

# 《卫生统计学》考试题库

华中科技大学同济医学院  
公共卫生学院流行病与卫生统计学系

## 《卫生统计学》考试题库目录

[第一章 绪论](#)

[第二章 定量资料的统计描述](#)

[第三章 正态分布](#)

[第四章 总体均数的估计和假设检验](#)

[第五章 方差分析](#)

[第六章 分类资料的统计描述](#)

[第七章 二项分布与Poisson分布及其应用](#)

[第八章  \$\chi^2\$ 检验](#)

[第九章 秩和检验](#)

[第十章 回归与相关](#)

[第十一章 常用统计图表](#)

[第十二章 实验设计](#)

[第十三章 调查设计](#)

[第十四章 医学人口统计与疾病统计常用指标](#)

[第十五章 寿命表](#)

[第十六章 随访资料的生存分析](#)

[附录：单项选择题参考答案](#)

# 第一章 绪论

[返回](#)

## 一、名词解释

1. 参数 (parameter)
2. 统计量 (statistic)
3. 总体 (population)
4. 样本 (sample)
5. 同质 (homogeneity)
6. 变异 (variation)
7. 概率 (probability)
8. 抽样误差 (sampling error)

## 二、单选题

1. 在实际工作中, 同质是指:
  - A. 被研究指标的影响因素相同
  - B. 研究对象的有关情况一样
  - C. 被研究指标的主要影响因素相同
  - D. 研究对象的个体差异很小
  - E. 以上都对
2. 变异是指:
  - A. 各观察单位之间的差异
  - B. 同质基础上, 各观察单位之间的差异
  - C. 各观察单位某测定值差异较大
  - D. 各观察单位有关情况不同
  - E. 以上都对
3. 统计中所说的总体是指:
  - A. 根据研究目的而确定的同质的个体之全部
  - B. 根据地区划分的研究对象的全体
  - C. 根据时间划分的研究对象的全体
  - D. 随意想象的研究对象的全体
  - E. 根据人群划分的研究对象的全体
4. 统计中所说的样本是指:
  - A. 从总体中随意抽取一部分
  - B. 有意识地选择总体中的典型部分
  - C. 依照研究者的要求选取有意义的一部分
  - D. 从总体中随机抽取有代表性的一部分
  - E. 以上都不是
5. 按随机方法抽取的样本特点是:
  - A. 能消除系统误差
  - B. 能消除随机测量误差
  - C. 能消除抽样误差
  - D. 能减少样本偏性
  - E. 以上都对
6. 统计学上的系统误差、测量误差、抽样误差在实际工作中:
  - A. 均不可避免
  - B. 系统误差和测量误差不可避免
  - C. 测量误差和抽样误差不可避免
  - D. 系统误差和抽样误差不可避免
  - E. 只有抽样误差不可避免
7. 统计工作的基本步骤是:
  - A. 设计、调查、审核、整理资料
  - B. 收集、审核、整理、分析资料
  - C. 设计、搜集、整理、分析资料
  - D. 调查、审核、整理、分析资料
  - E. 以上都不对
8. 统计工作的关键步骤是:
  - A. 调查或实验设计
  - B. 整理分组
  - C. 收集资料
  - D. 审核资料
  - E. 分析资料
9. 欲研究某种药物对高血压病的疗效, 临床观察 300 名病人的血压情况, 确切地说, 研究总体是:
  - A. 这 300 名高血压患者
  - B. 这 300 名高血压患者的血压值
  - C. 所有的高血压患者
  - D. 所有的高血压患者的血压值
  - E. 这种药物

10. 抽样误差是由：  
 A. 计算引起                                      B. 测量引起                                      C. 抽样引起  
 D. 采样结果不准引起                              E. 试剂、仪器未经校正引起
11. 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. 不同的总体参数之差                                      E. 以上都不是
12. 习惯上, 下列属于小概率事件的为：  
 A.  $P=0.09$     B.  $P=0.10$     C.  $P=0.15$     D.  $P=0.03$     E. 以上都不是
13. 治疗效果判定资料属于  
 A. 计量资料    B. 计数资料    C. 等级资料    D. 无序分类资料    E. 以上都不是
14. 概率  $P$  的范围：  
 A.  $-1 \leq P \leq 1$     B.  $0 < P < 1$     C.  $P \geq 1$     D.  $-1 \leq P \leq 0$     E.  $0 \leq P \leq 1$

### 三、简答题

- 1、统计学的基本步骤有哪些？
- 2、总体与样本的区别与关系？
- 3、抽样误差产生的原因有哪些？可以避免抽样误差吗？
- 4、何为概率及小概率事件？

## 第二章 定量资料的统计描述

### 第三章 正态分布

[返回](#)

#### 一、名词解释

1. 正态分布 (normal distribution)
2. 中位数 (median)
3. 四分位数间距 (quartile interval)
4. 方差 (variance)
5. 正偏态分布 (positively skewed distribution)
6. 负偏态分布 (negatively skewed distribution)
7. 对数正态分布 (logarithmic normal distribution)
8. 医学参考值范围 (medical reference range)

#### 二、单选题

1.  $\mu$  确定后,  $\delta$  越大, 则正态曲线：  
 A. 越陡峭    B. 形状不变    C. 越平缓    D. 向左移动    E. 向右移动
2. 平均数可用于分析下列哪种资料：  
 A. 统计资料    B. 等级资料    C. 计数资料    D. 计量资料    E. 调查资料
3. 常用的平均数指标有：  
 A. 样本均数、总体均数、中位数    B. 算术均数、总体均数、几何均数  
 C. 算术均数、几何均数、中位数    D. 中位数、样本均数、几何均数

- E. 以上都不对
4. 描述一组正态或近似正态分布资料的平均水平用：
    - A. 算术均数
    - B. 几何均数
    - C. 中位数
    - D. 平均数
    - E. 以上均是
  5. 用  $\bar{x} = \sum x/n$  公式计算均数的方法称为：
    - A. 加权法
    - B. 简捷法
    - C. 目测法
    - D. 平均法
    - E. 直接法
  6. 用频数表计算均数时，若以各组段下限值作为组中值计算均数，要使所得值等于原均数，则应：
    - A. 减一个组距
    - B. 加一个组距
    - C. 减半个组距
    - D. 加半个组距
    - E. 以上均不对
  7. 对于一组呈负偏态分布的资料，反映其平均水平应用哪个指标：
    - A. 几何均数
    - B. 中位数
    - C. 平均数
    - D. 均数
    - E. 算术均数
  8. 用频数表法计算均数时，组中值应为：
    - A. (本组段下限值+本组段上限值)/2
    - B. (本组下限值+下组下限值)/2
    - C. (本组下限值+下组上限值)/2
    - D. 本组段的上限值
    - E. 本组段的下限值
  9. 原始数据加上一个不为 0 的常数后：
    - A.  $\bar{x}$  不变、CV 变
    - B.  $\bar{x}$  变或 CV 变
    - C.  $\bar{x}$  不变、CV 不变
    - D.  $\bar{x}$  变、CV 不变
    - E.  $\bar{x}$ 、CV 均改变
  10. 对于对称分布的资料来说：
    - A. 均数比中位数大
    - B. 均数比中位数小
    - C. 均数等于中位数
    - D. 均数与中位数无法确定孰大孰小
    - E. 以上说法均不准确
  11. 血清学滴度资料最常计算\_\_\_\_\_以表示其平均水平。
    - A. 均数
    - B. 中位数
    - C. 几何均数
    - D. 全距
    - E. 标准差
  12. 利用频数表及公式  $M=L+i/f_n (n/2- f_L)$  计算中位数时：
    - A. 要求组距相等
    - B. 不要求组距相等
    - C. 要求数据呈对称分布
    - D. 要求变量值都比较接近
    - E. 要求数据呈对数正态分布
  13. 对于同一份正偏态的资料，求得的几何均数与算术均数：
    - A. 几何均数大于算术均数
    - B. 几何均数小于算术均数
    - C. 几何均数等于算术均数
    - D. 几何均数可以大于算术均数，也可以小于算术均数
    - E. 以上说法都不对
  14. 原始数据同时乘以一个既不等于零也不等于 1 的常数后：
    - A.  $\bar{x}$  不变，M 变 (M 为中位数)
    - B.  $\bar{x}$  与 M 都不变
    - C.  $\bar{x}$  变、M 不变
    - D.  $\bar{x}$  与 M 都变
    - E. 以上都不对
  15. 表示变量值变异情况的指标最常用的是：
    - A. 四分位数间距
    - B. 全距
    - C. 标准差
    - D. 变异系数
    - E. 方差
  16. 一组计量资料中，每个变量值与均数之差的平方和、除以变量值的个数-1，再开平方所得的值为：
    - A. 方差
    - B. 全距
    - C. 四分位数间距
    - D. 变异系数
    - E. 标准差

17. 标准差一定：  
 A. 不比均数大 B. 不比均数小 C. 取决于均数 D. 比标准误小 E. 以上都不对
18. 变异系数 CV 的数值\_\_\_\_\_。  
 A. 一定小于 1 B. 一定大于 1 C. 可大于 1;也可小于 1  
 D. 一定不会等于零 E. 一定比 S 小
19. 比较身高和体重两组数据变异度大小宜采用：  
 A. 全距 B. 四分位数间距 C. 方差 D. 变异系数 E. 标准差
20. 描述一组偏态分布资料的变异度,以\_\_\_\_\_指标较好。  
 A. R B.  $S^2$  C. Q D. S E. CV
21. 当两组(或几组)资料均数相近、度量单位相同时,标准差大的那组资料：  
 A. 均数的代表性最差 B. 均数的代表性最好  
 C. 无法据此判断出均数的代表性 D. 均数也最大  
 E. 以上都不对
22. 一组数据呈正态分布,其中小于  $\bar{x} + 2.58S$  的变量值有：  
 A. 5% B. 95% C. 97.5% D. 92.5% E. 99.5%
23. 若正常人血铅含量近似对数正态分布,拟用 300 名正常成人血铅确定 99%正常值范围,最好采用下列哪个公式：  
 A.  $\bar{x} + 2.58S$  B.  $\lg^{-1}(\bar{x}_{\lg x} + 2.58s_{\lg x})$  C.  $\bar{x} \pm 2.58S$   
 D.  $P_{99} = L + i/f_{99} (300 \times 99/100 - f_L)$  E.  $\lg^{-1}(\bar{x}_{\lg x} + 2.33s_{\lg x})$
24. 某市 120 名 12 岁男孩平均身高  $\bar{x} = 143.10\text{cm}$ ,  $S = 5.67\text{cm}$ ,则身高在 128.24–157.96cm 范围内的理论频数最接近下列哪个值?  
 A. 114 B. 119 C. 64 D. 72 E. 96
25. 若上海市健康女工 744 人血红蛋白含量的均数为 12.239g%,标准差为 0.998g%,则下列哪个最有理由认为是正常范围：  
 A. 11.24–11.3237 B. 9.654–14.814 C. 10.283–14.195  
 D. 10.592–13.886 E. 10.952–13.516

### 三、简答题

1. 描述数值变量资料集中趋势的指标有哪些?其适用范围有何异同?
2. 描述数值变量资料离散趋势的指标有哪些?其适用范围有何异同?
3. 医学中参考值范围的涵义是什么?确定的步骤和方法是什么?

### 四、计算分析题

1. 某市 100 名 7 岁男童的坐高 (cm) 如下:

63.8 64.5 66.8 66.5 66.3 68.3 67.2 68.0 67.9 69.7 63.2 64.6 64.8 66.2  
 68.0 66.7 67.4 68.6 66.8 66.9 63.2 61.1 65.0 65.0 66.4 69.1 66.8 66.4  
 67.5 68.1 69.7 62.5 64.3 66.3 66.6 67.8 65.9 67.9 65.9 69.8 71.1 70.1  
 64.9 66.1 67.3 66.8 65.0 65.7 68.4 67.6 69.5 67.5 62.4 62.6 66.5 67.2

64.5 65.7 67.0 65.1 70.0 69.6 64.7 65.8 64.2 67.3 65.0 65.0 67.2 70.2  
 68.0 68.2 63.2 64.6 64.2 64.5 65.9 66.6 69.2 71.2 68.3 70.8 65.3 64.2  
 68.0 66.7 65.6 66.8 67.9 67.6 70.4 68.4 64.3 66.0 67.3 65.6 66.0 66.9  
 67.4 68.5

- (1) 编制其频数分布表并绘制直方图，简述其分布特征；
- (2) 计算中位数、均数、几何均数，并说明用其中哪一种来表示这组数据的集中趋势为好？
- (3) 计算极差、四分位数间距、标准差，并说明用其中哪一种来表示这组数据的离散趋势为好？
- (4) 计算坐高在 64.0cm 到 68.0cm 范围内的 7 岁男童比例。
- (5) 计算 100 名 7 岁男童中坐高在 64.0cm 到 68.0cm 范围内的人数。

2. 用玫瑰花结形成试验检查 13 名流行性出血热患者的抗体滴度，结果如下，求平均滴度。

1:20    1:20    1:80    1:80    1:320    1:320    1:320  
 1:160    1:160    1:80    1:80    1:40    1:40

3. 调查某地 145 名正常人尿铅含量 (mg/L) 如下：

尿铅含量	0~	4~	8~	12~	16~	20~	24~	28~
例数	18	26	39	28	25	6	1	2

- (1) 求中位数；
- (2) 求正常人尿铅含量 95% 的正常值范围。

## 第四章 总体均数的估计和假设检验

### 第五章 方差分析

[返回](#)

#### 一、名词解释

- |                            |                                  |
|----------------------------|----------------------------------|
| 1 标准误(standard error)      | 2 可信区间(confidence interval)      |
| 3 假设检验(hypothesis testing) | 4 统计推断(statistical inference)    |
| 5 I 型错误(type I error)      | 6 II 型错误(type II error)          |
| 7 检验效能(power of test)      | 8 变量变换 (variable transformation) |

#### 二、单选题

- 1 \_\_\_\_\_ 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.  $\sigma_{\bar{x}}$     D. R    E. 四分位数间距
- 2 统计推断的内容是\_\_\_\_\_。  
 A. 用样本指标估计相应总体指标    B. 检验统计上的“假设”  
 C. A, B 均不是    D. A, B 均是    E. 估计参考值范围
- 3 两样本均数比较时，分别取以下检验水准，以\_\_\_\_\_所对应的第二类错误最小。  
 A. 0.01    B.  $\alpha=0.05$     C.  $\alpha=0.10$     D.  $\alpha=0.20$     E.  $\alpha=0.25$
- 4 方差分析中，当  $P<0.05$  时，结果\_\_\_\_\_。  
 A. 可认为各样本均数都不相等    B. 可认为各总体均数不等或不全等



- D. 减小几何均数 E. 以上都不对
15. 以下关于抽样误差，正确的是\_\_\_\_\_。  
 抽样误差仅是由个体变异产生的，抽样造成的样本统计量与总体参数的差异
- A. 抽样研究中，抽样误差是可以避免的  
 B. 对于同一总体的若干样本统计量间，也存在抽样误差  
 C. 抽样误差的大小可用标准差来说明  
 D. 以上均不对
16. 以下关于 t 分布不正确的是\_\_\_\_\_。
- A. 在相同自由度时，|t|值越大，概率 P 越小  
 B. 在相同 t 值时，双尾概率 P 为单尾概率 P 的两倍  
 C. t 分布曲线是一条曲线  
 D. t 分布的极限分布是标准正态分布  
 E. 标准正态分布可看作是 t 分布的特例
17. 以下关于可信区间，正确的是\_\_\_\_\_。
- A. 可信区间是包含未知总体参数的一个范围  
 B. 可信区间包含可信区间上下限两个值  
 C. 可信区间的确切含义是指有 (1- $\alpha$ ) 的可能认为计算出的可信区间包含了总体参数  
 D. 可信区间的确切含义也可理解是总体参数落在该范围的可能性为 1- $\alpha$   
 E. 以上说法均不对
18. 为了解某地 1 岁婴儿的血红蛋白浓度，从该地随机抽取了 1 岁婴儿 n (<50) 人，测得其血红蛋白的平均数为  $\bar{x}$ ，标准差为 S，则该地 1 岁婴儿血红蛋白的平均浓度的 95%可信区间为\_\_\_\_\_。(总体标准差  $\sigma$  未知)
- A.  $(\bar{X} - t_{\alpha/2} \frac{S}{\sqrt{n}}, \infty)$                       B.  $(-\infty, \bar{X} + t_{\alpha/2} \frac{S}{\sqrt{n}})$   
 C.  $(\bar{X} \pm t_{\alpha/2} \frac{S}{\sqrt{n}})$                       D.  $(\bar{X} \pm u_{\alpha/2} \frac{S}{\sqrt{n}})$   
 E.  $(\bar{X} \pm u_{\alpha/2} \frac{\sigma}{\sqrt{n}})$
19. 当两总体确有差异，按规定的检验水准  $\alpha$  所能发现该差异的能力的是\_\_\_\_\_。
- A.  $\alpha$       B.  $\beta$       C. 1- $\alpha$       D. 1- $\beta$       E. 以上都不对
20. 设  $\alpha$  为 I 型错误的概率， $\beta$  为 II 型错误的概率，当两总体均数确定且抽取的样本含量不变时，有\_\_\_\_\_。
- A.  $\alpha$  增大， $\beta$  减小                      B.  $\alpha$  增大， $\beta$  增大  
 C.  $\alpha$  减小， $\beta$  减小                      D.  $\alpha$  的改变不影响  $\beta$  的大小  
 E. 以上均不对
21. 在两个样本均数的假设检验中，若要同时减小 I 型错误和 II 型错误，则必须\_\_\_\_\_。
- A. 增加样本含量      B. 减小容许误差      C. 减小总体标准差  
 D. A 和 C              E. 以上都不对
22. 关于假设检验，以下不对的是\_\_\_\_\_。
- A. 根据样本统计量作出的推断结论具有概率性，因此其结论有可能出现判断错误



- B. 当根据样本作出的结论是拒绝 $H_0$ 时, 只可能犯 I 型错误, 不可能犯 II 型错误  
 C. 当根据样本作出的结论是接受 $H_0$ 时, 只可能犯 II 型错误, 不可能犯 I 型错误  
 D. 当两总体确有差异时, 按规定的检验水准所发现该差异的能力是  $\beta$   
 E. 以上均不对
23. 关于可信区间和假设检验, 不正确的是\_\_\_\_\_。  
 A. 可信区间可回答假设检验的问题  
 B. 可信区间用于说明量的大小, 假设检验用于推断质的不同  
 C. 可信区间比假设检验可提供更多的信息  
 D. 可信区间亦可回答假设检验的问题, 但可信区间不可完全代替假设检验  
 E. 以上结论均不对
24. 方差分析中对数据的要求是\_\_\_\_\_。  
 A. 任何两个观察值之间均不相关  
 B. 每一水平下的观察值分别服从总体均数为  $\mu_i$  的正态分布  
 C. 各总体的方差齐性  
 D. A, B 和 C 均对  
 E. B 和 C 对
25. 若假设检验结果为  $|t| \geq t_{0.05(\nu)}$ , 则说明\_\_\_\_\_。  
 A. 差异由抽样误差所致的概率等于或小于 0.05  
 B. 差异由抽样误差所致的概率大于 0.05  
 C. 差异由抽样误差所致的概率等于或大于 0.05  
 D. 差异是由于本质上有所不同所致的概率等于或小于 0.05  
 E. 差异是由于本质上有所不同所致的概率等于 0.05
26. 要比较的两组数值型资料呈明显偏态分布,  $n_1, n_2$  均小于 30, 且经统计检验  $\sigma_1^2 \neq \sigma_2^2$ , 此时宜采用哪种检验方法?  
 A. t 检验    B. u 检验    C. t' 检验    D. 秩和检验    E. F 检验
27. 成组资料方差分析中, 若  $SS_{组内} > SS_{组间}$ , 则\_\_\_\_\_。  
 A.  $MS_{组间} > MS_{组内}$     B.  $MS_{组间} < MS_{组内}$     C.  $MS_{组间} = MS_{组内}$   
 D.  $F \neq 1$     E. 尚不能作结论
28. 统计推断中, 可信度是指\_\_\_\_\_。  
 A.  $\alpha$     B.  $\beta$     C.  $1 - \alpha$     D.  $1 - \beta$     E. 以上均不对
29. 两样本均数的比较, 可用\_\_\_\_\_。  
 A. 方差分析    B. t 检验    C. u 检验  
 D. A, B, C 均可    E. 仅 A, B 都可
30. 两小样本数值变量资料比较的假设检验, 首先应考虑\_\_\_\_\_。  
 A. 用 u 检验    B. 用 t 检验    C. 用秩和检验  
 D. 资料符合秩和检验还是 t 检验的条件    E. 以上均不对
31. 两样本均数比较, 经 t 检验, 差别有显著性时, P 越小, 说明\_\_\_\_\_。  
 A. 两总体均数差别越大    B. 两总体均数差别越小

- 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  
E. 以上均不对

32 对 15 名 4~6 月贫血儿童用枸橼酸铁铵+维生素 C 治疗前后侧的血色素含量(g%) 见下表。

	血色素含量			
	治疗前	治疗 1 月	治疗 2 月	治疗 3 月
1	11.5	10.4	12.0	12.0
2	10.7	11.0	12.0	13.0
...	...	...	...	...
14	9.5	10.1	12.1	13.5
15	10.0	10.5	11.3	12.5

假定资料满足各种参数检验所要求的前提条件，那么，适合分析该资料的方法是

- A.  $\chi^2$  检验      B. 单因素方差分析      C. t 检验  
D. 配伍组设计方差分析      E. 秩和检验

### 三、简答题

- 1 假设检验的基本原理和步骤。
- 2 标准差和标准误的异同。
- 3 参考值范围和置信区间有何区别？
- 4  $t$  检验和方差分析的基本思想各是什么？二者的区别是什么？
- 5  $t$  检验和方差分析的应用条件有何异同？
- 6 以随机区组设计的两因素方差分析为例，简述方差分析的基本原理。
- 7 可信区间和假设检验的区别和联系。
- 8 配对  $t$  检验与两样本  $t$  检验的基本原理有何不同。

### 四、计算分析题

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

- (1) 甲、乙两药是否均有效？
- (2) 甲、乙两药的疗效是否有无差别？

		甲、乙两药治疗前后的血沉 (mm/ 小时)										
		病人号	1	2	3	4	5	6	7	8	9	10
甲	治疗前		20	23	16	21	20	17	18	18	15	19
	治疗后		16	19	13	20	20	14	12	15	13	13
乙	治疗前		19	20	19	23	18	16	20	21	20	20
	治疗后		16	13	15	13	13	15	18	12	17	14

2. 胃溃疡患者 12 人在施行胃次全切除术的前后，测定体重 (kg) 如下，问手术前后体重有无变化？

患者	1	2	3	4	5	6	7	8	9	10	11	12
术前	52.5	48	39	46	58.5	47.5	49	58	51	43	13	50
术后	72.5	51.5	40	52.5	49	55	52	52	50.5	50	41	54

3. 某医师研究血清转铁蛋白测定对病毒性肝炎诊断的临床意义, 测得 11 名正常人和 13 名病毒性肝炎患者血清转铁蛋白的含量 (U/L), 结果如下, 问患者和健康人转铁蛋白含量是否有差异?

正常人 ( $n=11$ ) 260.5 271.6 264.1 273.2 270.8 284.6 291.3 254.8 275.9  
281.7 268.6

病毒性肝炎患者 ( $n=13$ ) 221.7 218.8 233.8 230.9 240.7 256.9 253.0  
224.4 260.7 215.4 251.8 224.7 228.3

4. 某地区 1999 年测定了 30 岁以上正常人与冠心病病人的血清总胆固醇含量, 资料如下表。试检验正常人与冠心病病人血清总胆固醇含量的差别有无显著性。

正常人与冠心病病人血清总胆固醇 (mmol/L) 含量

组别	测定人数	均数	标准差	标准误
正常人	56	4.67	0.88	0.12
病人	142	5.78	1.18	0.10

5. 为试验三种镇咳药, 先以  $\text{NH}_4\text{OH}$  0.2ml 对小白鼠喷雾, 测定其发生咳嗽的时间, 然后分别用药灌胃, 在同样条件下再测定发生咳嗽的时间, 并以“用药前时间减去用药后时间”为指标, 计算延迟发生咳嗽的时间 (秒), 数据如下。试比较三种药的镇咳作用。

可待因 60 30 100 85 20 55 45 30 105

复方 2 号 50 20 45 55 20 15 80 10 75 10 60 45 40 30

复方 1 号 40 10 35 25 20 15 35 15 30 25 70 65 45 50

6. 经产科大量调查得知, 某市婴儿出生体重均数为 3.32kg, 标准差为 0.38kg, 今随机测得 36 名难产儿的平均体重为 3.43kg, 问该市难产儿出生体重的均数是否比一般婴儿出生体重均数高?

7. 已知某地 120 名正常成人脉搏均数为 73.2 次/分, 标准差为 8.1 次/分, 试估计该地正常成人脉搏总体均数的 95% 可信区间。

8. 请根据所给基本数据完成下列方差分析表。

分组	n	$\bar{X}$	S
A	8	24.375	3.543
B	6	32.00	5.329
C	5	22.00	5.431
D	8	21.00	7.463

方差分析表

变异来源	SS	v	MS	F
------	----	---	----	---

---

组间变异

组内变异

---

总变异

---

## 第六章 分类资料的统计描述

[返回](#)

### 一、名词解释

1. 率 (rate)
2. 构成比 (constituent ratio)
3. 相对比 (relative ratio)
4. 标准化率 (standardized rate)
5. 标化发病比 (standardized incidence ratio)
6. 标化死亡比 (standardized mortality ratio)

### 二、单选题

1. 某医院的资料, 计算了各种疾病所占的比例, 该指标为:  
A. 发病率 B. 构成比 C. 相对比 D. 标化发病率 E. 标化发病比
2. 男性吸烟率是女性的 10 倍, 该指标为:  
A. 相对比 B. 流行率 C. 构成比 D. 罹患率 E. 标化流行率
3. 对两地的高血压患病率进行标准化, 仅有两地各年龄组人口数和总患病率资料, 可采用的标准化方法为:  
A. 直接法 B. 间接法 C. 倒求法 D. 等比法 E. 以上都不是
4. 只有各年龄组的发病率资料, 用\_\_\_方法计算标化发病率。  
A. 直接法 B. 间接法 C. 倒求法 D. 等比法 E. 以上都不是
5. 欲反映某种疾病对人群的威胁程度, 计算\_\_\_\_\_指标。  
A. 某病病死率 B. 某病患病率 C. 某病死亡率 D. 某病构成比 E. 以上都不是
6. 各年龄组人口数和死亡率资料均有, 最好用\_\_\_\_\_方法计算标化死亡率。  
A. 直接法 B. 间接法 C. 倒求法 D. 等比法 E. 以上都不是
7. SMR 为:  
A. 实际死亡数/预期死亡数 B. 预期死亡数/实际死亡数  
C. 实际发病数/预期发病数 D. 预期发病数/实际发病数  
E. 以上都不是
8. SIR 为:  
A. 实际死亡数/预期死亡数 B. 预期死亡数/实际死亡数  
C. 实际发病数/预期发病数 D. 预期发病数/实际发病数  
E. 以上都不是
9. 间接标准化法计算标化死亡率是:  
A.  $SIR \times \text{标准人口死亡率}$  B.  $SIR \times \text{标准人口总数}$   
C.  $SMR \times \text{标准人口死亡率}$  D.  $SMR \times \text{标准人口总数}$

- E. 以上都不是
10. 间接标准化法计算标化发病率是：
- A.  $SIR \times \text{标准人口发病率}$       B.  $SIR \times \text{标准人口总数}$   
 C.  $SMR \times \text{标准人口死亡率}$       D.  $SMR \times \text{标准人口总数}$   
 E. 以上都不是
11. 对甲乙两县的死亡率进行标准化，若两县均有各年龄组死亡率，最好选择\_\_\_\_\_为标准进行标化。
- A. 两县各年龄组人口数合并      B. 甲县的人口构成资料  
 C. 乙县的人口构成资料      D. 全国或省人口普查所得的人口构成  
 E. 以上都不是
12. 观察何种疾病死因是造成当地居民死亡的主要死因，用
- A. 死亡率      B. 死因构成比      C. 疾病别死亡率  
 D. 某病病死率      E. 生存率
13. 衡量爆发性疾病发病的频度用
- A. 发病率      B. 罹患率      C. 患病率      D. 二代罹患率      E. 感染率
14. 某厂男职工 370 人，女职工 456 人，慢性苯中毒人数男女分别为 8 和 10 人，  
 $(456/370) \times 100\%$  为
- A. 率      B. 构成比      C. 相对比      D. 平均率      E. 标化患病比

### 三、简答题

- 某地 1956 年婴儿死亡人数中死于肺炎者占总死亡数的 16%，1976 年则占 18%，某医师认为该地 20 年来对婴儿肺炎的防治效果不明显。你同意他的观点吗？为什么？
- 已知甲乙两矿总职工数分别为 20856 人和 3911 人，甲矿工龄 <6 年、6~9 年、10 年以上的职工数分别为 14029、4285、2542 人，乙矿三个工龄组的职工数分别为 992、1905、1014 人，甲乙两矿尘肺总人数分别为 604 人和 127 人。两矿的尘肺发病率能否直接比较？为什么？怎样才能得出正确结论？
- 某医师研究了两种疗法对糖尿病的治疗效果，甲疗法治疗了 50 例，好转 30 例，乙疗法治疗了 10 例，好转 5 例，计算两种疗法的平均好转率为 55%，你认为正确吗？为什么？
- 某医师用自拟药方治疗肿瘤患者，治疗了 5 人，均有明显疗效，该医生报告该配方的显效率为 100%。你认为合适吗？为什么？怎样才能得出可靠的结论？

### 四、计算题

- 某研究组调查了城镇 25 岁以上居民高血压患病率，在北方城镇检查了 8450 人，其中 976 人被确诊为高血压，在南方城镇检查了 10806 人，有 1052 人被确诊为高血压，试计算南北方城镇的高血压患病率。
- 1995 年某研究组对武汉市江汉区中学生的吸烟情况、吸烟原因进行了调查，共查 1722 人，男生 839 人，172 人吸烟，女生 883 人，17 人吸烟；抽烟的主要原因有解除烦恼（64 人）、显示气派（45 人）、帮助社交（43 人）、帮助思考（16 人）、显示富有（12 人）和其它（9

人), 试计算: ①男女生吸烟率, ②男女生吸烟率之比, ③计算各种吸烟原因所占的百分构成比, 找出前三位的吸烟原因, 并描述分析结果。

3. 某医生欲研究甲乙两所大学 35 岁及以上的知识分子中高血压的患病情况, 检查了两校 35 岁及以上的全部知识分子, 得到如下资料。试对两校的患病情况进行比较, 得出正确的结论。

甲乙两校 35 岁及以上知识分子的高血压患病率

年龄 (岁)	甲校			乙校		
	检查人数	病人数	患病率 (%)	检查人数	病人数	患病率 (%)
35~	236	16	6.78	478	33	6.90
45~	375	27	7.20	379	28	7.39
55~	384	38	9.90	235	24	10.21
65~80	402	59	14.68	157	24	15.29
合计	1397	140	10.02	1249	109	8.73

4. 某研究组欲研究经常在街头小餐点就餐 (平均每天一次及以上) 的中学生是否乙肝病毒的感染率较高。在某地随机抽取了 200 名中学生, 询问他们是否经常在小餐点就餐, 并检查乙肝病毒感染情况。结果发现经常在小餐点就餐者 89 人, 乙肝感染率为 6.74%, 不经常者 111 人, 感染率为 4.50%, 试计算两类中学生乙肝感染率的标准误及总体乙肝感染率 95% 的可信区间。

## 第七章 二项分布与 Poisson 分布及其应用

### 第八章 $\chi^2$ 检验

### 第九章 秩和检验

[返回](#)

#### 一、单选题

- 在以下检验方法中, \_\_\_\_\_ 不属于非参数统计方法。  
A. t 检验    B. H 检验    C. T 检验    D.  $\chi^2$  检验    E. t 检验和  $\chi^2$  检验
- 两数值变量资料的小样本比较的假设检验, 首先应考虑  
A. 用 t 检验    B. 用 u 检验    C. 用秩和检验    D. t 检验和秩和检验均可  
E. 资料是否符合 t 检验的条件
- 等级资料比较时应选用 \_\_\_\_\_。  
A. t 检验    B. t' 检验    C.  $\chi^2$  检验    D. 秩和检验    E. 方差分析
- 在作两样本均数比较时,  $n_1$ 、 $n_2$  均小于 30、总体方差不齐且极度偏态的资料宜用 \_\_\_\_\_。  
A. t' 检验    B. t 检验    C. u 检验    D. 秩和检验    E. t' 检验、秩和检验均可

5. 设有X、Y两组数据，求得  $\hat{y} = a + bx$ ，经统计学检验，在  $\alpha = 0.05$  水平上拒绝  $H_0: \beta = 0$ ，则至少有 95% 的把握断言y与x之间在专业上有直线关系。这一结论

- A. 根据充分    B. 很有科学性    C. 脱离实际    D. 略有问题    E. 毫无根据

6. 在针刺麻醉下对三组患者进行肺切除手术，效果分四级，资料见下表(表中的数字为例数)。试分析三组患者的差别有无显著性意义。

针麻效果	患者分组			
	肺癌组	肺化脓症组	肺结核组	合计
I	10	24	48	82
II	17	41	65	123
III	19	33	36	88
IV	4	7	8	19
合计	50	105	157	312

问最适合于分析该资料的方法是

- A.  $\chi^2$  检验    B. Q 检验    C. 秩和检验    D. F 检验    E. t 检验

7. 设 x 是一个服从泊松分布的随机变量，已知  $\bar{x} = 25$ ，试计算标准差 s。

- A. 25    B. 5    C. 25/n    D. 5/n    E. n/25

8. 当自由度趋向无穷大，且  $\pi$  不接近于 0 也不接近于 1 时，二项分布趋向于\_\_\_\_\_。

- A.  $\chi^2$  分布    B. F 分布    C. 正态分布    D. t 分布    E. u 分布

9. 用大剂量 Vit. E 治疗产后缺乳，以安慰剂作对照。Vit. E 组中有效者 12 例，无效者 6 例；安慰剂组中有效者 3 例，无效者 9 例。问 Vit. E 是否有效？为了回答这个问题，应选用

- A.  $\chi^2$  检验    B. t 检验    C. F 检验    D. Fisher 精确检验    E. Q 检验

10. 设有资料如下表(表中数字为各种血型的人数)：

ABO 血型	MN 血型		
	M	N	MN
O	431	490	902
A	388	410	800
B	495	587	950
AB	137	179	32

问两种血型系统之间是否有联系，应选用的统计分析方法是

- A. 秩和检验    B.  $\chi^2$  检验    C. Ridit 分析    D. Kappa 检验    E. 方差分析

11. 用非若洛治疗不同类型关节炎的结果如下：

组别	有效例数	总例数	有效率(%)
类风湿性关节炎	140	165	84.85
风湿性关节炎	50	77	64.94
骨性关节炎	18	40	45.00

判断该药治疗不同类型关节炎的疗效之间的差别有无显著性意义，应选用的统计分析方法是

- A. 等级相关分析    B. Kappa 检验    C.  $\chi^2$  检验    D. 线形趋势检验    E. t 检验

12. 设事件 A 是一个稀有事件，在大量试验中，它发生 x 次的概率为  $P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$ ，其中  $\lambda$  为事件 A 的平均发生次数。问 x 是一个服从什么分布的随机变量？

- A. 正态分布    B. 二项分布    C. t 分布    D. 泊松分布    E.  $\chi^2$  分布

13. 设某事件在每次试验中成功的概率为  $p$ ，失败的概率为  $q=1-p$ ，在  $n$  次试验中，该事件成功  $k$  次的概率为： $P_n(k) = C_n^k p^k (1-p)^{n-k}$ ，问成功次数  $k$  服从什么分布？

- A. 泊松分布    B. 二项分布    C. 正态分布    D. F 分布    E.  $\chi^2$  分布

14. 设 1000 有名受试者，分别接受 ABO 血型系统和 MN 血型系统的检查，根据检查结果，按 (O、A、B、AB) 和 (M、N、MN) 的 12 种组合分别计数，得到一个  $4 \times 3$  列联表。为检查两种血型系统之间是否独立，需要某种检验方法，其自由度应为：

- A. 998    B. 6    C. 999    D. 11    E. 12

15. 为观察药物 A、B 对某病治愈率的差异有无显著性意义，某医生将 100 例该病患者随机分成两组，其中一组 40 人，服用 A 药；另一组 60 人，服用 B 药。结果发现，服用 A 药的人中有 30 人治愈；服用 B 药的人中有 11 人治愈。应选用的统计学方法是：

- A. 回归分析    B. 加权  $\chi^2$  检验    C.  $\chi^2$  检验    D. Kappa 检验    E. 方差分析

16. 为了解两种治疗方法对原发性肝癌的疗效，将病人随机分成两组，一组使用 5-氟尿嘧啶 + 辅助疗法 (简称 5-氟组)，另一组使用安慰剂 + 辅助疗法 (简称安慰组)。治疗结果按缓解、死亡划分。5-氟组 12 人，其中 7 人缓解，5 人死亡；安慰组 11 人，其中 4 人缓解，7 人死亡。在分析两种疗法的疗效差异有无统计学意义时，应选用的统计学分析方法是

- A.  $\chi^2$  检验    B. Fisher 精确检验    C. logistic 回归分析  
D. Ridit 分析    E. 方差分析

17. 某医院用三种方案治疗急性无黄疸型病毒性肝炎 254 例，观察结果见下表：

治疗方案	组别				
	无效	好转	显效	痊愈	合计
西药组	49	31	5	15	100
中药组	45	9	22	4	80
中西医结合组	15	28	11	20	74
合计	109	68	38	39	254

欲比较三组疗效之间的差异有无统计学意义，应选用

- A.  $\chi^2$  检验    B. Ridit 分析    C. 方差分析  
D. Fisher 精确检验    E. Kappa 检验

18. 某医院为了探讨流行性出血热的临床发病情况与其治疗原则之间的关系，收集了 446 例病情转化资料 (见下表)，试选择适当的分析方法来研究早期分度与最后定型之间的关系。

早期分度	最后定型			
	轻型	中型	危重型	合计
轻度	111	21	1	133
中度	5	163	20	188
重度	0	1	124	125
合计	116	185	145	446

- A.  $\chi^2$  检验    B. Ridit 分析    C. Fisher 精确检验    D. Kappa 检验    E. 方差分析

19. 对于以下资料，要分析不同类型气管炎疗效间的差异是否有显著性意义，应该选用什么统计分析方法？



类型	疗效				合计
	控制	显效	有效	无效	
甲型	65	18	30	13	126
乙型	42	6	23	11	82
合计	107	24	53	24	208

- A.  $\chi^2$  检验    B. Ridit 分析    C. Fisher 精确检验    D. Kappa 检验    E. 方差分析

20. 用 6.5Gy 不均匀照射狗，照射后早期动物呕吐发生例数如下。要了解三组呕吐发生率之间的差异是否有显著性意义，应选用什么统计分析方法？

- A. 回归分析    B.  $\chi^2$  检验    C. 方差分析    D. Fisher 精确检验    E. Kappa 检验

21. 某医院三年间四种甲状腺疾病在四个季节中发病情况如下表。为了解四种甲状腺疾病在四个季节中发病人数分布之间的差异是否有显著性意义，应选用什么统计分析方法？

甲状腺疾病分类	季节				合计
	春	夏	秋	冬	
甲亢	411	451	294	284	1440
亚甲炎	249	329	331	204	1113
甲低	60	61	59	52	232
甲状腺肿瘤	45	50	46	40	181
合计	765	891	730	580	2966

- A.  $\chi^2$  检验    B. Ridit 分析    C. 秩和检验  
D. 线性趋势检验    E. Fisher 精确检验

22. 在四格表  $\chi^2$  检验中，若  $\chi^2$  值为 6.86，则

- A.  $P > 0.05$     B.  $P < 0.05$     C.  $P = 0.01$     D.  $P < 0.01$     E.  $P > 0.05$

23. 以下关于非参数检验的描述哪一项是错误的

- A. 非参数方法不依赖于总体分布类型  
B. 应用非参数检验时不考虑被研究对象的分布类型  
C. 非参数检验的效能低于参数检验  
D. 一般情况下非参数检验犯第二类错误的概率小于参数检验  
E. 非参数检验方法用于分布间的比较

24. 对于满足参数检验条件的数值变量资料，如果采用秩和检验，则

- A. 第一类错误率增大    B. 第二类错误率增大    C. 第一类错误率减小  
D. 第二类错误率减小    E. 两类错误率同时减小

25. 两组或几组有序分类变量资料的比较宜用

- A.  $t$  检验    B. 回归分析    C.  $F$  检验    D. 四格表  $\chi^2$  检验    E. 秩和检验或 Ridit 分析

26. 在进行成组设计两样本秩和检验时，以下哪种检验假设是正确的

- A.  $H_0$ : 两总体均数相等     $H_1$ : 两总体均数不等  
B.  $H_0$ : 两样本均数相等     $H_1$ : 两样本均数不等  
C.  $H_0$ : 两总体分布相同     $H_1$ : 两总体分布不同

- D.  $H_0$ : 两样本分布相同  $H_1$ : 两样本分布不同  
E. 以上均不对
27. 在进行Wilcoxon配对法秩和检验时, 以下何种检验假设是正确的  
A.  $H_0$ : 两样本对应的总体均数相同  
B.  $H_0$ : 两样本均数相同  
C.  $H_0$ : 两样本对应的总体分布相同  
D.  $H_0$ : 两样本对应总体的中位数相同  
E.  $H_0$ : 差值的总体中位数为0
28. 进行两样本比较的假设检验时, 应首先考虑选用  
A.  $t$ 检验 B. 秩和检验 C.  $\chi^2$ 检验 D. 任选一种检验方法  
E. 根据资料满足哪种检验的条件来决定
29. Ridit分析适用于:  
A. 偏态分布的数值变量资料的假设检验  
B. 方差不齐的几种正态分布资料的比较  
C. 有序分类变量资料的比较  
D. 多个样本率的比较  
E. 多组样本构成比的比较
30. Ridit分析属于:  
A. 参数检验 B. 非参数检验 C. 秩和检验 D. 描述性分析 E. 相关回归分析
31. 设某地人群中糖尿病患病率为  $\pi$ , 由该地随机抽查  $n$  人, 则\_\_\_\_\_。  
A.  $n$  人中患糖尿病的人数  $x$  服从二项分布  $B(n, \pi)$   
B. 样本患病率  $p=X/n$  服从  $B(n, \pi)$   
C. 患病人数与样本患病率均服从二项分布  $B(n, \pi)$   
D. 患病人数与样本患病率均不服从二项分布  $B(n, \pi)$   
E. 患病人数或样本患病率服从二项分布  $B(n, \pi)$
32. 设  $X_1, X_2$  分别服从以  $\mu_1, \mu_2$  为均数的Poisson分布, 且  $X_1$  与  $X_2$  独立, 则\_\_\_\_\_服从以  $\mu_1 + \mu_2$  为均数的Poisson分布。  
A.  $X_1 + X_2$  B.  $X_1 - X_2$  C.  $X_1 + X_2$  与  $X_1 - X_2$  均  
D.  $X_1 + X_2$  与  $X_1 - X_2$  均不 E.  $X_1 + X_2$  或  $X_1 - X_2$
33. \_\_\_\_\_的均数等于方差。  
A. 正态分布 B. 对数正态分布 C. Poisson分布 D. 二项分布 E. 对称分布
34. \_\_\_\_\_时, 二项分布  $B(n, \pi)$  近似正态分布。  
A.  $n$  较大且  $\pi$  接近 0 B.  $n$  较大且  $\pi$  接近 1  
C.  $n$  较大且  $\pi$  接近 0 或 1 D.  $n$  较大且  $\pi$  接近 0.5 E.  $\pi$  接近 0.5
35. \_\_\_\_\_时, 二项分布  $B(n, \pi)$  近似于以  $n\pi$  为参数的 Poisson 分布。  
A.  $n$  较大且  $\pi$  接近 0 B.  $n$  较大且  $\pi$  接近 1  
C.  $n$  较大且  $\pi$  接近 0 或 1 D.  $n$  较大且  $\pi$  接近 0.5 E.  $\pi$  接近 0.5
36. 用二项分布直接计算概率法检验  $H_0: \pi = 0.4$   $H_1: \pi > 0.4$ 。当随机样本含量  $n=10$ , 阳性

数  $X=6$  时, 为作统计推断应将概率  $p=$ \_\_\_\_\_与检验水准  $\alpha$  比较。

- A.  $p(X=6)$
- B.  $p(X=6) + p(X=7) + \dots + p(X=10)$
- C.  $p(X=7) + p(X=8) + \dots + p(X=10)$
- D.  $p(X=6) + p(X=5) + \dots + p(X=0)$
- E.  $p(X=5) + p(X=4) + \dots + p(X=0)$

37. 在样本率  $p$  与总体率  $\pi_0$  比较时, 用近似正态  $u$  检验的条件是\_\_\_\_\_。

- A. 样本含量  $n$  较大、总体率  $\pi_0$  接近 0.5
- B. 样本含量  $n$  较大、总体率  $\pi_0$  接近 1
- C. 样本含量  $n$  较大、样本率  $p$  接近 0.5
- D. 样本含量  $n$  较大、样本率  $p$  接近 1
- E. 样本率  $p$  不太小

38. 从甲、乙两文中, 查到同类研究的两个率比较的四格表资料, 其  $\chi^2$  检验甲文  $\chi^2 > \chi^2_{0.01(1)}$ , 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. 甲文结果更可信
- D. 甲文结果不可信
- E. 甲文说明总体的差别大

39. 四个样本率作比较,  $\chi^2 > \chi^2_{0.01(3)}$ , 可认为\_\_\_\_\_。

- A. 各总体率不等或不全相等
- B. 各总体率均不相等
- C. 各样本率均不相等
- D. 各样本率不等或不全相等
- E. 两个总体率相等

40. 四格表的自由度\_\_\_\_\_。

- A. 不一定等于 1
- B. 一定等于 1
- C. 等于行数  $\times$  列数
- D. 等于样本含量减 1
- E. 以上都不对

41.  $\chi^2$  值的取值范围为\_\_\_\_\_。

- A.  $-\infty < \chi^2 < +\infty$
- B.  $\chi^2 \leq 1$
- C.  $0 \leq \chi^2 \leq +\infty$
- D.  $\chi^2 \geq 1$
- E.  $-\infty \leq \chi^2 \leq 0$

42. 当自由度不变时, 关于  $\chi^2$  值与  $P$  值的关系, 下列哪一项是正确的

- A.  $\chi^2$  值越大,  $P$  值越大。
- B.  $\chi^2$  值越大,  $P$  值越小。
- C.  $\chi^2$  值变化时,  $P$  值不变。
- D.  $\chi^2$  值变化时,  $P$  值变大或变小。
- E.  $\chi^2$  值变化时,  $P$  值可能变化也可能不变。

43. 在  $\chi^2$  检验中, 自由度的计算为

- A.  $R \times C$
- B.  $n$
- C.  $n-1$
- D.  $(R-1) \times (C-1)$
- E.  $R \times C - 2$

44. 作四格表  $\chi^2$  检验时, 需进行连续性矫正的条件是

- A.  $1 < T < 5$  且  $n > 40$
- B.  $b+c < 40$
- C.  $T < 1$  或  $n < 40$
- D.  $T > 5$  且  $n > 40$
- E.  $b+c > 40$

45. 检验计数资料的两种属性或特征之间有无关联时, 常用的方法为

- A.  $t$  检验
- B.  $u$  检验
- C.  $\chi^2$  检验
- D. 秩和检验
- E. 正态性检验

## 二、分析计算题

1. 试就下表资料分析比较甲、乙两医院乳腺癌手术后的 5 年生存率(%)。

甲、乙两医院乳腺癌手术后的 5 年生存率(%)

腋下淋巴 结转移	甲医院			乙医院		
	病例数	生存数	生存率	病例数	生存数	生存率
无	45	35	77.77	300	215	71.67
有	710	450	68.38	83	42	50.60
合计	755	485	64.24	383	257	67.10

2. 已知某药的治愈率为 60%。现欲研究在用此药的同时加用维生素 C 是否有增效作用，某医生抽取 10 名病人试用此药加用维生素 C，结果 8 人治愈，请作统计推断。

3. 欲了解 A、B 两地妇女的乳腺癌患病率是否相同，分别从两地抽样调查 1 万妇女，结果 A 地患病人数为 80 人，B 地 102 人。请作统计分析。若 A 地调查 2 万名妇女，160 人患病，B 地调查 1 万名妇女，102 人患病，则统计结论是什么。

4. 按国家规定平均每毫升饮用水中的细菌数不得超过 100 个，现从某饮用水随机抽取 2 毫升水测得细菌 215 个，问该水源是否符合饮用水的国家卫生标准。

5. 按国家规定平均每升饮用水中的大肠菌数不得超过 3 个，现从某饮用水随机抽取 1 升水测得大肠菌 4 个，问该饮用水是否符合饮用水的国家卫生标准。

6. 根据下表数据，试比较使用含氟牙膏与使用一般牙膏者的患龋率有无不同？

使用含氟牙膏与使用一般牙膏者的患龋率

牙膏类型	调查人数	患龋齿人数	患龋率(%)
含氟牙膏	200	70	35
一般牙膏	100	50	50
合计	300	120	40

7. 某卫生防疫站在中小学观察三种矫正治疗近视眼措施的效果，近期疗效数据见下表。结论为“近期疗效以“夏天无”眼药水为最好，眼保健操次之，新医疗法最差”。试对此说法作出分析评价。

三种措施的近期有效率(%)

矫正方法	观察例数	近期有效率(%)
“夏天无”眼药水	135	37.78
新医疗法	32	18.75
眼保健操	18	27.78

8. 某厂在冠心病普查中研究冠心病与眼底动脉硬化的关系，资料整理如下。问两者之间是否存在一定的关系？

冠心病诊断结果与眼底动脉硬化级别的关系

眼底动脉 硬化级别	冠心病诊断结果			合计
	正常	可疑	冠心病	
0	340	11	6	357
I	73	13	6	92
II	18	18	133	

III	2	1	6	
合计	513	44	31	588

- 用两种方法检查已确诊的乳腺癌患者 120 名。甲法的检出率为 60%，乙法的检出率为 50%，甲、乙两法一致的检出率为 35%，试问两种方法何者更优？
- 在某克山病区作中小学生心肌受损情况检查，277 名男生中检出率为 48.74%，147 名女生中检出率为 57.10%，问男女生心肌受损率是否不同？
- 用某种化学物质进行诱发肿瘤试验，实验组 15 只小白鼠中 4 只发生癌变，对照组 10 只无一发生癌变，问两组发癌率有无差别？

### 三、简答题

- 二项分布、Poisson 分布与正态分布间有何关系？
- $\chi^2$  检验的应用条件有哪些？
- $\chi^2$  检验用于解决哪些问题？
- 四格表的 u 检验与  $\chi^2$  检验有何异同？
- 请列举 R×C 表  $\chi^2$  检验的注意事项。
- R×C 行列表与 R×C 双向有序分类列联表  $\chi^2$  检验的区别是什么？
- 某医生观察某新药预防流感的效果，并用四格表专用公式计算出  $\chi^2=6.63$ ,  $P=0.01$ 。问：
  - 此表的四个基本数字 a, b, c, d 各等于多少？
  - 你对此有何看法？

分组	观察人数	未发病	有效率 (%)
服药组	180	130	72.2
未服药组	230	190	82.6

## 第十章 回归与相关

[返回](#)

### 一、名词解释

- 自变量 (independent variable)
- 因变量 (dependent variable)
- 简单相关分析 (simple correlation analysis)
- 简单线性回归分析 (simple linear regression analysis)
- 偏相关分析 (partial correlation analysis)
- 多元线性回归分析 (multivariate linear regression analysis)
- 最小二乘法 (least square method)
- Pearson 相关分析 (Pearson correlation analysis)
- Spearman 相关分析 (Spearman correlation analysis)
- 确定系数 (determinant coefficient)

## 二、单选题

- 1、相关系数检验的无效假设 $H_0$ 是. \_\_\_\_\_  
A.  $\rho > 0$ ,      B.  $\rho = 0$ ,      C.  $\rho = 1$ ,      D.  $\rho \neq 0$
- 2、由样本计算两个随机变量  $x$  和  $y$  之间的简单相关系数  $r$  的值近似等于零, 经统计检验得到  $p=0.90$ 。作结论时, 正确的表述应该是\_\_\_\_\_。  
A.  $x$  与  $y$  之间呈直线关系                      B.  $x$  与  $y$  之间呈曲线关系  
C.  $x$  与  $y$  之间没有直线关系                      D.  $x$  与  $y$  之间毫无关系
- 3、线性回归分析的原理是对因变量  $Y$  的总体变异进行分解。它可能出现\_\_\_\_\_。  
A.  $SS_{剩} = SS_{回}$                       B.  $SS_{总} > SS_{剩}$   
C.  $SS_{总} = SS_{回}$                       D.  $SS_{剩} < SS_{回}$                       E. 以上均可
- 4、对变量  $X$  和  $Y$  同时进行简单相关分析和简单回归分析, 其结果有\_\_\_\_\_。  
A.  $r > 0, b < 0$                       B.  $r < 0, b > 0$   
C.  $r > 0, b > 0$                       D.  $r = b$                       E.  $r$  与  $b$  符号无关
- 5、已知  $r=1$ , 则一定有\_\_\_\_\_。  
A.  $b=1$                       B.  $a=1$   
C.  $SS_{剩}=0$                       D.  $SS_{总}=SS_{剩}$                       E.  $SS_{剩}=SS_{回}$
- 6、含有常数项的直线回归系数假设检验, 其自由度是\_\_\_\_\_。  
A.  $n$                       B.  $n-1$   
C.  $n-2$                       D.  $2n-1$                       E.  $2n-2$
- 7、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. 相关分析或回归分析                      D.  $\chi^2$  检验
- 8、对简单线性回归模型进行显著性检验的目的是对\_\_\_\_\_作出统计推断。  
A. 样本斜率      B. 总体斜率      C. 样本均数      D. 总体均数
- 9、如果对简单线性回归模型进行显著性检验的结果是不能拒绝 $H_0$ , 这就意味着\_\_\_\_\_。  
A. 该模型有应用价值                      B. 该模型无应用价值  
C. 该模型求错了                      D.  $X$  与  $Y$  之间毫无关系
- 10、对两个数值变量同时进行了相关和回归分析,  $r$  有统计学意义 ( $P < 0.05$ ), 则  
A.  $b$  无统计学意义  
B.  $b$  有高度的统计学意义。  
C.  $b$  有统计学意义  
D. 不能确定  $b$  有无统计学意义。  
E. 以上都不是
- 11、设有  $X, Y$  两组数据, 求得  $\hat{y} = a + bx$ , 经统计学检验, 在  $\alpha = 0.05$  水平上拒绝 $H_0: \beta = 0$ , 则至少有 95% 的把握断言  $y$  与  $x$  之间在专业上有直线关系。这一结论 ( )。  
A. 根据充分      B. 很有科学性      C. 脱离实际      D. 略有问题      E. 毫无根据
- 12、在对两个变量  $x$  与  $y$  进行直线相关分析后发现, 相关系数  $r$  约等于 0, 经检验, 得  $P > 0.9$ 。

在下专业结论时，正确的表述应该是

- A.  $x$  与  $y$  之间呈直线关系    B.  $x$  与  $y$  之间呈曲线关系    C.  $x$  与  $y$  之间呈没有关系  
D.  $x$  与  $y$  之间有无关系尚未确定    E.  $x$  与  $y$  之间存在某种关系
- 13、 $x_1$ 、 $x_2$  分别代表给大鼠注射一定剂量的类毒素后测得的大鼠红细胞与血红蛋白含量。求得直线回归方程为： $\hat{x}_2 = 6.188441 + 0.199953x_1$ ， $n=19$ ， $r=0.82031$ ，检验该直线回归方程是否有显著性意义的统计学结论应该为  
A.  $P>0.05$     B.  $0.01<P<0.05$     C.  $0.005<P<0.01$     D.  $P<0.001$     E.  $P<0.01$
- 14、求得  $y$  倚  $x$  变化的直线回归方程后，必须对回归方程作显著性检验，其目的是为了对（ ）作出统计推断。  
A. 样本斜率    B. 总体斜率    C. 样本均数    D. 总体均数    E. 样本分布
- 15、在求出直线回归方程后，如果检验结果是接受无效假设，那就意味着  
A. 此直线方程有应用价值    B. 此直线方程无应用价值  
C. 此直线方程并非所求    D.  $x$  与  $y$  之间毫无关系    E.  $x$  与  $y$  之间呈直线关系
- 16、在求出  $y$  倚  $x$  变化的直线回归方程  $\hat{y} = a + bx$  后，发现将原始数据中的某一点  $(x_k, y_k)$  的横坐标值代入方程所得的  $\hat{y}_k \neq y_k$ ，这说明  
A. 此现象无法解释    B. 正常现象    C. 计算有错  
D.  $x$  与  $y$  之间呈曲线关系    E.  $x$  与  $y$  之间呈直线关系
- 17、对同一个资料，根据最小平方原则求出两个直线方程  $\hat{y} = a + bx$  和  $\hat{x} = a + by$ ，在一般情况下，它们之间的关系是  
A. 重合    B. 垂直    C. 平行    D. 相交    E. 不确定
- 18、两因素之间有显著性的交互作用，意味着  
A. 因素 A 的作用随因素 B 的作用增强而增强  
B. 因素 A 的作用随因素 B 的作用减弱而减弱  
C. 因素 A 的作用随因素 B 的作用变化而变化  
D. 一个因素的水平对实验结果的影响随另一个因素水平的改变而改变  
E. 一个因素的水平对实验结果的影响不随另一个因素水平的改变而改变
- 19、根据 10 对  $(x, y)$  的某资料算得  $y$  倚  $x$  变化的直线回归方程为  $\hat{y} = 10.376109 + 0.199008x$ 。 $x$  与  $y$  之间的直线相关系数  $r=0.848915$ 。现在原始数据已丢失，问若提供以下的哪一项，就可准确判定上述直线回归方程是否具有显著性意义？  
A. 至少 8 对  $(x, y)$  的值    B. F 临界值表    C.  $\chi^2$  临界值表  
D.  $r$  临界值表    E.  $t$  临界值表
- 20、有两组适合于作直线相关分析的试验资料（按专业知识都应取双侧检验），第一组资料： $n_1=5$ ，相关系数  $r_1=0.857$ ；第二组资料： $n_2=8$ ，相关系数  $r_2=0.712$ 。在没有详细资料和各种统计用表的条件下，可作出什么统计推断？  
A. 因  $r_1>r_2$ ，故  $r_1$  有显著性意义  
B. 因  $n_1>n_2$ ，故  $r_2$  有显著性意义  
C.  $r_1$  和  $r_2$  都有显著性意义

- D.  $r_1$ 和 $r_2$ 都没有显著性意义
- E. 缺少作出明确推断的依据

21、关于相关系数 $r$ 与决定系数 $r^2$ 含义上的区别，下面哪一项表述最合适？

- A.  $r$  值的大小反映两个变量之间是否有密切的关系
- B.  $r^2$ 值的大小反映两个变量之间呈直线关系的密切程度和方向
- C.  $r$  值接近于 0 表示两个变量之间没有任何关系
- D.  $r^2$ 值接近于 0 表示直线回归的贡献很小
- E.  $r$ 与 $r^2$ 在含义上没有区别

22、不同地区水中平均碘含量与地方性甲状腺肿患病率的资料如下。为了通过测定碘含量来预测地方性甲状腺肿的患病率，应选用

	地区编号					
	1	2	3	4	...	17
碘含量(单位)	10.0	2.0	2.5	3.5	...	24.5
患病率(%)	40.5	37.5	39.0	20.0	...	0.0

- A. 相关分析
- B. 回归分析
- C. Ridit 分析
- D. 聚类分析
- E. 因子分析

23、对含有两个随机变量的同一批资料，即作直线相关，又作直线回归分析。令对相关系数检验的 $t$ 值为 $t_r$ ，对回归系数检验的 $t$ 值为 $t_b$ ，二者之间存在什么关系？

- A.  $t_r > t_b$
- B.  $t_r < t_b$
- C.  $t_r = t_b$
- D.  $t_r \geq t_b$
- E.  $t_r \leq t_b$

24、设  $\rho$  为总体相关系数，根据实际资料算得样本相关系数  $r$  后，需进行显著性检验，其无效假设应为：

- A.  $H_0: r=0$
- B.  $H_0: r \neq 0$
- C.  $H_0: \rho=0$
- D.  $H_0: \rho \neq 0$
- E.  $H_0: \rho \leq 0$

25、某人把若干只钉螺按同样条件同时埋入土中，每个一个月取出一部分，检测其存活率，直至第 12 个月。分别用  $x$ 、 $y$  表示埋螺月数和存活率，发现随  $x$  值逐渐增大， $y$  值由开始的 100%逐渐下降。此人用直线回归方程描述了  $y$  随  $x$  变化而变化的关系，经统计学检验， $P < 0.01$ ，结论是此资料可用直线回归方程来描述。你对此有何评论？

- A. 有统计学检验为依据，故此结论可信
- B. 试验点太少，无法下结论
- C. 数据未全部写出，无法判断
- D. 没有散点图，单靠  $P$  值是不行的
- E. 此资料应该用曲线回归方程而不是直线回归方程来描述

26、给 10 只中年大鼠注射类毒素后，测得每只大鼠的红细胞数 $x$  ( $\times 10^4/\text{mm}^3$ )与血红蛋白含量 $y$  (g/L)分别如下：

鼠号	1	2	3	4	5	6	7	8	9	10
X	654	786	667	605	761	642	652	706	602	539
Y	130	168	143	130	158	129	151	153	149	109

为了对两个变量  $x$ 、 $y$  进行线性回归分析(设  $x$  为自变量， $y$  为因变量)，首先要从专业上考虑是否有理由研究它们之间的关系，然后应当做的事就是



- A. 求出直线回归方程并检验                      B. 求出相关系数并检验  
a) 绘制散点图并分析散点的变化趋势 D. 应用回归方法解决实际问题  
E. 求出回归系数并检验
- 27、某人测得 20 人的身高和体重数据，求出了二者之间的直线相关系数，查相关系数临界值表时，其自由度应为  
A. 20          B. 19          C. 18          D. 17          E. 16
- 28、测的一组正常人的胆固醇值和血磷值，要对二者同时进行统计分析，可选用那种方法？  
A. 配对设计计量资料的 t 检验  
B.  $\chi^2$  检验                                      C. 相关回归分析  
D. 多元方差分析                              E. 单因素方差分析
- 29、若  $r > r_{0.05(v)}$ ，则  
A.  $P > 0.05$           B.  $P < 0.05$           C.  $P \leq 0.05$           D.  $P \geq 0.05$           E.  $P > 0.01$
- 30、对两个变量进行直线相关分析， $r = 0.39$ ， $P > 0.05$ ，说明两个变量之间  
A. 有相关关系                      B. 有数量关系                      C. 有因果关系  
D. 有伴随关系                      E. 无线性相关关系
- 31、对相关系数  $r$  进行显著性检验，当  $r > r_{0.05(v)}$ ，则 \_\_\_\_\_。  
A. 两变量之间关系密切  
B. 两变量之间相关有统计学意义  
C. 两变量之间关系不密切  
D. 两变量之间相关无统计学意义  
E. 以上都不是

### 三、简答题

- 1、简单相关分析与简单线性回归分析的联系和区别是什么？
- 2、Pearson 相关分析和 Spearman 相关分析各有什么优缺点？
- 3、简单线性回归分析和多元线性回归分析对数据有什么要求？
- 4、简单线性回归模型中的回归系数和多元线性回归模型中的偏回归系数有什么异同点？
- 5、线性回归分析要求因变量是服从正态分布的随机变量。如果该条件不满足，会导致什么不良后果？说出一个解决这种问题的有效方法。
- 6、多元线性回归分析中要求自变量之间不存在多重共线性。如果该条件不满足，会导致什么不良后果？说出一个解决这种问题的有效方法。
- 7、多元线性回归分析模型中，标准偏回归系数和非标准偏回归系数的作用区别是什么？
- 8、在一个因变量 Y 关于自变量 X 的简单线性回归分析模型中，自变量 X 的回归系数估计值是 b，并且经过统计检验，X 的回归系数具有显著性统计意义。试解释 b 的统计意义。
- 9、在一个因变量 Y 关于子变量 X1 和 X2 的二元线性回归分析模型中，自变量 X1 和 X2 的回归系数估计值分别是 b1 和 b2，并且经过统计检验，X1 和 X2 的回归系数均具有显

著性统计意义。试解释  $b_1$  和  $b_2$  的统计意义。

10、已知一个因变量  $Y$  关于子变量  $X_1$  和  $X_2$  的二元线性回归分析模型的复确定系数  $R^2=0.85$ ，试解释它的统计意义。

#### 四、计算题

1、随机抽得 10 名 20 岁男性青年的身高和前臂长如下：

身高 (cm)  $Y$ : 170, 173, 160, 155, 173, 188, 178, 183, 180, 165

前臂长 (cm)  $X$ : 45, 42, 44, 41, 47, 50, 47, 46, 49, 43

a) 计算相关系数  $r$ ，并对相关系数  $\rho$  进行检验；

b) 计算  $Y$  对  $X$  的回归方程，并对回归系数和总体模型进行统计检验。

2、已知随机变量  $X$  和  $Y$  的方差估计值分别是  $\text{Var}(X)=0.25$  和  $\text{Var}(Y)=0.16$ ，且变量  $X$  和  $Y$  之间的协方差估计值是  $\text{Cov}(X, Y)=0.1$ ，试估计变量  $X$  和  $Y$  之间的相关系数  $r(X, Y)$ 。

3、根据回归分析原理，填充下面回归分析结果表。

Source	DF	Sum of squares	Mean square	F-value
Model	2			
Residual		1.82		
Total	12	38.22		

## 第十一章 常用统计图表

## 第十二章 实验设计

## 第十三章 调查设计

[返回](#)

### 一、名词解释

1. 实验研究设计 (experiment design)
2. 随机对照试验 (random-control trail)
3. 配对设计 (Paired design)
4. 交叉设计 (Cross-over design)
5. 盲法 (blind-method)
6. 临床实验 (clinical trial)
7. 社区干预实验 (community intervention trial)
8. 简单随机抽样 (simple random sampling)
9. 样本量估计 (sampling survey)
10. 整群抽样 (cluster sampling)
11. 系统抽样 (systematic sampling)

### 二、单选题

1、研究吸烟与肺癌的关系中年龄与性别常常与肺癌有关联，是肺癌的危险因素，因此在该项研究中，它们是\_\_\_\_\_。

A、处理因素 B、混杂因素 C、研究因素 D、观察因素 E、实验因素

2、在实验设计研究中，随机化分组可以减少或控制\_\_\_\_\_。

A. 系统误差 B. 测量误差 C. 实验误差 D. 随机误差

- 3、下列各种实验设计中，在相同条件下最节约样本含量的是\_\_\_\_\_。
  - A. 完全随机设计
  - B. 配伍组设计
  - C. 配对设计
  - D. 交叉设计
- 4、设立对照应满足\_\_\_\_\_原则，才能显示对照的作用。
  - A. 随机化分组
  - B. 重复
  - C. 均衡
  - D. 随机数字表
  - E. ABCD 均不是
- 5、实验设计和调查设计的根本区别是\_\_\_\_\_。
  - A. 实验设计以动物为对象
  - B. 调查设计以人为对象
  - C. 实验设计可随机分组
  - D. 实验设计可人为设置处理因素
  - E. 二者无区别
- 6、某医生在研究某药物治疗糖尿病的疗效时，试验组服用该药物，对照组服用淀粉，这属于：
  - A. 实验对照
  - B. 历史对照
  - C. 安慰剂对照
  - D. 空白对照
  - E. 标准对照
- 7、下列哪种研究可以认为地设置各种处理因素：
  - A. 调查研究
  - B. 横断面研究
  - C. 实验研究
  - D. 病例对照研究
- 8、下列抽样方法，在相同条件下，抽样误差最大的为：
  - A. 单纯随机抽样
  - B. 系统抽样
  - C. 整群抽样
  - D. 分层抽样
  - E. 分层随机抽样
- 9、有一组某地居民 1950~1968 年伤寒与结核病死亡率（1/10 万）资料，下列哪种统计图可以形象化地反映两种疾病的死亡率随着时间推移的变化速度。
  - A. 条图或圆图
  - B. 直方图
  - C. 散点图
  - D. 普通线图
  - E. 半对数线图
- 10、实验设计的基本原则是：
  - A. 重复、随机、配对
  - B. 随机、盲法、对照
  - C. 对照、随机、重复
  - D. 随机、盲法、配对
  - E. 齐同、对照、随机

### 三、简答题

- 1、实验设计的基本原则是什么？其各自目的何在？
- 2、确定样本含量应该具备什么条件？
- 3、请列举常见的实验设计方法，并简述其优缺点。
- 4、常用的统计图有哪些？它们的适用条件是什么？
- 5、实验误差有哪几类？如何控制实验误差？
- 6、普通线图和半对数线图的主要区别是什么？
- 7、请简述四种基本抽样方法的优缺点和适用条件。
- 8、完全随机化设计与随机化区组设计各有什么优缺点？
- 9、对照、随机化、均衡和重复各自的目的何在？请举例说明。
- 10、直方图、条图和百分条图有什么区别？

### 四、计算题

1. 某种心血管疾病的 55 岁以上男性患者的平均体重为 90kg，标准差为 20kg，现采用

- 一种新的疗法，欲以  $\alpha=0.05$ ，检验效能为 90%：①估计病人的体重能降至 85kg，需多大的样本？②估计病人的体重改变 5kg，需多大的样本？
2. 已经知道某民族中 HBsAg 阳性率为 12.45%，现欲抽样调查该民族 HBsAg 阳性率，要求误差不超过 2%，问需要调查多少人？
  3. 在新生儿某种病毒爆发期间，某地区已发现 1000 名活产婴儿中有 160 名感染，现经一段时间治疗，欲了解目前感染率是否降至 0.10，取  $\alpha=0.05$ ， $\beta=0.10$ ，问需要抽取多大的样本？
  4. 欲了解某地菜农钩虫感染率是否高于粮农，估计两总体率约为 60%、85%，今指定  $\alpha=0.05$ ， $\beta=0.10$ ，则每组要查多少人？
  5. 为研究孕妇在怀孕期间实用某种试验饮食是否会增加新生儿出生时的体重，欲抽取两组孕妇，一为普通饮食组，一为试验饮食组。据以往经验，新生儿体重的标准差为 500g，设  $\alpha=0.05$ ， $\beta=0.20$ ，检验效能为 80%，检测出试验饮食会使新生儿体重增加 100g，问需要多大的样本？

## 五、设计与分析题

1. 某地 1960 年和 1995 年三种死因别死亡率如下表，试将该资料绘制成统计图，并作简要分析。

死因	1960 年	1995 年
肺结核	165.2	20.1
心脏病	72.5	87.1
恶性肿瘤	53.6	124.9

2. 一次临床试验中需要 100 例病人，根据就诊先后顺序，将先就诊的 50 名病人分入 A 组，后就诊的 50 名病人分入 B 组。你认为这种设计方案是否遵循了有关统计学原则？具体理由是什么？
3. 某单位研究菊花艾叶香预防感冒和空气消毒效果。对象为某幼儿园分住三个楼的儿童。中间楼是中班儿童（160 人），东楼是小班儿童（110 人），这两栋楼燃香；西楼是大班儿童（170 人），作为对照组，不燃香。结论为菊花艾叶香无预防感冒效果，但有空气消毒作用（肉汤平板上菌落数较少）。请按调查设计和实验设计的原则对此加以评述。
4. 在生化实验中要评定 A、B、C、D 四种检验方法，操作者有甲、乙、丙、丁四人，每人作四次实验，请设计这一实验。
5. 某研究所开展一项新药临床试验，共收病例 200 名进入试验，如果将病例平分为新药组与对照组，请列举可能的几种设计方案，并分析其各自的特点。

## 第十四章 医学人口统计与疾病统计常用指标

[返回](#)

### 一. 名词解释

1. 负担系数
2. 人口金字塔
3. 总生育率 (general fertility rate, GFR)
4. pearl 怀孕率 (pear pregnancy rate)
5. 婴儿死亡率 (infant mortality rate, IMR)
6. 新生儿死亡率 (neonatal mortality rate, NMR)
7. 围生儿死亡率 (perinatal mortality)
8. 孕产妇死亡率 (maternal mortality rate)
9. ICD (International Classification of Disease)

### 二. 选择题

1. 为了对某种疾病进行研究出发, 要获得某地人口总数, 用什么方法比较好?  
A. 法定制 B. 实际制 C. 时点制 D. 实数制
2. 下列说法, 错误的是?  
A. 老年人口系数是构成比  
B. 少年儿童人口系数是构成比  
C. 负担系数是构成比  
D. 性别比是相对比

### 三. 简答题

1. 简述医学人口统计资料的来源及特点。
2. 简述反映生育水平的常用指标及其计算方法。
3. 简述反映人口再生产的常用统计指标及其计算方法。
4. 简述反映计划生育工作情况的常用统计指标及其计算方法。
5. 测量死亡水平的常用指标有哪些? 如何计算?
6. 疾病统计的主要资料来源有哪些?
7. 常用的疾病统计指标有哪些? 如何计算?
8. 患病率和发病率有什么区别?

## 第十五章 寿命表

[返回](#)

### 一. 名词解释

1. 寿命表 (life table)
2. 现时寿命表 (current life table)

3. 定群寿命表(cohort life table)
4. 年龄组死亡概率(age specific probability of dying)
5. 去死因寿命表 (cause eliminated life table)
6. 平均预期寿命 (life expectancy)

## 二. 选择题

1. 寿命表中所含有的指标可不包括：
  - A. 年龄
  - B. 年龄组死亡概率
  - C. 尚存人数
  - D. 生存人年数
  - E. 总死亡率
2. 如要进行某种药物对肿瘤疗效的分析，应选用？
  - A. 现时寿命表
  - B. 完全寿命表
  - C. 简略寿命表
  - D. 定群寿命表
3. 寿命表中的“年龄组死亡概率”实质上是一个：
  - A. 条件概率
  - B. 频率
  - C. 构成比
  - D. 相对比
4. 以下关于寿命表的说法，错误的是：
  - A. 预期寿命可能随着年龄的增加而增加
  - B. 可用校正婴儿死亡率作为 0 岁组死亡概率的估计值
  - C.  $x$  岁尚存者在今后  $n$  年内的生存概率为  ${}_n p_x = 1 - nq_x$
  - D. 两地平均寿命的比较应该考虑人口构成不同的问题
5. 寿命表制作时的关键指标是：
  - A. 年龄
  - B. 年龄组死亡概率
  - C. 尚存人数
  - D. 生存人年数
  - E. 平均预期寿命
6. 研究某种疾病导致的死亡对居民生命的影响时，下列那种方法最好：
  - A. 简略寿命表
  - B. 现时寿命表
  - C. 定群寿命表
  - D. 去死因寿命表
7. 各年龄的死亡水平发生改变都会影响平均寿命的大小，其中哪个年龄别的影响最大？
  - A. 65 岁到 70 岁以前组
  - B. 30 岁到 35 岁以前组
  - C. 零岁到 1 岁以前组
  - D. 25 岁到 30 岁以前组
8. 如果寿命表中死亡人数分布中婴幼儿峰度上升，老年峰位置前移是说明居民的健康水平：
  - A. 提高
  - B. 降低
  - C. 没有变化
  - D. 还不能下结论

## 三. 简答题

1. 平均死亡年龄和平均预期寿命是否相等，为什么？
2. 简述寿命表制作需要哪些指标？如何求？（即寿命表制作的原理与方法）
3. 简述寿命表的用途。

## 四. 计算题

1. 根据所学知识完成表 15.2。
2. 根据所学知识，完成下表，其中年龄组平均人口数同表 15.2。

表 15.2 某地 1982 年男性简略寿命表

年龄组 (岁)	平均 人口数	实际死 亡人数	年龄组 死亡率	死亡 概率	尚存 人数	死亡 人数	生存 人年数	生存总 人年数	预期 寿命
$X \sim$	${}_n P_x$	${}_n D_x$	${}_n m_x$	${}_n q_x$	$l_x$	${}_n d_x$	${}_n L_x$	$T_x$	$e_x$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
0~	30005 *	429	—	0.014298	100000	1430	98785	6895566	68.96
1~	86920	105	0.001208	0.004820	98570	475	393330	6796781	68.95
5~	102502	81		0.003942	98095	387	489508	6403451	65.28
10~	151494	113	0.000746	0.003723	97708	364	487630	5913943	60.53
15~	182932	157	0.000858		97344	417	485678	5426313	55.74
20~	203107	215	0.001059	0.005281	96927	512	483355	4940635	50.97
25~	190289	221			96415	558	480680	4457280	46.23
30~	147076	181	0.001231	0.006136		588	477815	3976600	41.48
35~	99665	160	0.001605	0.007993	95269			3498785	36.73
40~	90891	234	0.002575	0.012793	94508	1209	469518	3024342	32.00
45~	105382	417	0.003957						
50~	86789	602	0.006936	0.034089	91471	3118	449560	2092899	22.88
55~	69368	919	0.013248	0.064116	88353	5665	427603	1643339	18.60
60~	51207	1328	0.025934	0.121775	82688	10069	388268	1215736	14.70
65~	39112	1691	0.043235	0.195088	72619	14167	327678	827468	11.39
70~	20509	1561	0.076113	0.319727	58452	18689	245538	499790	8.55
75~	9301	1126	0.121062	0.464674	39763	18477	152623	254252	6.39
80~	4297	900	0.209448	1.000000	21286	21286	101629	101629	4.77

表 15.3 某地 1982 年男性去肿瘤死因寿命表

年龄 (岁)	死亡 人数	肿瘤死 亡人数	${}_n r_x$	全死因寿命 表生存概率	去肿瘤死亡寿命表				
					${}_n P_x^*$	$l_x^*$	${}_n L_x^*$	$T_x^*$	$e_x^*$
$X \sim$	${}_n D_x$	${}_n D_x^*$	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
0~	429	2	0.995338	0.985702	0.985768	100000	98790	7130335	71.30
1~	105	4		0.995180	0.995363	98577	393394	7031545	71.33
5~	81	8		0.996058	0.996447	98120	489728	6638151	67.65
10~	113	11	0.902655	0.996277	0.996639	97771	488033	6148423	62.89
15~	157	13	0.917197	0.995719		97442	486253	5660390	58.09
20~	215	21	0.902326	0.994719	0.995234		484138	5174137	53.31
25~	221	36	0.837104	0.994212	0.995153	96596	481810	4689999	48.55
30~	181	41	0.773481	0.993864	0.995251	96128	479498	4208189	43.78
35~	160	44	0.725000	0.992007					
40~	234	80	0.658120	0.987207	0.991562	95116	473573	3251723	34.19
45~	417	142	0.659472	0.980409	0.987037	94313	468508	2778150	29.46
50~	602	210	0.651163	0.965911	0.977668	93090	460253	2309642	24.81
55~	919	315	0.657236	0.935884	0.957384	91011	445358	1849389	
60~	1328	360	0.728916	0.878225	0.909690	87132		1404031	16.11
65~	1691	381	0.774690	0.804912	0.845248	79263	365650	988043	12.47
70~	1561	248	0.841127	0.680273	0.723211	66997	288625		9.29
75~	1126	127	0.887211	0.535326	0.574417	48453	190713	333768	6.89
80~	900	64	0.928889	0.000000	0.000000	27832	143055	143055	5.14

## 第十六章 随访资料的生存分析

[返回](#)

### 一、名词解释

1. 生存分析(survival analysis)
2. 生存时间(survival time)
3. 完全数据(complete data)
4. 截尾数据(censored data)
5. 生存率(survival rate)
6. 生存概率(survival probability)

### 二、单选题

1. 进行生存分析时, \_\_\_\_\_ 不属截尾资料
  - A. 随访结束时仍存活者
  - B. 随访期内找不到者
  - C. 随访期内死于本病者
  - D. 随访期内死于其他病因者
  - E. 以上均不是
2. 生存分析中的生存时间是指
  - A. 观察开始至终止的时间
  - B. 观察开始至失访的时间
  - C. 确诊至死亡的时间
  - D. 手术至死亡的时间
  - E. 出院至失访的时间
3. 生存曲线下下降的坡度越陡, 表示
  - A. 生存时间越短
  - B. 生存时间越长
  - C. 生存概率越大
  - D. 与生存时间无关
  - E. 以上均不是
4. 随访资料作生存分析的条件是
  - A. 有一定的例数
  - B. 有一定的死亡数
  - C. 自变量取值不随时间变化
  - D. 自变量的作用大小不随时间变化
  - E. 死亡比例不能太小
5. 不宜作生存分析的情况是
  - A. 一批女性乳腺癌患者手术之后的存活时间
  - B. 一批患者确诊肝炎后肝功能异常的持续时间
  - C. 一批患者胃出血的持续时间
  - D. 一次体格检查中查出肿瘤患者的时间
  - E. 一批妇女宫内节育器在宫内的保存时间
6. Cox 回归的因变量是
  - A. 生存时间
  - B. 结局变量
  - C. 生存时间和结局变量
  - D. 完全数据
  - E. 正态和方差齐性
7. Cox 回归方法属于
  - A. 参数法
  - B. 非参数法
  - C. 半参数法
  - D. logistic 回归
  - E. 以上均不是
8. Kaplan-Meier 法属于
  - A. 参数法
  - B. 非参数法
  - C. 半参数法
  - D. logistic 回归
  - E. 以上均不是
9. log-rank 检验属于



- A. 参数法            B. 非参数法            C. 半参数法  
D. logistic 回归      E. 以上均不是
10. 生存分析中，描述生存时间的集中趋势宜用
- A. 算术平均数      B. 几何均数      C. 中位数  
D. 百分位数          E. 众数

### 三、简述题

1. 生存时间数据有哪几类？为何会产生截尾数据？
2. 在非参数法中，生存率之间比较的方法有哪几种？各有何特点？
3. 随访资料记录的主要内容有哪些？分析随访资料生存时间的方法主要有哪几种？

### 四、计算题

某疾病受吸烟（ $X_1$ ）和饮酒（ $X_2$ ）的影响，其中， $X_1=1$  为吸烟， $X_1=0$  为不吸烟； $X_2=1$  为饮酒， $X_2=0$  为不饮酒。建立了两种因子的 COX 模型，经分析得  $X_1$  与  $X_2$  的回归系数分别为  $\beta_1=0.8755$ ， $\beta_2=0.5108$ 。

- (1) 试说明  $\beta_1$  与  $\beta_2$  的意义。
- (2) 试写出 COX 模型。
- (3) 试求既吸烟又饮酒者发病的危险度与既不吸烟又不饮酒者发病的危险度之比，即比例风险度。

## 单项选择题参考答案

[返回](#)

第一章 1-5 CBADD 6-10 CCADC 11-14 CDCE

第二章, 第三章 1-5 CDCAE 6-10 DBBEC 11-15 CBBDC 16-20 EECDC  
21-25 AEEBC

第四章, 第五章 1-5 CDEBD 6-10 DADBD 11-15 EBBAC 16-20 CCCDA  
21-25 AEEDA 26-30 DECDD 31-32 CD

第六章 1-5 BABAC 6-10 AACCA 11-14 DBBC

第七章, 第八章, 第九章 1-5 AEDAC 6-10 CBCAB 11-15 CDBBC 16-20 ABBBB  
21-25 ADDBE 26-30 CEECB 31-35 AACDA 36-40 BABAB  
41-45 CBDAC

第十章 1-5 DBECC 6-10 CCBAC 11-15 CDDBB 16-20 BADDE 21-25 DBCCD  
26-30 CCCBE 31 B

第十二章、第十三章、第十四章 1-5 BDDAD 6-10 CCCEC

第十五章 1-2 BC

第十六章 1-5 EDADB 6-8 DCB

第十七章 1-5 CAABD 6-10 CCBBC

整理者: 钟文明  
2006年3月24日