $\mu$  到  $\mu + 1.96$  倍标准差的面积为\_\_\_\_\_
- A. 95%                      B. 45%                      C. 97.5%                      D. 47.5%
31. 标准正态分布曲线下中间 90% 的面积所对应的横轴尺度  $u$  的范围是\_\_\_\_\_
- A.  $-1.645$  到  $+1.645$                       B.  $-\infty$  到  $+1.645$   
C.  $\infty$  到  $+1.282$                       D.  $-1.282$  到  $+1.282$
32. 设  $X$  符合均数为  $\mu$ 、标准差为  $\sigma$  的正态分布, 作  $u = \frac{x - \mu}{\sigma}$  的变量变换则\_\_\_\_\_
- A. 符合正态分布, 且均数不变                      B. 符合正态分布,  $\sigma$  且标准差不变  
C.  $u$  符合正态分布, 且均数和标准差都不变                      D.  $u$  不符合正态分布  
E.  $u$  符合正态分布, 但均数和标准差都改变
33. 正态分布是以\_\_\_\_\_
- A. 标准差为中心的频数分布                      B.  $t$  值为中心的频数分布  
C. 组距为中心的频数分布                      D. 均数为中心的频数分布  
E. 观察例数为中心的频数分布
34. 用变异系数比较变异程度, 适于\_\_\_\_\_
- A 两组观察值单位不同, 或两均数相差较大  
B 两组观察值单位相同, 标准误相差较大  
C 两均数相差较大, 标准误相差较大  
D 以上都不是
35. 决定个体值正态分布的参数是\_\_\_\_\_
- A. 变异系数                      B. 全距                      C. 标准误                      D. 标准差                      E. 以上都不是
36. 正态分布是以\_\_\_\_\_
- A. 标准差为中心的频数分布                      B.  $t$  值为中心的频数分布  
C. 组距为中心的频数分布                      D. 均数为中心的频数分布  
E. 观察例数为中心的频数分布
37. 正偏态资料计算平均水平, 首选\_\_\_\_\_
- A. 算术均数                      B. 几何均数                      C. 中位数                      D. 加权均数                      E. 百分位数
38. 均数与标准差的关系是\_\_\_\_\_
- A. 均数越大, 标准差越大                      B. 均数越大, 标准差越小  
C. 标准差越大, 均数代表性越好                      D. 标准差越小, 均数代表性越差  
E. 标准差越小, 均数代表性越好
39. 计量资料的标准差\_\_\_\_\_
- A. 不会比均数大                      B. 不会比均数小                      C. 要比标准误小  
D. 不决定于均数                      E. 以上都不对
40. 有 9 名某传染病病人, 潜伏期分别为 (天): 2, 1, 21, 7, 12, 1, 4, 13, 24 其平均潜伏期为\_\_\_\_\_
- A. 4                      B. 5.5                      C. 7                      D. 12                      E. 9.5

41. 表示变异程度的指标中\_\_\_\_\_
- A. 标准差越大, 变异程度越小  
 B. 标准差越小, 变异程度越大  
 C. 变异系数越大, 变异程度越大  
 D. 变异系数越大, 变异程度越小  
 E. 全距越大, 变异程度越小
42. 正态资料的变异系数应\_\_\_\_\_
- A. 一定 $>1$       B. 一定 $<1$       C. 可能 $>1$ , 也可能 $<1$   
 D. 一定 $<$ 标准差      E. 一定 $>$ 标准差
43. 五小鼠出生体重分别为4, 5, 6, 7, 8 (g); 染毒后存活日数分别为2, 5, 6, 7, 1 (天), 问以何种方式说明两组数据变异大小\_\_\_\_\_
- A. S      B.  $S_{\bar{x}}$       C. 全距      D. CV      E. 自由度
44. 调查50例链球菌咽峡炎患者潜伏期如下

潜伏期	12—24—36—48—60—72—84—96—108—120	合计
病例数	1    7    11    11    7    5    4    2    2	50

- 求平均潜伏期, 应首选\_\_\_\_\_
- A. 算术均数      B. 几何均数      C. 中位数  
 D. 百分位数      E. 以上均可以
45. 102名健康人钩端螺旋体血液抗体滴度分布如下, 欲表示其平均水平, 宜用\_\_\_\_\_

抗体滴度	1:100	1:200	1:400	1:800	1:1600	合计
人数	7	19	34	29	13	102

- A. 算术均数    B. 几何均数    C. 中位数    D. 百分位数    E. 以上均可以

## 五、问答题:

- 均数、几何均数和中位数的适用范围有何异同?
- 中位数与百分位数在意义上、计算和应用上有何区别与联系?
- 同一资料的标准差是否一定小于均数?
- 测得一组资料, 如身高或体重等, 从统计上讲, 影响其标准差大小的因素有哪些?
- 正态分布、标准正态分布与对数正态分布在概念上和应用上有何异同?
- 医学中参考值范围的含义是什么? 确定的原则和方法是什么?
- 对称分布资料在“均数 $\pm 1.96$ 倍标准差”的范围内, 也包括95%的观察值吗?

## 六、计算题

- 某地101例30~49岁健康男子血清总胆固醇值 (mmol/L) 测定结果如下:
 

4.77	3.37	6.14	3.95	3.56	4.23	4.31	4.71	5.69	4.12	4.56	4.37
5.39	6.30	5.21	7.22	5.54	3.93	5.21	6.51	5.18	5.77	4.79	5.12
5.20	5.10	4.70	4.07	3.50	4.69	4.38	4.89	6.25	5.32	4.50	4.63
3.61	4.44	4.43	4.25	4.03	5.85	4.09	3.35	4.08	4.79	5.30	4.97
3.18	3.97	5.16	5.10	5.86	4.79	5.34	4.24	4.32	4.77	6.36	6.38
4.86	5.55	3.04	4.55	3.35	4.87	4.17	5.85	5.16	5.09	4.52	4.38
4.31	4.58	5.72	6.55	4.76	4.61	4.17	4.03	4.47	3.04	3.91	2.70

---

---

4.60 4.09 5.96 5.48 4.40 4.55 5.38 3.89 4.60 4.47 3.64  
4.34 5.18 6.14 3.24 4.90 3.05

- (1) 编制频数分布表并绘制直方图，简述其分布特征。
  - (2) 计算均数  $\bar{X}$ 、标准  $s$ 、变异系数  $CV$ 。
  - (3) 计算中位数  $M$ ，并与均数  $\bar{X}$  比较，
  - (4) 计算  $P_{2.5}$  及  $P_{97.5}$  并与  $\bar{X} \pm 1.96s$  的范围比较。
  - (5) 分别考察  $\bar{X} \pm 1S$ 、 $\bar{X} \pm 1.96S$ 、 $\bar{X} \pm 2.58S$  范围内的实际频数与理论分布是否基本一致？
  - (6) 现测得一 40 岁男子的血清总胆固醇值为 6.993 (mmol/L)，若按 95% 正常值范围估计，其血清总胆固醇值是否正常？估计该地 30~49 岁健康男子中，还有百分之几的人血清总胆固醇值比他高？
2. 某地卫生防疫站，对 30 名麻疹易感儿童经气溶胶免疫一个月后，测得其得血凝抑制抗体滴度资料如表。

---

表 2-1: 平均滴度计算表

---

抗体滴度	人数 $f$
1: 8	2
1: 16	6
1: 32	5
1: 64	10
1: 128	4
1: 256	2
1: 512	1
合计	30

---

- (1) 试计算其平均滴度。
  - (2) 有人发现本例用抗体滴度稀释倍数和直接用滴度倒数算得几何标准差的对数值相同，为什么？
3. 50 例链球菌咽峡炎患者的潜伏期如表，说明用均数、中位数或几何均数，何者的代表性较好？并作计算。

---

表 2-2: 50 例链球菌 咽峡炎患者的潜伏期

---

潜伏期(小时)	病例数 f
12~	1
24~	7
36~	11
48~	11
60~	7
72~	5
84~	4
96~	2
108~120	2
合计	50

4. 某市 1974 年为了解该地居民发汞的基础水平, 为汞污染的环境监测积累资料, 调查了留住该市一年以上, 无明显肝、肾疾病, 无汞作业接触史的居民 238 人, 发汞含量如表:

表 2-3: 238 人发汞含量频数计算表

发汞值 ( $\mu\text{mol/kg}$ )	人数 f
1.5~	20
3.5~	66
5.5~	60
7.5~	48
9.5~	18
11.5~	16
13.5~	6
15.5~	1
17.5~	0
19.5~21.5	3
合计	238

- (1). 说明此频数分布的特征,
- (2). 计算均数和中位数, 何者较大? 为什么? 何者用于说明本资料的集中位置较合适?
- (3). 选用何种指标描述其离散程度较好?
- (4). 估计该地居民发汞值的 95%参考值范围

## 答案

### 名词解释:

1. 平均数 是描述数据分布集中趋势 (中心位置) 和平均水平的指标
2. 标准差 是描述数据分布离散程度 (或变量变化的变异程度) 的指标

3. 标准正态分布 以  $\mu$  服从均数为 0、标准差为 1 的正态分布, 这种正态分布称为标准状态分布。
4. 参考值范围 参考值范围也称正常值范围, 医学上常把把绝大多数的某指标范围称为指标的正常值范围。

**填空题:**

1. 计量, 计数, 等级
2. 设计, 收集资料, 分析资料, 整理资料。
3.  $u = \frac{x - \mu}{\sigma}$  (变量变换) 标准正态分布、0、1
4.  $\pm \sigma$      $\pm 1.96\sigma$      $\pm 2.58\sigma$     68.27%    95%    99%
5. 47.5%
6. 均数、标准差
7. 全距、方差、标准差、变异系数
8.  $\mu \pm 1.96\sigma$      $\mu \pm 2.58\sigma$
9. 全距    R
10. 检验水准、显著性水准、0.05、0.01 (0.1)
11. 80%    90%    95%    99%    95%
12. 95%    99%
13. 集中趋势、离散趋势
14. 中位数
15. 同质基础, 合理分组
16. 均数, 均数,  $\mu$ ,  $\sigma$ , 规律性
17. 标准差
18. 单位不同, 均数相差较大

**是非题:**

1. ×    2. √    3. ×    4. ×    5. ×    6. √    7. √    8. √    9. √    10. √  
 11. √    12. √    13. ×    14. √    15. √    16. ×    17. ×    18. ×    19. √    20. √  
 21. √

**单选题:**

1. B    2. D    3. C    4. A    5. C    6. D    7. E    8. A    9. C    10. D  
 11. B    12. C    13. C    14. C    15. A    16. C    17. E    18. C    19. D    20. C  
 21. B    22. B    23. E    24. C    25. A    26. C    27. B    28. D    29. D    30. D  
 31. A    32. E    33. D    34. A    35. D    36. D    37. C    38. E    39. D    40. B  
 41. C    42. B    43. D    44. C    45. B

**问答题:**

1. 均数、几何均数和中位数的适用范围有何异同?  
 答: 相同点, 均表示计量资料集中趋势的指标。  
 不同点: 表 2-5.

**表 2-5 均数, 几何均数和中位数的相异点**

平均数	意义	应用场合
均数	平均数量水平	应用甚广, 最适用于对称分布, 特别是正态分布
几何均数	平均增减倍数	①等比资料; ②对数正态分布资料
中位数	位次居中的观	①偏态资料; ②分布不明资料; ③分布一端或两

2. 中位数与百分位数在意义上、计算和应用上有何区别与联系？

答：

1) 意义：中位数是百分位中的第 50 分位数，常用于描述偏态分布资料的集中位置，反映位次居中的观察值水平。百分位数是用于描述样本或总体观察值序列在某百分位置的水平，最常用的百分位是  $P_{50}$  即中位数。多个百分位数结合使用，可更全面地描述总体或样本的分布特征。

(2) 计算：中位数和百分位数均可用同一公式计算，即

$$P_x = L + (i/f_x)(n \cdot x\% - \sum f_l)$$

可根据研究目的选择不同的百分位数代入公式进行计算分析。

(3) 应用：中位数常用于描述偏态分布资料的集中趋势；百分位数常用于医学参考值范围的确定。中位数常和其它分位数结合起来描述分布的特征，在实际工作中更为常用。百分位数还可以用来描述变量值的离散趋势（四分位数间距）。

3. 同一资料的标准差是否一定小于均数？

答：不一定。同一资料的标准差的大小与均数无关，主要与本资料的变异度有关。变异大，标准差就大，有时比均数大；变异小，标准差小。

4. 测得一组资料，如身高或体重等，从统计上讲，影响其标准差大小的因素有哪些？

(1) 样本含量的大小，样本含量越大，标准差越稳定。

(2) 分组的多少

(3) 分布形状的影响，偏态分布的标准差较近似正态分布大

(4) 随机测量误差大小的影响

(5) 研究总体中观察值之间变异程度大小

5. 正态分布、标准正态分布与对数正态分布在概念上和应用上有何异同？

(1) 概念上：①相同点：正态分布、标准正态分布与对数正态分布都是变量的连续型分布。其特征是：分布曲线在横轴上方，略呈钟型，以均数为中心，两边对称，均数处最高，两边逐渐减小，向外延伸，不与横轴相交。②相异点：表示方法不同，正态分布用  $N(\mu, \sigma^2)$  表示，标准正态分布用  $N(0, 1)$  表示，对数正态分布  $N(\mu_{lgx}, \sigma_{lgx}^2)$  表示。

(2) 应用上：①相同点：正态分布、对数正态分布都可以转换为标准正态分布。

②相异点：标准正态分布是标准正态变量  $u$  的分布，标准正态曲线下的面积唯一的由  $u$  决定，给应用带来极大方便。对医学资料呈偏态分布的数据，有的经对数变换后服从正态分布。正态分布、对数正态分布可描述变量值的分布特征，可用于正常值范围估计和质量控制等。正态分布是很多统计方法的理论基础。

6. 医学中参考值范围的含义是什么？确定的原则和方法是什么？

含义：参考值范围亦称正常值范围，它是指特定健康状况人群（排除了有关疾病和因素对所研究指标有影响的所谓“正常人”不同于“健康人”概念）的解剖、生理、生化等数据绝大多数人的波动范围。

(2) 原则：

① 抽取有代表性的足够例数的正常人群样本，样本分布越接近总体，所得结果越可靠。一般认为样本含量最好在 100 例以上，以能得到一个分布较为稳定的样本为原则。

② 对选定的正常人进行准确而统一的测定，保证测定数据可靠是确定正常值范围的前提。

③ 判定是否要分组（如男女、年龄、地区等）确定正常值范围。

④ 决定取双侧范围值还是单侧范围值。

⑤ 选择适当的百分范围

- ⑥ 确定可疑范围
- ⑦ 估计界值
- (3) 方法:
  - ① 百分位数法:  $P_x = L + (i/f_x)(n \cdot x\% - \sum f_l)$
  - ② 正态分布法 (对数正态分布):

双侧

$$\bar{X} \pm u_{\alpha} S$$

$$\lg^{-1}(\bar{X}_{\lg X} \pm u_{\alpha} S_{\lg X})$$

百分位数法用于各种分布型 (或分布不明) 资料; 正态分布法用于服从或近似正态分布 (服从对数正态分布) 的资料。

7. 对称分布资料在“均数±1.96 倍标准差”的范围内, 也包括 95% 的观察值吗?

答: 不一定。均数±1.96 倍标准差是正态分布的分布规律, 对称分布不一定是正态分布。

### 计算题:

1. 某地 101 例 30~49 岁健康男子血清总胆固醇值 (mmol/L) 测定结果如下:

4.77 3.37 6.14 3.95 3.56 4.23 4.31 4.71 5.69 4.12  
 4.56 4.37 5.39 6.30 5.21 7.22 5.54 3.93 5.21 6.51  
 5.18 5.77 4.79 5.12 5.20 5.10 4.70 4.074 3.50 4.69  
 4.38 4.89 6.25 5.32 4.50 4.63 3.61 4.44 4.43 4.25  
 4.03 5.85 4.09 3.35 4.08 4.79 5.30 4.97 3.18 3.97  
 5.16 5.10 5.86 4.79 5.34 4.24 4.32 4.77 6.36 6.38  
 4.86 5.55 3.04 4.55 3.35 4.87 4.17 5.85 5.16 5.09  
 4.52 4.38 4.31 4.58 5.72 6.55 4.76 4.61 4.17 4.03  
 4.47 3.04 3.91 2.70 4.60 4.09 5.96 5.48 4.40 4.55  
 5.38 3.89 4.60 4.47 3.64 4.34 5.18 6.14 3.24 4.90  
 3.05

(1) 编制频数分布表, 简述其分布特征。

- ① 找出最大值、最小值求全距 (R):  
全距=最大值-最小值=7.22-2.70=4.50 (mmol/L)
- ② 求组距:  $I = \text{全距} / \text{组数} = 4.52 / 10 = 0.452 \approx 0.5$  (mmol/L)
- ③ 分组段, 划记 (表 1-1)

表 2-6 某地 101 例 30~49 岁健康男子血清总胆固醇值划记表

组段 (mmol/L)	划记	频数
2.5~		1

3.0~	8
3.5~	9
4.0~	23
4.5~	25
5.0~	17
5.5~	9
6.0~	6
6.5~	2
7.0~7.5	1
合计	101

由表 2-6 可知，本例频数分布中间局多，两侧逐渐减少，左右基本对称。

表 2-7 某地 101 例 30~49 岁健康男子血清总胆固醇值 (mmol/L)  $\bar{X}$ 、s 计算表

血清总胆 固醇值	组中值 X	频数 f	fX	fX <sup>2</sup>	累计 频数	累计频数 (实际)
2.5~	2.75	1	2.75	7.563	1	0.0099
3.0~	3.25	8	26.00	84.500	9	0.0891
3.5~	3.75	9	33.75	126.563	18	0.1782
4.0~	4.25	23	97.75	415.438	41	0.4059
4.5~	4.75	25	118.75	564.063	66	0.6535
5.0~	5.25	17	89.25	468.563	83	0.8218
5.5~	5.75	9	51.75	297.563	92	0.9109
6.0~	6.25	6	37.50	234.375	98	0.9703
6.5~	6.75	2	13.50	91.125	100	0.9901
7.0~7.5	7.25	1	7.25	52.563	101	1.0000
			478.25	2242.315		

注：Xu 为组段上限值

(2) 计算均数  $\bar{X}$ 、标准 s、变异系数 CV。

由上计算表 1-2 可见： $\bar{X} = \Sigma fX / \Sigma f = 478.25 / 101 = 4.735$  (mmol/L)

$$s = \frac{\sqrt{\sum fX^2 - (\sum fX)^2 / \sum f}}{\sqrt{\sum f - 1}}$$

$$= \frac{\sqrt{2342.313 - (478.25)^2 / 101}}{\sqrt{101 - 1}} = 0.882 \text{ (mmol/L)}$$

$$CV = s / \bar{x} \times 100\% = 0.882 / 4.735 \times 100\% = 18.627\%$$

(3) 计算中位数 M, 并与均数  $\bar{X}$  比较, 利用前表计算中位数 M

$$M = L + (i/f_{50}) (n \times 50\% - \sum f_L)$$

$$= 4.5 + (0.5/25) (101 \times 50\% - 41) = 4.69 \text{ (mmol/L)}$$

本题算术均数为 4.735 (mmol/L), 与中位数 4.69 (mmol/L) 很接近, 这也是资料服从正态分布的特征之一。

(4) 计算 P2.5 及 P97.5 并与  $\bar{X} \pm 1.96s$  的范围比较。

$$P_{2.5} = 3.0 + (0.5/8) \times (101 \times 2.5\% - 1) = 3.095 \text{ (mmol/L)}$$

$$P_{97.5} = 6.5 + (0.5/2) \times (101 \times 97.5\% - 98) = 6.619 \text{ (mmol/L)}$$

$$\bar{X} \pm 1.96s = 4.735 \pm 1.96 \times 0.882 = 3.01 \sim 6.46 \text{ (mmol/L)}$$

用百分位数法求得 101 例 30~49 岁健康男子血清总胆固醇值 95% 分布范围 3.095~6.619 (mmol/L), 与正态分布法求得的 95% 分布范围 3.01~6.46 (mmol/L) 基本一致。

(5) 分别考察  $\bar{X} \pm 1s$ 、 $\bar{X} \pm 1.96s$ 、 $\bar{X} \pm 2.58s$  范围内的实际频数与理论分布是否基本一致 (表 1-3)

表 2-8 某地 101 例 30~49 岁健康男子血清总胆固醇值理论分布与实际分布比较

$\bar{X} \pm us$	血清总胆固醇	实际分布		理论分布
		人数	%	%
$\bar{X} \pm 1s$	3.85~5.62	72	71.29	68.27
$\bar{X} \pm 1.96s$	3.01~6.46	97	96.04	95.00
$\bar{X} \pm 2.58s$	2.46~7.01	100	99.01	99.00

由上表,  $\bar{X} \pm 1s$  范围内, 实际分布与理论分布略有不同, 而  $\bar{X} \pm 1.96s$ 、 $\bar{X} \pm 2.58s$  范围内, 实际分布与理论分布基本一致。

(6) 现测得一 40 岁男子的血清总胆固醇值为 6.993 (mmol/L), 若按 95% 正常值范围估计, 其血清总胆固醇值是否正常? 估计该地 30~49 岁健康男子中, 还有百分之几的人血清总胆固醇值比他高?

前计算得 95% 正常值为 3.01~6.46 (mmol/L) 现测得一 40 岁男子的血清总胆固醇值为 6.993 (mmol/L), 在 95% 范围以外, 故属于异常

$$u = (X - \mu) / \sigma = (6.993 - 4.735) / 0.882 = 2.56$$

因  $\Phi(2.56) = \Phi(-2.56)$ , 查表 1 得  $\Phi(-2.56) = 0.0052$

估计该地 30~49 健康男子中约有 0.52% 的人血清总胆固醇值比他高。

2. 某地卫生防疫站, 对 30 名麻疹易感儿童经气溶胶免疫一个月后, 测得其得血凝抑制抗体滴度资料如表 2-9 第 (1) (2) 栏。

表 2-9 平均滴度计算表

抗体滴度 (1)	人数 f (2)	滴度倒数 $X_1$ (3)	$\lg X_1$ (4)	$f \lg X_1$ (5) = (2) × (4)
1: 8	2	8	0.9031	1.8062
1: 16	6	16	1.2041	7.2247
1: 32	5	32	1.5051	7.5257
1: 64	10	64	1.8062	18.0618
1: 128	4	128	2.1072	8.4288
1: 256	2	256	2.4082	4.8165
1: 512	1	512	2.7093	2.7093
合计	30			50.5730

(1) 试计算其平均滴度。

由表 1-4 得,  $G = 1g^{-1}(50.5730/30) = 1g^{-1}1.6858 = 48.5$

该站 30 名麻疹易感儿童经气溶胶免疫一个月后, 测得血凝抑制抗体平均滴度为 1: 48.50

表 2-10 平均滴度计算表

抗体滴度 (1)	人数 f (2)	滴度倒数 $X_1$ (3)	$\lg X_1$ (4)	$f \lg X_1$ (5) = (2) × (4)
1: 8	2	8	0.9031	1.8062
1: 16	6	16	1.2041	7.2247
1: 32	5	32	1.5051	7.5257
1: 64	10	64	1.8062	18.0618
1: 128	4	128	2.1072	8.4288
1: 256	2	256	2.4082	4.8165
1: 512	1	512	2.7093	2.7093
合计	30			50.5730

(2) 有人发现本例用抗体滴度稀释倍数和直接用滴度 (原书误为倒数) 算得对数值的标准差相同, 为什么?

表 2-11 滴度对数值计算表

抗体滴度 $X_2$	人数 f	$\lg X_2$	$f \lg X_2$
1: 8	2	-0.9031	-1.8062
1: 16	6	-1.2041	-7.2247
1: 32	5	-1.5051	-7.5257
1: 64	10	-1.8062	-18.0618
1: 128	4	-2.1072	-8.4288
1: 256	2	-2.4082	-4.8165
1: 512	1	-2.7093	-2.7093

合计	30	-50.5730
----	----	----------

1) 由表 1-4 中数据计算标准差为:  $s\lg x_1 = \lg^{-1} 0.4444 = 2.7823$

2) 由表 1-5 中数据计算标准差为:  $s\lg x_2 = \lg^{-1} 0.4444 = 2.7823$

直接用抗体滴度的对数  $\lg x_2$  与稀释倍数的对数  $\lg x_1$  计算标准差是相等的, 因为由上表可见  $\lg x_2 = \lg 1 - \lg X_1 = -\lg x_1$ , 而  $\lg x_1$  与  $-\lg x_1$  的离散程度是相同的, 所以用抗体滴度稀释倍数和直接用滴度算得对数值的标准差是相同的。

3. 50 例链球菌咽峡炎患者的潜伏期如表 2-12, 说明用均数、中位数或几何均数, 何者的代表性较好? 并作计算。

**表 2-12 50 例链球菌咽峡炎患者的潜伏期的中位数计算表**

潜伏期(小时)	病例数 f	累计频数
12~	1	1
24~	7	8
36~	11	19
48~	11	30
60~	7	
72~	5	
84~	4	
96~	2	
108~120	2	
合计	50	

本例目测频数分布为偏态分布, 长尾拖向右侧, 故为正偏态, 宜用中位数及几何均数表示其平均水平。

如上表, 经计算中位数, 几何均数、算术均数分别为:

$M=54.55$  (小时),  $G=54.08$  (小时),  $\bar{X}=58.56$  (小时) 显然, 算术均数受长潜

伏期的影响使其偏大, 中位数  $M$  与几何均数  $G$  接近, 故描述链球菌咽峡炎患者潜伏期的集中趋势指标使用中位数  $M$  或几何均数  $G$  均可。

4. 某市 1974 年为了解该地居民发汞的基础水平, 为汞污染的环境监测积累资料, 调查了留住该市一年以上, 无明显肝、肾疾病, 无汞作业接触史的居民 238 人, 发汞含量如表 2-13:

**表 2-13 238 人发汞含量频数计算表**

发汞值 ( $\mu\text{mol/kg}$ )	人数 f	组中值 X	fX	fX <sup>2</sup>	累计频数	累计频率
1.5~	20	2.5	50.0	125.00	20	8.40
3.5~	66	4.5	297.0	1336.50	86	36.10
5.5~	60	6.5	390.0	2535.00	146	61.34
7.5~	48	8.5	408.0	3468.00	194	81.50
9.5~	18	10.5	189.0	1984.50	212	89.08
11.5~	16	12.5	200.0	2500.00	228	95.80
13.5~	6	14.5	87.0	1261.50	234	98.32
15.5~	1	16.5	16.5	272.25	235	98.74
17.5~	0	18.5	0.0	0.00	235	98.74
19.5~21.5	3	20.5	61.5	1260.75	238	100.00

---

---

合计	238	1699.0	14743.50
----	-----	--------	----------

---

- (1). 说明此频数分布的特征: 可见发汞值的频数分布高峰位于第 2 个组段。前 4 个组段的频数占总频数的 81.5%, 长尾拖向右侧, 呈极度正偏态。
- (2). 计算均数  $\bar{X}$  和中位数  $M$ , 何者较大? 为什么? 何者用语说明本资料的集中位置较合适?

$$\bar{X} = \Sigma fX / \Sigma f = 1699 / 238 = 7.139 \text{ (} \mu \text{ mol/kg)}$$

$$M = L + (i/f_{50}) (n \times 50\% - \Sigma f_L) \\ = 5.5 + 2/60 (238 \times 50\% - 86) = 6.6 \text{ (} \mu \text{ mol/kg)}$$

由计算结果得知,  $\bar{X} > M$  其原因因为本例呈正态分布, 均数计算结果受到少数较大发汞值的影响, 使得  $\bar{X}$  偏向大发汞值一边. 本例用中位数描述偏态资料的集中趋势较好, 它不受两端较大值和极小值的影响.

- (3). 选用何种指标描述其离散程度较好?  
选用四分位数间距描述其离散程度较好.
- (4). 估计该地居民发汞值的 95% 参考值范围  
本资料应选用单侧 95% 上界值, 本例是正偏态分布. 而且样本含量较大,  $n=238$ , 保证获得一个较为稳定的分布, 故采用百分位数法计算的参考值范围较为合适.

$$P_{95} = L + (i/f_{95}) (n \times 95\% - \Sigma f_L) \\ = 11.5 + (2/16) (238 \times 95\% - 212) = 13.2625 \text{ (} \mu \text{ mol/kg)}$$

考试科目：卫生统计 科目代码：613 考试时间： 月 日

(注：特别提醒所有答案一律写在答题纸上，直接写在试题或草稿纸上的无效!)

### 一、 名词解释：

- |            |             |         |
|------------|-------------|---------|
| 1. 标准误     | 2. 总体均数可信区间 | 3. 假设检验 |
| 4. 检验水准    | 5. 检验统计量    | 6. P 值  |
| 7. 第 I 类错误 | 8. 第 II 类错误 |         |

### 二、 填空题：

1. 用样本均数估计总体均数的可靠程度宜选用\_\_\_\_\_。
2. 一般将概率小于\_\_\_\_\_和\_\_\_\_\_的随机事件称为小概率事件。
3. 对总体所作的假设进行统计推断, 作出拒绝或接受假设的结论的方法, 称\_\_\_\_\_。
4. 样本均数比较, 经 t 检验差别有显著性时, P 越小, 说明\_\_\_\_\_。
5. t 分布的图型形状与\_\_\_\_\_有关。
6. 可信区间的两个要素是\_\_\_\_\_，\_\_\_\_\_。
7. 准误差愈小, 表示\_\_\_\_\_愈小, \_\_\_\_\_对\_\_\_\_\_估计愈可靠。
8. 统计推断包括\_\_\_\_\_和\_\_\_\_\_两方面。
9. 设检验中, 不拒绝  $H_0$  时, 可发生\_\_\_\_\_错误, 其概率用\_\_\_\_\_表示。

### 三、 是非题：

1. 区间估计以预先给定的概率 (可信度) 估计总体参数在哪个范围内的方法, 称区间估计。 ( )
2. 样本标准误  $S_{\bar{x}} = \frac{S}{\sqrt{n}}$  反映了抽样误差大小, 显然 n 越大抽样误差越大, 用样本推断总体的精度越高。 ( )
3. 抽样误差的大小可以用标准差来表示。 ( )
4. 参数估计有(1)点 (值) 估计——用样本统计量值估计相应的总体参数。(2)区间估计——估计总体参数在哪个范围, 它不涉及抽样误差, 所以比点 (值) 估计更为重要。 ( )
5. 从正态总体  $N(\mu, \sigma)$  中, 随机抽取例数为 n 的样本, 则样本均数  $\bar{X}$  也服从  $N(\mu, \sigma)$  的正态分布。 ( )
6. 抽样研究时, 可通过增加样本含量来减少抽样误差。 ( )
7. 成组比较的 t 检验要求两组样本例数一定相等。 ( )
8. 可信区间比假设检验还可以提供更多信息, 不但能回答差别有无统计意义, 还能提出差别有无实际意义。 ( )
9. t 检验结果  $t > 1.96$ , 可以认为两样本均数不同。 ( )
10. 样本含量相同时, 配对设计与成组设计相比, 前者统计效率较高 ( )
11. 在配对 t 检验中, 用药前数据减去用药后数据, 与用药后数据减去用药前数据作 t 检验后的结论是相同的 ( )
12. t 分布曲线的形状与标准差有关 ( )
13. 拒绝了实际上是成立的  $H_0$ , 这类“弃真”的错误称为第一类错误或 I 型错误

- ( )
14. 抽样调查是从总体中随机抽取一定数量的观察单位组成样本, 然后用样本信息来推断总体特征 ( )
15. 变量变换是各组观察值经变量转换后达到方差齐性 ( )
16. 样本均数的 99% 可信区间可用  $\bar{X} \pm 2.58 S_{\bar{x}}$  表示 ( )
17. 计算总体均数可信区间的通式为  $(\bar{X} - t_{\alpha} \cdot v \cdot S_{\bar{x}}, \bar{x} + t_{\alpha} v S_{\bar{x}})$   
缩写为  $\bar{X} \pm t_{\alpha} v S_{\bar{x}}$  ( )
18. 当  $P \leq \alpha$  时, 结论为按所取  $\alpha$  检验水准拒绝  $H_0$ , 接受  $H_1$ ; 如  $P > \alpha$  时, 即样本信息支持  $H_0$ , 就无理由拒绝它, 此时只好接受它 ( )
19. 当  $t_{0.01v} > t > t_{0.05v}$  时, 则  $0.01 > P > 0.05$  ( )
20. 两个均数比较只能用 t 检验或 u 检验 ( )
21. t 检验可用于同一批对象的身高和体重均数差别的假设检验 ( )
22.  $\beta$  为第二类错误的概率  $(1 - \beta)$  越小, 所需的样本例数越多 ( )

#### 四、 单选题:

1. 从一个计量资料的总体中抽样, 产生的抽样误差的原因是: ( )
- A. 总体中的个体值存在差异      B. 总体均数不等于零  
C. 样本中的个体值存在差异      D. 样本均数不等于零  
E. 样本只包含总体的一部分个体
2.  $\sigma_{\bar{x}}$  是指: ( )
- A. 所有个体值对总体均数的离散程度    B. 某一个样本均数对总体均数的离散程度  
C. 所有样本均数对总体均数的离散程度    D. 一些样本均数对总体均数的离散程度  
E. 所有某个含量相同的样本均数对总体均数的离散程度
3. 在同一正态总体中随机抽样, 有 99% 的样本均数在下述范围内: ( )
- A.  $\bar{x} \pm 2.58 s_{\bar{x}}$       B.  $\mu \pm 2.58 s_{\bar{x}}$       C.  $\bar{x} \pm 2.58 \sigma_{\bar{x}}$   
D.  $\mu \pm 2.58 \sigma_{\bar{x}}$       E. 以上都不是
4. 在同一总体中随机抽取多个样本, 用样本来估计总体均数的 95% 可信区间, 则估计精密的是: ( )
- A. 均数大的样本      B. 均数小的样本      C. 标准差小的样本  
D. 标准误大的样本      E. 标准误小的样本
5.  $\bar{X} \pm t_{0.05}(v) S_{\bar{x}} = \bar{x} - t_{0.05}(v) S_{\bar{x}} \sim \bar{x} + t_{0.05}(v) S_{\bar{x}}$  表示: ( )
- A. 总体的 95% 个体值在该区间内  
B. 样本的 95% 个体值在该区间内  
C. 平均每 100 个总体均数, 有 95 个总体均数在该区间内  
D. 平均每 100 个样本 (含量相同) 均数, 有 95 个样本均数在该区间内  
E. 平均每 100 个样本 (含量相同) 有 95 个样本所得出的该区间包括总体均数
6. 总体均数的 99% 可信区间为: ( )
- A.  $\bar{x} \pm 1.96 s$       B.  $\bar{x} \pm 1.96 s_{\bar{x}}$       C.  $\bar{x} \pm t_{0.05}(v) s$

D.  $\bar{x} \pm t_{0.05}(\nu)s_{\bar{x}}$  E. 以上都不是

7. 在由两样本均数的差别推断两总体均数的差别的 t 检验中, 检验假设的无效假设是: ( )
- A. 两样本均数差别无统计意义 B. 两总体均数差别无统计意义  
C. 两样本均数相等 D. 两总体均数相等  
E. 两总体均数不等
8. 由两样本均数的差别推断两总体均数的差别, 所谓差别有显著性是指: ( )
- A. 两样本均数差别有显著性 B. 两总体均数差别有显著性  
C. 两样本均数和两总体均数的差别都有显著性  
D. 其中一个样本均数和总体均数的差别有显著性 E. 以上都不是
9. 在样本均数和总体均数差别的显著性检验中,  $H_0$  (无效假设):  $\mu = \mu_0$ ;  $H_1$  (备择假设):  $\mu \neq \mu_0$ ; 结果因为  $P < 0.05$  而拒绝  $H_0$  接受  $H_1$ , 是由于: ( )
- A. 无效假设成立的可能性小于 5% B. 备择假设成立的可能性大于 5%  
C. 无效假设成立的可能性小于 5% 和备择假设成立的可能性大于 95%  
D. 该样本来自该总体 ( $\mu = \mu_0$ ) 的可能性小于 5%  
E. 该样本来自另一个总体 ( $\mu \neq \mu_0$ ) 的可能性大于 95%
10. 与标准正态分布 (u 分布) 比较, t 分布的: ( )
- A. 中心位置左移 B. 中心位置右移 C. 分布曲线峻峭一些  
D. 分布曲线平坦一些 E. 以上都不是
11. 与标准正态分布 (u 分布) 比较, t 分布的: ( )
- A. 均数要小些 B. 均数要大些 C. 标准差要小些  
D. 标准差要大些 E. 以上都不是
12. 若总例数相同, 则两组计量资料的 t 检验与配对计量资料的 t 检验相比较, 一般为: ( )
- A. 两组计量资料的 t 检验的效率要高些 B. 配对计量资料的 t 检验的效率要高些  
C. 两者效率相等 D. 两者效率相差不大 E. 两者效率不可比
13. 在两个均数比较的 t 检验中, 计算合并方差的公式为: ( )
- A.  $S_c^2 = S_1^2 + S_2^2$  B.  $S_c^2 = \frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}$   
C.  $S_c^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2}$  D.  $S_c^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}$   
E.  $S_c^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$
14. 推断样本率为 16.8% 与 14.5% 所代表的总体有无差别, 选用的方法是: ( )
- A. 样本均数与总体均数比较的 t 检验 B. 配对 t 检验  
C. 成组 t 检验 D. 以上都不是
15. 在比较两个小样本的均数时, 需改用校正 t 检验的情况是: ( )
- A. 两总体均数相等 B. 两总体均数不等 C. 两总体方差相等  
D. 两总体方差不等 E. 两样本方差不等

16. 在两组资料比较的  $t$  检验中, 结果为  $P < 0.05$ , 差别有显著性。  $P$  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. 说明两样本均数有差别的可能性愈大          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
17.  $t$  检验中, 不同类型资料的  $t$  检验的区别是: ( )
- A. 检验步骤不同                      B. 统计量  $t$  的计算公式不同  
C. 确定  $P$  值时查的表不同          D. 根据  $P$  值判断结果的方法不同  
E. 以上都不对
18. 两组同质资料中, \_\_\_\_\_ 小的那个样本均数更有代表性 ( )
- A.  $S$               B.  $CV$               C.  $S_{\bar{x}}$               D.  $\bar{X}$
19. \_\_\_\_\_ 小, 表示用该样本均数估计总体均数的可靠性大。
- A.  $CV$               B.  $S$               C.  $S_{\bar{x}}$               D.  $R$
20. 统计推断的内容\_\_\_\_\_。
- A. 是用样本指标估计相应的总体指标              B. 是检验统计上的“假设”  
C. A、B 均不是                                      D. A、B 均是
21. 两样本均数比较, 经  $t$  检验, 差别有显著性时,  $P$  越小, 说明\_\_\_\_\_。
- A. 两样本均数差别越大  
B. 两总体均数差别越大  
C. 越有理由认为两总体均数不同  
D. 越有理由认为两样本均数不同

## 五、 问答题:

1. 标准差和标准误有何区别和联系?
2. 可信区间和参考值范围有何不同?
3. 假设检验和区间估计有何联系?
4. 假设检验时, 一般当  $P < 0.05$  时, 则拒绝  $H_0$ , 理论依据是什么?
5.  $t$  检验和方差分析的应用条件有何异同?
6. 怎样正确使用单侧检验和双侧检验?
7. 验假设中  $\alpha$  和  $P$  的区别何在?

## 六、 计算题

1. 某地抽样调查了部分成人的红细胞数和血红蛋白量, 结果如表:

表: 3-1 健康成人的红细胞和血红蛋白测得值及标准误与变异系数的计算

	性别	例数	均数	标准差	标准值
红细胞数 ( $\times 10^{12}/L$ )	男	360	4.66	0.58	4.84
	女	225	4.18	0.29	4.33
血红蛋白 ( $g/L$ )	男	360	134.5	7.1	140.2
	女	255	117.6	10.2	124.7

- (1) 说明女性的红细胞数与血红蛋白量的变异程度何者为大?
- (2) 分别计算男、女两项指标的抽样误差。
- (3) 试估计该地健康成年男、女红细胞数的均数。

(4) 该地健康成年男、女间血红蛋白含量有无差别?

2. 将20名某病患者随机分为两组, 分别用甲、乙两药治疗, 测得治疗前及治疗后一个月的血沉(mm/小时)如下表, 问:

(1) 甲, 乙两药是否均有效?

(2) 甲, 乙两药的疗效有无差别?

表3-2 甲, 乙两药治疗前后的血沉

病人号		1	2	3	4	5	6	7	8	9	10
甲药	治疗前	10	13	6	11	10	7	8	8	5	9
	治疗后	6	9	3	10	10	4	2	5	3	3
病人号		1	2	3	4	5	6	7	8	9	10
乙药	治疗前	9	10	9	13	8	6	10	11	10	10
	治疗后	6	3	5	3	3	5	8	2	7	4

(1) 甲, 乙两药是否均有效?

(2) 甲, 乙两药的疗效有无差别?

3. 将钩端螺旋体病人的血清分别用标准株和水生株作凝溶试验, 测得稀释倍数如下, 问两组的平均效价有无差别?

标准株 (11 人)	100	200	400	400	400	400	800	1600	1600	1600	3200
水生株 (9 人)	100	100	100	200	200	200	200	400	400		

4. 表 3-3 为抽样调查资料, 可做那些统计分析?

表 3-3 某地健康成人的第一秒肺通气量 (FEV<sub>1</sub>) (L)

FEV <sub>1</sub>	人 数	
	男	女
2.0~	1	4
2.5~	3	8
3.0~	11	23
3.5~	27	33
4.0~	36	20
4.5~	26	10
5.0~	10	2
5.5~	3	0
6.0~6.5	1	0
合计	118	100

5. 某医师就表 3-4 资料, 对比用胎盘浸液钩端螺旋体菌苗对 328 名农民接种前,



单选题:

1. A 2. E 3. D 4. E 5. E 6. E 7. D 8. A 9. D 10. D  
 11. D 12. B 13. E 14. D 15. D 16. E 17. B 18. C 19. C 20. D  
 21. C

问答题:

1. 标准差和标准误有何区别和联系?

表 3-6 标准差与标准误的区别

	标准差 ( $\sigma$ 或 $s$ )	标准误 ( $\sigma_{\bar{x}}$ 或 $s_{\bar{x}}$ )
意义上	描述一组变量值之间的离散趋势	描述样本均数间的离散趋势
应用上	① $s$ 越小, 表示变量值围绕均值分布越密集, 说明均数的代表性越好。 ② 可用 $\bar{X} \pm u_{\alpha} s$ 估计变量值分布范围	① $s_{\bar{x}}$ 越小, 表示样本均数与总体均数越接近, 说明样本均数推断总体均数可靠性越大。 ② 可用 $\bar{X} \pm t_{\alpha, \nu} s_{\bar{x}}$ 估计总体均数可信区间
与 $n$ 的关系	$n$ 越大, $s$ 越趋于稳定	$n$ 越大, $s_{\bar{x}}$ 越小

(2) 联系

- ① 二者均是表示变异度大小的统计指标。  
 ② 标准误  $\sigma_{\bar{x}} = \sigma / \sqrt{n}$  与标准差大小成正比, 与抽样例数  $n$  的平方根成反比。  
 ③ 当  $n$  一定时, 同一份资料, 标准差越大, 标准误也越大。

2. 可信区间和参考值范围有何不同?

参考值范围是指同质总体中个体变量值的分布范围, 如  $X \pm 1.96s$  说明有 95% 的变量值分布在此范围内, 它与标准差的大小有关, 若个体变异越大, 该范围越宽, 分布也就越散。而可信区间是指在可信度为  $(1-\alpha)$  时, 估计总体参数可能存在的范围。即从同一总体中随机抽样, 当  $n$  一定时, 每抽一次即可得一个样本均值, 以  $\bar{X} \pm t_{\alpha, \nu} s_{\bar{x}}$  计算可信区间, 如 95% 可信区间, 类似的随机抽样进行一百次, 平均有 95 次, 即有 95 个可信区间包括了总体均数, 有 5 次没有包括括总体均数, 5% 是小概率事件, 实际发生的可能性很小, 因此实际应用中就认为总体均数在求得的可信区间。这种估计方法犯错误的可能性最大不超过 5%。可信区间与标准误大小有关, 标准误越大, 可信区间则越大。

3. 假设检验和区间估计有何联系?

---

假设检验和区间估计都属于统计推断的内容。假设检验用以推断总体参数间是否有质的区别，并可获得样本统计量，以得到相对精确的概率值。而可信区间用于推断总体参数的大小，它不仅可用于回答假设检验的问题，尚可比假设检验提供更多的信息。但这并不意味着用可信区间代替假设检验，因为假设检验可得到P值，比较精确地说明结论的概率保证，而可信区间只能告诉我们在某 $\alpha$ 水准上是否有统计意义，却不能像P那样提供精确的概率。因此，只有将二者有机地结合起来，相互补充，才是完整的分析。

4. 假设检验时，一般当 $P < 0.05$ 时，则拒绝 $H_0$ ，理论依据是什么？

假设检验时，当 $P < 0.05$ ，则拒绝 $H_0$ ，其理论依据是在 $H_0$ 成立的条件下，出现大于等于现有检验统计量的概率 $P < 0.05$ ，它是小概率事件，即在一次抽样中得到这么小概率是事件是不大可能发生的，因而拒绝它。由此可见，假设检验的结论是具有概率性的，它存在犯错误的可能性小于等于0.05。

5. t检验和方差分析的应用条件有何异同？

(1) 相同点：在均数比较中，t检验和方差分析均要求各样本来自正态总体；各处理组总体方差齐且各随机样本间相互独立，尤在小样本时更需注意。

(1) 不同点：t检验仅用于两组资料的比较，除双侧检验外，尚可进行单侧检验，亦可计算一定可信度的可信区间，提示差别有无实际意义。而方差分析用于两组及两组以上均数的比较，亦可用于两组资料的方差齐性检验。

6. 怎样正确使用单侧检验和双侧检验？

根据专业知识推断两个总体是否有差别时，是甲高于乙，还是乙高于甲，两种可能都存在时，一般选双侧；若根据专业知识，如果甲不会低于乙，或研究者仅关心其中一种可能时，可选用单侧。一般来讲，双侧检验较稳妥故较多用，在预实验有探索性质时，应以专业知识为依据，它充分利用了另一侧的不可能性，故检出效率高，但应慎用。

7. 第一类错误与第二类错误的区别及联系何在？了解这两类错误有何实际意义？

(1) 假设检验中I、II型错误的区别。

I型错误是拒绝了实际上成立的 $H_0$ ，也称为“弃真”错误，用 $\alpha$ 表示。统计推断时，根据研究者的要求来确定。

II型错误是不拒绝实际上不成立的 $H_0$ ，也称为“存伪”错误，用 $\beta$ 表示。它只能与特定的 $H_1$ 结合起来才有意义，一般难以确切估计。

(2) I、II型错误的联系。

① 当抽样例数一定时， $\alpha$ 越大， $\beta$ 越小；反之， $\alpha$ 越小， $\beta$ 越大。

② 统计推断中，I、II型错误均有可能发生，若要使两者都减小，可适当增加样本含量。

③ 根据研究者要求， $n$ 一定时，可通过确定 $\alpha$ 水平来控制 $\beta$ 大小。

(3) 了解两类错误的实际意义。

① 可用于样本含量的估计。

② 用来计算可信度 $(1-\alpha)$ ，表明统计推断可靠性的大小。

③ 可用于计算把握度 $(1-\beta)$ ，来评价检验方法的效能等。

④ 有助于研究者选择适当的检验水准。

⑤ 可以说明统计结论的概率保证。

**计算题：**

1. 某地抽样调查了部分成人的红细胞数和血红蛋白量，结果如表：



表 3-7: 健康成人的红细胞和血红蛋白测得值及标准误与变异系数的计算

	性别	例数	均数	标准差	标准值	变异系数 (%)	标准误
红细胞数 ( $\times 10^{12}/L$ )	男	360	4.66	0.58	4.84	12.45	0.0306
	女	225	4.18	0.29	4.33	6.94	0.0182
血红蛋白 (g/L)	男	360	134.5	7.1	140.2	5.28	0.3742
	女	255	117.6	10.2	124.7	8.67	0.6387

(1) 说明女性的红细胞数与血红蛋白量的变异程度何者为大?

$$\text{女性 } CV_{RBC} = S/\bar{x} \times 100\% = 0.29/4.18 \times 100\% = 6.49\%$$

$$CV_{HB} = S/\bar{x} \times 100\% = 10.2/117.6 \times 100\% = 8.67\%$$

由上计算可知该地女性血红蛋白量比红细胞数变异度大

(2) 分别计算男、女两项指标的抽样误差。

见上表最后一栏, 标准误计算公式  $s_{\bar{x}} = s/\sqrt{n}$ 。

(3) 试估计该地健康成年男、女红细胞数的均数。

健康成年男子红细胞数总体均数 95%可信区间为:

$$\bar{X} \pm 1.96S_x = 4.66 \pm 1.96 \times 0.0306 = 4.60 \sim 4.72 \text{ (} 10^{12}/L \text{)}$$

其中  $n=360$  故近似按  $v=\infty$ 。同理健康成年女子红细胞数总体均数 95%可信区间为 4.14~4.22 ( $10^{12}/L$ )

(4) 该地健康成年男、女间血红蛋白含量有无差别?

$$H_0: \mu_{\text{男}} = \mu_{\text{女}}$$

$$H_1: \mu_{\text{男}} \neq \mu_{\text{女}}$$

$$\alpha = 0.05$$

$$u = (\bar{X}_1 - \bar{X}_2) / (s\bar{x}_1 - \bar{x}_2) = (134.5 - 117.6) / \sqrt{7.2^2/360 + 10.2^2/255} = 22.83$$

按  $v=\infty$ , 查附表 2, 得  $P < 0.0005$ , 按  $\alpha = 0.05$  水准, 拒绝  $H_0$ , 接受  $H_1$ , 可以认为男女间血红蛋白含量不同, 男高于女。

2. 将 20 名某病患者随机分为两组, 分别用甲、乙两药治疗, 测得治疗前及治疗后一个月的血沉(mm/小时)如下表, 问:

(1) 甲, 乙两药是否均有效?

(2) 甲, 乙两药的疗效有无差别?

表 3-8 甲, 乙两药治疗前后的血沉

病人号	1	2	3	4	5	6	7	8	9	10
甲										
药										
治疗前	10	13	6	11	10	7	8	8	5	9
治疗后	6	9	3	10	10	4	2	5	3	3
差 值	4	4	3	1	0	3	6	3	2	6
乙										
病人号	1	2	3	4	5	6	7	8	9	10
药										
治疗前	9	10	9	13	8	6	10	11	10	10
治疗后	6	3	5	3	3	5	8	2	7	4
差 值	3	7	4	10	5	1	2	9	3	6

---

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要  
下载或阅读全文，请访问：

<https://d.book118.com/188115134030006073>