

医学统计学·形成性考核任务一（第1-2章 25%）

1. 名词解释

匹配题 (20.0 分) (计分规则:按匹配正确项计分) (难易度:中)

提示	答案	答案池
总体		测量值与事实真相之间的差值。
误差		是指根据研究目的确定的研究对象的全体。
抽样研究		又称全距，是所有观察值中最大值和最小值之差。
极差		是指以样本特征推论总体特征的研究。
变异系数		是标准差与均数的比值。

正确答案:

总体

是指根据研究目的确定的研究对象的全体。

误差

测量值与事实真相之间的差值。

抽样研究

是指以样本特征推论总体特征的研究。

极差

又称全距，是所有观察值中最大值和最小值之差。

变异系数

是标准差与均数的比值。

答案解析:

暂无

2. 下面有关抽样误差的叙述，正确的是（ ）。

单选题 (2.0 分) (难易度:中)

- A. 严格设计和严格实施的研究可以避免抽样误差
- B. 样本量越大，抽样误差越大
- C. 抽样误差是由于测量人员测量技术不合格导致的误差
- D. 抽样误差与研究特征的个体差异有关

正确答案: D

答案解析: 暂无

3. “是否吸烟”的变量类型是（ ）。

单选题 (2.0 分) (难易度:中)

- A. 数值型变量资料
- B. 多分类变量资料
- C. 等级资料
- D. 二分类变量资料

正确答案: D

答案解析: 暂无

4. 下面关于样本量的陈述，正确的是（ ）。

单选题 (2.0 分) (难度:中)

- A. 样本量与总体规模有关
- B. 抽样误差与样本量无关
- C. 样本量与应答率水平无关
- D. 样本量需要专门的公式估计

正确答案: D

答案解析: 暂无

5. 下面关于研究对象的陈述, 错误的是 ( )。

单选题 (2.0 分) (难度:中)

- A. 研究对象与研究目的有关
- B. 研究对象可以是人, 也可以是动物
- C. 研究对象不需要来自研究总体
- D. 研究对象是研究设计的内容

正确答案: C

答案解析: 暂无

6. 下面有关总体的叙述, 正确的是 ( )。

单选题 (2.0 分) (难度:中)

- A. 总体是由根据研究目的所确定的全部研究对象
- B. 总体与研究目的无关
- C. 总体由样本量决定
- D. 总体由统计分析方法决定

正确答案: A

答案解析: 暂无

7. 下列选项中, 属于数值变量的是 ( )。

单选题 (2.0 分) (难度:中)

- A. 民族
- B. 体重
- C. 血型
- D. 性别

正确答案: B

答案解析: 暂无

8. 数据录入时, 部分数据录入有误, 误差的类型属于 ( )。

单选题 (2.0 分) (难度:中)

- A. 样本与总体之差
- B. 系统误差
- C. 随机测量误差
- D. 抽样误差

正确答案: B

答案解析: 暂无

9. 统计量是 ( )。

单选题 (2.0 分) (难度度:中)

- A. 统计总体数据得到的量
- B. 反映总体特征的的量
- C. 使用样本数据计算出来的统计指标
- D. 使用参数估计出来的

正确答案: c

答案解析: 暂无

10. 某病房记录了50名病人的护理等级，其中特级护理1名，一级护理3名，二级护理12名，三级护理34名，此资料属于（     ）。

单选题 (2.0 分) (难度度:中)

- A. 分类变量资料
- B. 二分类资料
- C. 有序分类变量资料
- D. 数值变量资料

正确答案: c

答案解析: 暂无

11. 下面有关误差的叙述，错误的是（     ）。

单选题 (2.0 分) (难度度:中)

- A. 随机误差不可避免
- B. 系统误差一定要避免发生
- C. 抽样误差包含个体差异
- D. 因为样本含量越大，抽样误差越小，所样本含量越大越好

正确答案: D

答案解析: 暂无

12. 某药物临床试验数据的两端均没有确定数值，描述其中心位置适用的最佳指标是（     ）。

单选题 (2.0 分) (难度度:中)

- A. 中位数
- B. 几何均数
- C. 均数
- D. 第55百分位数

正确答案: A

答案解析: 暂无

13. 接种某种疫苗一个月后，测定8名患者的抗体滴度分别为1：10、1：10、1：20、1：20、1：40、1：80、1：80、1：160，表示平均滴度的最佳指标是（     ）。

单选题 (2.0 分) (难度度:中)

- A. 均数
- B. 几何均数
- C. 中位数
- D. 百分位数

正确答案: B

答案解析：暂无

14. 对于正态分布的资料，描述分散程度的适宜指标是（     ）。

单选题 (2.0 分)（难度度:中）

- A. 标准差
- B. 离均差平方和
- C. 极差
- D. 四分位数间距

正确答案：A

答案解析：暂无

15. 已知某数据呈明显单峰偏态分布，描述其分散程度的最佳统计指标是（     ）。

单选题 (2.0 分)（难度度:中）

- A. 全距
- B. 标准差
- C. 变异系数
- D. 四分位数间距

正确答案：D

答案解析：暂无

16. 频数表中的组段的下限是（     ）。

单选题 (2.0 分)（难度度:中）

- A. 最后一个组段的终点
- B. 第一个组段的起点
- C. 每个组段的起点
- D. 每个组段的终点

正确答案：C

答案解析：暂无

17. 正偏态分布是指（     ）。

单选题 (2.0 分)（难度度:中）

- A. 集中位置偏向一侧，两侧频数分布不对称
- B. 集中位置在正中，左右两侧频数分布大体对称
- C. 集中位置偏向数值小的一侧
- D. 集中位置偏向数值大的一侧

正确答案：C

答案解析：暂无

18. 某校6岁女童身高均数为114cm，标准差为10cm，下列正确的选项是（     ）。

单选题 (2.0 分)（难度度:中）

- A. 5%的6岁女童身高 $\leq$ 94cm
- B. 5%的6岁女童身高 $\geq$ 94cm
- C. 2.5%的6岁女童身高 $\geq$ 133.6cm
- D. 2.5%的6岁女童身高 $\geq$ 124cm

正确答案：C

答案解析：暂无

19. 标准正态分布曲线的特征是（     ）。

单选题 (2.0 分)（难度度:中）

A.  $\mu=0$ ， $\sigma=1$

B.  $\mu=1$ ， $\sigma=0$

C.  $\mu=1$ ， $\sigma=1$

D.  $\mu=0$ ， $\sigma=0$

正确答案：A

答案解析：暂无

20. 下列有关均数和标准差的叙述，正确的是（     ）。

单选题 (2.0 分)（难度度:中）

A. 标准差越大，均数对各变量值的代表性越好

B. 标准差越小，均数对各变量值的代表性越好

C. 均数越大，标准差越小

D. 均数越小，标准差越大

正确答案：B

答案解析：暂无

21. 某组数据26例，变量的平方和以及变量的和分别是675和130，则均数和标准差分别为（     ）。

单选题 (2.0 分)（难度度:中）

A. 5.1、25

B. 26、5.8

C. 5、1

D. 5、2.7

正确答案：C

答案解析：暂无

22. 请简述均数和中位数的适用条件。

简答题 (10.0 分)（难度度:中）

答案解析：

答：算术均数简称均数，适于描述正态分布或近似正态分布资料的中心位置（5分）。中位数可表示任意分布资料的中心位置。表示偏离正态分布的单峰对称分布、单峰偏态分布以及分布末端无确切数值的单峰分布资料中心位置的最佳指标为中位数。（5分）

23. 请简述正态分布的特点。

简答题 (10.0 分)（难度度:中）

答案解析：

答：正态分布曲线具有如下特征：

- (1) 形态  
正态曲线呈钟形分布，高峰位于中央，高峰所在位置为分布的算术均数。分布以均数为中心向两端逐渐降低，完全对称于算术均数。曲线在横轴上方与横轴不相交。（3分）
- (2) 参数  
正态分布的两个参数是均数 $\mu$ （位置参数）和标准差 $\sigma$ （形状参数）。标准差 $\sigma$ 取值不变时，分布随均数 $\mu$ 的增大右移；反之左移。均数 $\mu$ 取值不变时，分布随标准差 $\sigma$ 增大，曲线变“矮胖”，反之变“瘦高”。（3分）
- (3) 正态曲线下面积的规律  
1)  $X$ 轴与正态曲线之间围绕的总面积为1（100%）。  
2) 区间 $(\mu-\sigma, \mu+\sigma)$ 内，曲线下的面积为68.27%，区间 $(\mu-1.96\sigma, \mu+1.96\sigma)$ 内，曲线下的面积为95%，区间 $(\mu-2.58\sigma, \mu+2.58\sigma)$ 内，曲线下的面积为99%。区间 $(\mu-1.65\sigma, +\infty)$ 和区间 $(-\infty, \mu+1.65\sigma)$ 的面积均为95%。（4分）

24. 计算题

综合题 (10.0 分)

- (1) 现测得A厂10个批次倍他米松片的质量指数分别为0.8652、0.6905、0.7265、0.6959、0.6706、0.4888、0.7419、0.5157、0.5865、0.7917。试计算均数、中位数和标准差。

简答题 (10.0 分)（难易度:中）

答案解析：  
答：（1）利用直接法的公式计算均数：

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{0.8652 + 0.6905 + \dots + 0.7917}{10} = \frac{6.773}{10} = 0.6773 \quad (3分)$$

- (2) 本研究中样本量为10，是偶数，将10个数字由小到大排列，排第5和第6的数值之和的一半，即为中位数。计算公式为：

$$M = \frac{1}{2} \left( X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)} \right) = \frac{1}{2} \left( X_{(5)} + X_{(6)} \right) = \frac{1}{2} (0.6905 + 0.6959) = 0.6932 \quad (3分)$$

- (3)  $\sum X^2 = 4.7132$ ， $\sum X = 6.773$

$$S = \sqrt{\frac{\sum X^2 - (\sum X)^2/n}{n-1}} = \sqrt{\frac{4.7132 - 6.773^2/10}{10-1}} = 0.1183$$

标准差为0.1183  
(4分)

25. 计算题

综合题 (10.0 分)

- (1) 某研究者用蒸汽杀青法和生晒法分别测量10份样品中连翘苷的含量，算得的均数分别为0.631 $\mu$ g/mL和0.095  $\mu$ g/mL，标准差分别为0.05587  $\mu$ g/mL和0.01958  $\mu$ g/mL，试比较两种方法测得的连翘苷含量的分散程度。

简答题 (10.0 分)（难易度:中）

答案解析：  
答：两组均数相差较大，故选择变异系数作为比较两组数据分散程度的指标。变异系数的计算公式为： $CV = \frac{S}{\bar{X}} \times 100\%$ 。  
两种方法测得连翘苷含量的变异系数分别为：  
蒸汽杀青法： $(0.05587/0.631) \times 100\% = 8.85\%$

生晒法：  $(0.01958/0.095) \times 100\% = 20.61\%$

因此蒸汽杀青法测得连翘苷含量的分散程度低于生晒法。

医学统计学·形成性考核任务二（第3-4章 25%）

1. 名词解释

匹配题 (20.0 分) (计分规则:按匹配正确项计分) (难易度:中)

提示	答案	答案池
相对数		即样本统计量的标准差，反映抽样误差的大小。
构成比		是由两个有联系的指标之比所构成的统计指标。常用的相对数包括率、构成比和相对比。
标准误		也称显著性水准，是预先规定的判断小概率事件的概率尺度，记为 $\alpha$ ，通常取 $\alpha=0.05$ 或 $\alpha=0.01$ 。
置信区间		按照一定的可信度 $(1-\alpha)$ ，估计包含未知总体参数的一个区间范围，即为置信区间。
检验水准		又称构成指标，它表示事物内部每一组成部分在整体中所占的比重。

正确答案:

相对数  
是由两个有联系的指标之比所构成的统计指标。常用的相对数包括率、构成比和相对比。

构成比  
又称构成指标，它表示事物内部每一组成部分在整体中所占的比重。

标准误  
即样本统计量的标准差，反映抽样误差的大小。

置信区间  
  
按照一定的可信度  $(1-\alpha)$  ，估计包含未知总体参数的一个区间范围，即为置信区间。

检验水准  
也称显著性水准，是预先规定的判断小概率事件的概率尺度，记为 $\alpha$ ，通常取 $\alpha=0.05$ 或 $\alpha=0.01$ 。

答案解析:  
暂无

2. 说明某现象发生强度的指标是（ ）。

单选题 (2.0 分) (难易度:中)

- A. 率
- B. 相对比
- C. 构成比
- D. 增长速度

正确答案: A  
答案解析: 暂无

3. 构成比所有组成部分的和（ ）。

单选题 (2.0 分) (难易度:中)

- A. 必小于1
- B. 必等于1
- C. 必大于1



D. 必等于100

正确答案： B

答案解析： 暂无

4. 某年两所企业不同年龄段高血压患病率，均为A企业高于B企业，但总的患病率却是B企业高于A企业，请问出现这种不同的可能原因是（ ）。

单选题 (2.0 分) (难度度:中)

- A. 两所企业病人的严重程度不同
- B. 两所企业的不同年龄段人数构成不同
- C. 两所企业的总人数不同
- D. 两所企业领导构成不同

正确答案： B

答案解析： 暂无

5. 随机抽取某地300名男性和150名女性为研究对象，结果发现男性和女性某病患病率分别为40%和30%，则合并患病率为（ ）。

单选题 (2.0 分) (难度度:中)

- A. 33.3%
- B. 36.7 %
- C. 35.0%
- D. 41.2%

正确答案： B

答案解析： 暂无

6. 下列选项中，不受年龄结构影响的指标是（ ）。

单选题 (2.0 分) (难度度:中)

- A. 粗死亡率
- B. 年龄别死亡率
- C. 标准化死亡率
- D. 死因构成

正确答案： C

答案解析： 暂无

7. 下列有关标化后的总患病率的叙述，正确的选项是（ ）。

单选题 (2.0 分) (难度度:中)

- A. 反映了患病的实际水平
- B. 不随所用标准的变化而变化
- C. 反映了事物实际发生的强度
- D. 反映了一种相对水平，仅可作为比较的基础

正确答案： D

答案解析： 暂无

8. 动态数列中定基比指标的类型是（ ）。

单选题 (2.0 分) (难度度:中)

- A. 频率
- B. 构成比

C. 相对比

D. 绝对数

正确答案: C

答案解析: 暂无

9. 计算脊髓灰质炎疫苗接种后血清检查的阳转率, 分母应该是 ( )。

单选题 (2.0 分) (难度度:中)

A. 脊髓灰质炎患儿

B. 脊髓灰质炎易感患儿

C. 脊髓灰质炎疫苗接种人数

D. 脊髓灰质炎疫苗接种后阳转人数

正确答案: C

答案解析: 暂无

10. 下列有关率的描述, 正确的是 ( )。

单选题 (2.0 分) (难度度:中)

A. 分母无论大小都可以计算率

B. 率是说明某现象发生的频率或强度的指标

C. 率表示事物内部各部分的比重大小

D. 以 $m/n$ 表示率, 可得出 $m$ 是 $n$ 的几分之几

正确答案: B

答案解析: 暂无

11. 动态数列中的绝对增长量指标是 ( )。

单选题 (2.0 分) (难度度:中)

A. 构成比

B. 相对比

C. 绝对数

D. 频率

正确答案: C

答案解析: 暂无

12. 统计推断的内容包括 ( )。

单选题 (2.0 分) (难度度:中)

A. 区间估计和点估计

B. 参数估计与假设检验

C. 统计预测和统计控制

D. 统计描述和统计图表

正确答案: B

答案解析: 暂无

13.  $|t| > t_{0.05/2, \nu}$  统计上认为 ( )。

单选题 (2.0 分) (难度度:中)

A. 两总体均数差别无统计学意义

B. 两样本均数差别无统计学意义

C. 两总体均数差别有统计学意义

D. 两样本均数差别有统计学意义

正确答案: C

答案解析: 暂无

14. 配对  $t$  检验的备择假设（双侧检验）可写为（ ）。

单选题 (2.0 分)（难度度:中）

A.  $\mu=\mu_0$

B.  $\mu_d=0$

C.  $\mu_d\neq 0$

D.  $\mu_1\neq\mu_2$

正确答案: C

答案解析: 暂无

15. 两样本均数比较，假设检验结果  $P < 0.05$  说明（ ）。

单选题 (2.0 分)（难度度:中）

A. 两总体均数的差别较小

B. 两总体均数的差别较大

C. 可以认为两总体无差别

D. 可以认为两总体有差别

正确答案: D

答案解析: 暂无

16. 两样本均数比较,  $P < 0.05$  时,  $P$  值越小说明（ ）。

单选题 (2.0 分)（难度度:中）

A. 两样本均数差别越大

B. 两总体均数差别越大

C. 越有理由认为两样本均数不同

D. 越有理由认为两总体均数不同

正确答案: D

答案解析: 暂无

17. 为研究新旧两种仪器测量血生化指标的差异，分别用这两台仪器测量同一批样品，可考虑采用的假设检验方法为（ ）。

单选题 (2.0 分)（难度度:中）

A. 成组设计  $t$  检验

B. 成组设计  $Z$  检验

C. 配对设计  $t$  检验

D. 配对设计  $\chi^2$  检验

正确答案: C

答案解析: 暂无

18. 方差分析的应用条件之一是方差齐性,它是指（ ）。

单选题 (2.0 分)（难度度:中）

A. 各比较组相应的样本方差相等

- B. 各比较组相应的总体方差相等
- C. 组内方差=组间方差
- D. 总方差=组内方差+组间方差

正确答案： B

答案解析： 暂无

19. 随机抽取某地300名健康成年女性的血红蛋白量，则其总体均数95%置信区间为（     ）。

单选题 (2.0 分) （难度度:中）

- A.  $\bar{p} \pm 2.58s$
- B.  $\bar{x} \pm 1.96s_x$
- C.  $p \pm 1.96s_p$
- D.  $\bar{x} \pm 2.58s_x$

正确答案： B

答案解析： 暂无

20. 表示均数抽样误差大小的统计指标是（     ）。

单选题 (2.0 分) （难度度:中）

- A. 标准差
- B. 标准误
- C. 方差
- D. 变异系数

正确答案： B

答案解析： 暂无

21. 完全随机设计的两样本均数进行比较，当方差不齐时，不可选择的检验方法是（     ）。

单选题 (2.0 分) （难度度:中）

- A.  $t'$  检验
- B. 秩和检验
- C. 变量变换
- D.  $t$ 检验

正确答案： D

答案解析： 暂无

22. 应用相对数时的注意事项有哪些？

简答题 (10.0 分) （难度度:中）

答案解析：

- 答：（1）应注意资料的可比性；（2分）
- （2）样本含量（分母）太小时，不宜计算相对数，最好用绝对数来表示；（2分）
- （3）构成比和率的意义不同，应避免混淆；（2分）

- (4) 对各组观察例数不等的几个率，不能直接相加求总率；（2分）
- (5) 样本率之间的比较由于存在抽样误差，需要假设检验方可对总体做出推断结论。（2分）

23. 标准差与标准误之间的联系与区别是什么？

简答题 (10.0 分)（难度度:中）

答案解析：

答：标准差是描述一个变量的所有观察值与均数离散程度的指标。（3分）样本统计量的标准差称为标准误。样本均数的标准差也称均数的标准误，它反映样本均数间的离散趋势，也反映样本均数与总体均数间的差异，说明了均数抽样误差的大小。（3

分）在计算上可以用标准差估计标准误，均数标准误的估计值  $S_{\bar{x}} = \frac{S}{\sqrt{n}}$ ，可见均数标准误的大小与标准差成正比。（4分）

24. 计算题

综合题 (10.0 分)

- (1) 某研究者随机抽取25名高原地区成年男子进行检查，得到血红蛋白均数为155g/L，标准差25g/L。试估计该高原地区成年男子的血红蛋白总体均数的95%置信区间。

简答题 (10.0 分)（难度度:中）

答案解析：

答：该研究总体标准差  $\sigma$  未知但  $S$  已知， $n$  较小（如  $n=25$ ）（2分）， $v=25-1=24$ ， $\bar{X}=155\text{g/L}$ ， $S=25\text{g/L}$ ，查附表， $t_{0.05/2,24}=2.064$ （1分），因此95%的置信区

间为  $\bar{X} \pm t_{0.05/2,24} S_{\bar{x}} = \bar{X} \pm t_{0.05/2,24} \frac{S}{\sqrt{n}} = 155 \pm 2.064 \times 25 / \sqrt{25}$ ，即（144.68，165.32）g/L

（公式2分，数据代入正确1分，计算结果正确1分，单位1分）  
该高原地区成年男子的血红蛋白总体均数的95%置信区间为（144.68，165.32）g/L。（2分）

25. 计算题

综合题 (10.0 分)

- (1) 2. 25例糖尿病患者随机分成两组，甲组单纯用药物治疗，乙组采用药物治疗合并饮食疗法，两个月后测空腹血糖（mmol/L）如下表所示，试分析两种疗法治疗后患者血糖值是否相同？

25 名糖尿病患者两种疗法治疗后二个月血糖值(mmol/L)			
编号	甲组血糖值（ $X_1$ ）	编号	乙组血糖值（ $X_2$ ）
1	8.4	1	5.4
2	10.5	2	6.4
3	12.0	3	6.4
4	12.0	4	7.5
5	13.9	5	7.6
6	15.3	6	8.1
7	16.7	7	11.6
8	18.0	8	12.0
9	18.7	9	13.4
10	20.7	10	13.5
11	21.1	11	14.8
12	15.2	12	15.6
		13	18.7

简答题 (10.0 分)（难度度:中）

答案解析:

答: 两种疗效治疗效果的比较可以进行两独立样本资料的  $t$  检验, 首先要看方差是否齐, 进行方差齐性检验。本研究中  $\bar{X}_1=15.21$ ,  $\bar{X}_2=10.85$ ,  $S_1=4.22$ ,

$$S_2=4.02, n_1=12, n_2=13$$

(1) 方差齐性检验

1) 建立检验假设, 确定检验水准

$H_0: \sigma_1^2 = \sigma_2^2$ , 两总体方差相等。

$H_1: \sigma_1^2 \neq \sigma_2^2$ , 两总体方差不相等。

$\alpha=0.05$

(1分)

2) 计算检验统计量

$$F = \frac{S_{\text{较大}}^2}{S_{\text{较小}}^2} = \frac{4.22^2}{4.02^2} = 1.102$$

$$\nu_{\text{较大}} = n_2 - 1 = 13 - 1 = 12, \nu_{\text{较小}} = n_1 - 1 = 12 - 1 = 11$$

(1分)

3) 确定  $P$  值, 作出推断结论

查  $F$  界值表 (方差齐性检验用),  $F_{0.05/2, (12, 11)} = 3.43$ , 由于  $F < F_{0.05/2, (12, 11)}$ , 故  $P > 0.05$ , 按  $\alpha=0.05$  的检验水准, 不拒绝  $H_0$ , 可以认为两总体方差相等。(1分)

因此可以用两独立样本资料的  $t$  检验来比较两种疗效的治疗效果。

(2) 两独立样本资料的  $t$  检验

1) 建立检验假设, 确定检验水准

$H_0: \mu_1 = \mu_2$ , 两种疗法治疗后患者血糖值相同

$H_1: \mu_1 \neq \mu_2$ , 两种疗法治疗后患者血糖值不相同

$\alpha=0.05$

(2分)

2) 计算检验统计量

$$S_c^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = \frac{(12 - 1) \times 4.22^2 + (13 - 1) \times 4.02^2}{12 + 13 - 2} = 17.02$$

(1分)

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{\bar{X}_1 - \bar{X}_2}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_c^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{15.21 - 10.85}{\sqrt{17.02 \times \left( \frac{1}{12} + \frac{1}{13} \right)}} = 2.64$$

(2分)

3) 确定  $P$  值, 作出推断结论

自由度  $\nu = 12 + 13 - 2 = 23$ , 查  $t$  界值表,  $t_{0.05/2, 23} = 2.069$ ,  $t > t_{0.05/2, 23}$ , 故  $P < 0.05$ , 差异有统计学意义, 可以认为两种疗法治疗后患者血糖值不同。(2分)

医学统计学·形成性考核任务三（第5，7章 25%）

1. 名词解释

匹配题 (20.0 分) (计分规则:按匹配正确项计分) (难易度:中)

提示	答案	答案池
$\chi^2$ 检验		也称卡方检验，是对分类变量资料进行假设检验的统计学方法。
理论频数		当行数或列数大于2时，称为行×列表或 $R\times C$ 表。行×列表 $\chi^2$ 检验主要用于多个独立样本率或多个独立构成比之间的比较。
行×列表		
统计表		是以点、线、面等几何图形客观展示数据的分布、水平、构成及关系等特征的重要工具。与统计表相比，统计图更直观。
统计图		第 $R$ 行第 $C$ 列的理论频数记为 $T_{RC}$ ，  计算公式为 $T_{RC}=\frac{n_{R.}n_{.C}}{n}$ ，其中 $n_{R.}$ 表示第 $R$ 行实际数的行合计， $n_{.C}$ 表示第 $C$ 列实际数的列合计， $n$ 表示总例数。  是以表格的形式客观展示数据、数据分析过程及统计分析结果的重要工具。与文字叙述相比，统计表更加直观，可提供更多的原创信息。

正确答案:

$\chi^2$ 检验  
也称卡方检验，是对分类变量资料进行假设检验的统计学方法。  
理论频数

第 $R$ 行第 $C$ 列的理论频数记为 $T_{RC}$ ，计算公式为  $T_{RC}=\frac{n_{R.}n_{.C}}{n}$  ，其中 $n_{R.}$ 表示第 $R$ 行实际数的行合计， $n_{.C}$ 表示第 $C$ 列实际数的列合计， $n$ 表

示总例数。

行×列表

当行数或列数大于2时，称为行×列表或 $R\times C$ 表。行×列表 $\chi^2$ 检验主要用于多个独立样本率或多个独立构成比之间的比较。

统计表

是以表格的形式客观展示数据、数据分析过程及统计分析结果的重要工具。与文字叙述相比，统计表更加直观，可提供更多的原创信息。

统计图

是以点、线、面等几何图形客观展示数据的分布、水平、构成及关系等特征的重要工具。与统计表相比，统计图更直观。

答案解析:

暂无

2. 三个独立样本率比较的 $\chi^2$ 检验，若 $\chi^2>\chi^2_{0.05,2}$ ，统计结论为（ ）。

单选题 (2.0 分) (难易度:中)

- A. 各样本率均不相同
- B. 各总体率均不相同
- C. 各样本率不同或不全相同
- D. 各总体率不同或不全相同

正确答案：D

答案解析：暂无

3. 某医生用甲药治疗15例病人，治愈8人；用乙药治疗20例病人，治愈2人。比较两药疗效时，可选用的最佳方法是（ ）。

单选题 (2.0 分)（难度度:中）

- A.  $\chi^2$ 检验
- B.  $t$ 检验
- C. 校正 $\chi^2$ 检验
- D. Fisher确切概率法

正确答案： D

答案解析：暂无

4. 两独立样本设计四格表资料中，如果行合计和列合计不变，实际频数发生变化，则理论频数相应（ ）。

单选题 (2.0 分)（难度度:中）

- A. 增大
- B. 减小
- C. 不变
- D. 先增大后减小

正确答案： C

答案解析：暂无

5. 两独立样本设计四格表资料 $\chi^2$ 检验校正公式的应用条件是（ ）。

单选题 (2.0 分)（难度度:中）

- A.  $n \geq 40$ 且 $1 \leq T_{\min} < 5$
- B.  $T_{\min} \geq 5$
- C.  $n \geq 40$ 或 $T_{\min} \geq 5$
- D.  $n \geq 40$ 且  $T_{\min} \geq 5$

正确答案： A

答案解析：暂无

6. 四格表资料的自由度是（ ）。

单选题 (2.0 分)（难度度:中）

- A. 一定不等于1
- B. 一定等于1
- C. 等于行数 $\times$ （列数-1）
- D. 等于样本含量-1

正确答案： B

答案解析：暂无

7. 多个样本率的假设检验，其无效假设是（ ）。

单选题 (2.0 分)（难度度:中）

- A.  $m_1 = m_2 = \dots = m_n$
- B.  $P_1 = P_2 = \dots = P_n$
- C.  $X_1 = X_2 = \dots = X_n$
- D.  $\pi_1 = \pi_2 = \dots = \pi_n$



正确答案：D

答案解析：暂无

8. 配对资料的卡方检验，其备择假设是（ ）。

单选题 (2.0 分) (难度度:中)

A.  $P_1=P_2$

B.  $P_1\neq P_2$

C.  $\pi_1\neq \pi_2$

D.  $\pi_1=\pi_2$

正确答案：C

答案解析：暂无

9. 采用配对计数资料的卡方检验时，若 $b+c<40$ 时，下列方法中正确的是（ ）。

单选题 (2.0 分) (难度度:中)

A. 校正 $\chi^2$ 检验

B. 不校正 $\chi^2$ 检验

C. 增加 $n$ 后再做检验

D. 确切概率法

正确答案：A

答案解析：暂无

10. 两个四格表资料的卡方检验，资料1的结果为 $\chi^2>\chi^2_{0.01,1}$ ，资料2的结果为 $\chi^2>\chi^2_{0.05,1}$ ，下列选项正确的是（ ）。

单选题 (2.0 分) (难度度:中)

A. 资料1两个百分数相差较大

B. 资料2两个百分数相差较大

C. 资料1更有理由认为两总体率不同

D. 资料2更有理由认为两总体率不同

正确答案：C

答案解析：暂无

11. 对 $5\times 4$ 表进行 $\chi^2$ 检验，当有5个格子的理论频数 $1<T<5$ 时，下列措施正确的是（ ）。

单选题 (2.0 分) (难度度:中)

A. 仍做 $\chi^2$ 检验

B. 作校正的 $\chi^2$ 检验

C. 应进行适当的合并

D. 增加样本例数

正确答案：C

答案解析：暂无

12. 可以用来描述身高与年龄关系的统计图是（ ）。

单选题 (2.0 分) (难度度:中)

A. 直条图

B. 直方图

C. 散点图

D. 普通线图

正确答案: C

答案解析: 暂无

13. 适宜绘制线图的资料类型是（ ）。

单选题 (2.0 分) (难度度:中)

- A. 体重和血压之间的关系
- B. 数值型变量资料的频数分布
- C. 10年高血压患病率
- D. 研究对象的受教育程度

正确答案: C

答案解析: 暂无

14. 可以表示原始数据测量值的统计图是（ ）。

单选题 (2.0 分) (难度度:中)

- A. 直条图
- B. 直方图
- C. 茎叶图
- D. 箱式图

正确答案: C

答案解析: 暂无

15. 可用于比较率的长期下降趋势的统计图是（ ）。

单选题 (2.0 分) (难度度:中)

- A. 直条图
- B. 直方图
- C. 半对数线图
- D. 茎叶图

正确答案: C

答案解析: 暂无

16. 下面有关直条图的叙述，正确的是（ ）。

单选题 (2.0 分) (难度度:中)

- A. 直条图纵坐标的取值必须始于0
- B. 直条图的纵坐标可始于任何数值
- C. 直条图各直条的宽度不必相等
- D. 直条图是用来表示数值型变量资料频数分布的统计图形

正确答案: A

答案解析: 暂无

17. 统计图中可展示中位数的统计图是（ ）。

单选题 (2.0 分) (难度度:中)

- A. 散点图
- B. 线图
- C. 箱式图
- D. 茎叶图

正确答案: c

答案解析: 暂无

18. 可以用以指明统计表内数字含义的选项是（     ）。

单选题 (2.0 分)（难度度:中）

- A. 标题
- B. 数字
- C. 标目
- D. 备注

正确答案: c

答案解析: 暂无

19. 用以说明统计表内容和选项是（     ）。

单选题 (2.0 分)（难度度:中）

- A. 备注
- B. 标题
- C. 标目
- D. 数字

正确答案: B

答案解析: 暂无

20. 根据某地7岁－12岁小学生的体质指数的调查资料，对调查者的身高分布进行统计描述，宜绘制的图形是（     ）。

单选题 (2.0 分)（难度度:中）

- A. 直条图
- B. 线图
- C. 直方图
- D. 散点图

正确答案: c

答案解析: 暂无

21. 表达某地两年几种疾病的患病率，宜绘制的图形是（     ）。

单选题 (2.0 分)（难度度:中）

- A. 直方图
- B. 百分直条图
- C. 复式直条图
- D. 散点图

正确答案: c

答案解析: 暂无

22. 简述两独立样本设计四格表资料 $\chi^2$ 检验的类型及应用条件。

简答题 (10.0 分)（难度度:中）

答案解析:

答：（1）当 $n \geq 40$ 且 $T_{\min} \geq 5$ 时，用两独立样本设计四格表资料 $\chi^2$ 检验的基本公式或专用基本公式计算 $\chi^2$ 统计量。（4分）

（2）当 $n \geq 40$ 且 $1 \leq T_{\min} < 5$ 时，用两独立样本设计四格表资料 $\chi^2$ 检验的校正公式或专用校正公式计算 $\chi^2$ 统计量。（3分）

（3）当 $n < 40$ ，或 $T_{\min} < 1$ 时，用两独立样本四格表资料的Fisher确切概率法获得 $P$ 值。（3分）

23. 简述统计表的制表原则。

简答题 (10.0 分) (难度度:中)

答案解析:

答：统计表的制表原则是重点突出（2分）、内容正确完整（2分）、结构简洁清晰（3分），便于阅读和数据信息的提取（3分）。

24. 计算题

综合题 (10.0 分)

（1）欲比较心电图和生化测定诊断低血钾的作用，某研究者分别采用两种方法对84名临床确诊的低血钾患者进行检查，结果如下表所示。试分析两种方法的检测结果有无差别。

两种方法检测低血钾的结果			
心电图	生化测定		合计
	+	-	
+	48	27	75
-	4	5	9
合计	52	32	84

简答题 (10.0 分) (难度度:中)

答案解析:

答：（1）建立检验假设，确定检验水准

$H_0: \pi_1 = \pi_2$ ，即两种方法的总体检出率相同  
 $H_1: \pi_1 \neq \pi_2$ ，即两种方法的总体检出率不同  
 $\alpha = 0.05$ （2分）

（2）计算检验统计量

已知  $b=27$ ， $c=4$ ， $b+c=31 < 40$ ，应采用 McNemar  $\chi^2$  检验的校正公式估计统计量

$$\chi_c^2 = \frac{(|b - c| - 1)^2}{b + c} = \frac{(|27 - 4| - 1)^2}{27 + 4} = 15.61$$
$$\nu = (2 - 1)(2 - 1) = 1 \quad (5 \text{ 分})$$

（3）确定  $P$  值，作出推断结论

查 $\chi^2$ 界值表， $\chi_{0.05,1}^2 = 3.84$ ，本例 $\chi^2 > \chi_{0.05,1}^2$ ，得  $P < 0.05$ ，按 $\alpha = 0.05$ 的检验水准，拒绝  $H_0$ ，认为两种方法的检出率不同，差别有统计学意义。（3分）

25. 制表题

综合题 (10.0 分)

(1) 某研究利用常规药为对照评价新药的疗效。服常规药228人中60人有效；服新药258人中80人有效。试将上述资料用规范的统计表表示。

简答题 (10.0 分) (难易度:中)

答案解析:

答：统计表显示如下。

服用新药和常规药的有效率				
药物	有效	无效	合计	有效率（%）
新药	80	178	258	31.01
常规药	60	168	228	26.32
合计	140	346	486	28.81

(评分标准：标题2分，线条正确2分，横标目正确2分，纵标目正确2分，数字正确2分)

医学统计学·形成性考核任务四（第6，8章 25%）

1. 名词解释

匹配题 (20.0 分) (计分规则:按匹配正确项计分) (难易度:中)

提示	答案	答案池
简单信息相关		指将同一批研究对象不同特征的数据进行合并的操作。
简单线性回归		用统计学模型定量地描述两个变量之间数量依存关系的方法。
Pearson相关系数		指把来自不同研究对象的数据文件合并为一个数据文件的操作。
添加变量的合并		描述两变量间线性相关方向与密切程度的值，Pearson总体相关系数用 <i>r</i> 表示，样本相关系数用 <i>r</i> 表示。
添加观测值的合并		用相关系数定量地描述两个变量之间关联强度与方向的方法。

正确答案:

简单信息相关

用相关系数定量地描述两个变量之间关联强度与方向的方法。

简单线性回归

用统计学模型定量地描述两个变量之间数量依存关系的方法。

Pearson相关系数

描述两变量间线性相关方向与密切程度的值，Pearson总体相关系数用*r*表示，样本相关系数用*r*表示。

添加变量的合并

指将同一批研究对象不同特征的数据进行合并的操作。

添加观测值的合并

指把来自不同研究对象的数据文件合并为一个数据文件的操作。

答案解析:

暂无

2. 对变量*X*和*Y*做线性相关分析时，资料需要符合的条件是（ ）。

单选题 (2.0 分) (难易度:中)

- A. *X*和*Y*有回归关系
- B. *X*服从正态分布
- C. *Y*服从正态分布
- D. *X*和*Y*服从双变量正态分布

正确答案: D

答案解析: 暂无

3. 线性回归系数*t*检验的自由度是（ ）。

单选题 (2.0 分) (难易度:中)

- A. *n*
- B. *n*-1
- C. *n*-2
- D. *n*-3

正确答案: c

答案解析：暂无

4. Pearson线性相关系数假设检验的无效假设 $H_0$ 是（ ）。

单选题 (2.0 分) (难度度:中)

A.  $r=0$

B.  $r\neq 0$

C.  $r=0$

D.  $r\neq 0$

正确答案：A

答案解析：暂无

5. 散点图中各散点基本呈直线共变趋势，部分数据点在拟合的直线上，且随 $X$ 值增大， $Y$ 值有增大趋势，说明两个变量之间可能有（ ）。

单选题 (2.0 分) (难度度:中)

A. 负相关关系

B. 正相关关系

C. 完全负相关关系

D. 完全正相关关系

正确答案：B

答案解析：暂无

6. 用最小二乘法估计线性回归模型参数的原则是，各观测点距回归直线的（ ）。

单选题 (2.0 分) (难度度:中)

A. 纵向距离之和最小

B. 垂直距离的平方和最小

C. 横向距离的平方和最小

D. 纵向距离的平方和最小

正确答案：D

答案解析：暂无

7. 对样本量 $n=10$ 的资料估计相关系数并进行假设检验，得 $t_r < t_{0.05/2, n-2}$ ，则下列各项中的最佳选项是（ ）。

单选题 (2.0 分) (难度度:中)

A. 两个变量之间不存在相关关系

B. 两个变量之间有线性相关关系的可能性小于5%

C. 尚不能认为两个变量之间存在相关关系

D. 两个变量之间的相关关系有统计学意义

正确答案：C

答案解析：暂无

8. 分析环境噪音强度与居民失眠患病率之间的关系可采用（ ）。

单选题 (2.0 分) (难度度:中)

A. Pearson相关分析

B. Spearman等级相关分析

C. 直线回归分析

D. 卡方检验

正确答案: B

答案解析: 暂无

9. 估计线性相关系数之前,应首先绘制的统计图是( )。

单选题 (2.0 分) (难度度:中)

A. 散点图

B. 饼图

C. 直方图

D. 直条图

正确答案: A

答案解析: 暂无

10. 下列各项中正确的是( )。

单选题 (2.0 分) (难度度:中)

A.  $r>0$ , 一定有  $b>0$

B.  $r>0$ , 一定有  $b<0$

C.  $r$ 的正负与 $b$ 的正负无关

D.  $r<0$ , 一定有  $b>0$

正确答案: A

答案解析: 暂无

11. 相关系数的取值范围是( )。

单选题 (2.0 分) (难度度:中)

A.  $r$ 取任意实数

B.  $-1< r \leq 1$

C.  $-1 \leq r \leq 1$

D.  $r$ 取非负数

正确答案: C

答案解析: 暂无

12. 下列有关 $b$ 和 $r$ 关系的叙述,正确的是( )。

单选题 (2.0 分) (难度度:中)

A. 同一样本 $b$ 和 $r$ 的显著性检验结果相同

B. 同一样本 $b$ 和 $r$ 的显著性检验结果不同

C. 同一样本 $b$ 和 $r$ 的显著性检验结果可能相同,也可能不同

D. 同一样本 $b$ 和 $r$ 的显著性检验结果是无关的

正确答案: A

答案解析: 暂无

13. 已知 $t_{b1}=t_{b2}$ ,则一定有( )。

单选题 (2.0 分) (难度度:中)

A.  $t_{r1}=t_{r2}$

B.  $b_1=b_2$

C.  $r_1=r_2$

D.  $a_1=a_2$



正确答案: A

答案解析: 暂无

14. 在数据中增加一个变量的操作所用到的菜单是（ ）。

单选题 (2.0 分) (难度:中)

A. Insert Variable

B. Insert Case

C. Go to Case

D. Weight Cases

正确答案: A

答案解析: 暂无

15. 在原有变量的基础之上，产生一个新变量的操作所用到的菜单是（ ）。

单选题 (2.0 分) (难度:中)

A. Sort Cases

B. Select Cases

C. Compute

D. Categorize Variables

正确答案: C

答案解析: 暂无

16. Recode菜单的功能是（ ）。

单选题 (2.0 分) (难度:中)

A. 建立一个新变量

B. 增加一个观察单位

C. 对变量重新赋值

D. 对数据进行加权

正确答案: C

答案解析: 暂无

17. SPSS默认的变量取值宽度是（ ）。

单选题 (2.0 分) (难度:中)

A. 12

B. 10

C. 8

D. 9

正确答案: C

答案解析: 暂无

18. 选取部分研究对象进行分析时的操作所用到的菜单是（ ）。

单选题 (2.0 分) (难度:中)

A. Match Cases

B. Active Cases

C. Split Cases

D. Select Cases

正确答案：D

答案解析：暂无

19. 直线回归的前提条件是（ ）。

单选题 (2.0 分) (难度度:中)

- A. 线性、独立、正态、等方差
- B. 线性、依赖、正态、等方差
- C. 线性、独立、偏态、等方差
- D. 非线性、独立、正态、等方差

正确答案：A

答案解析：暂无

20. 下列有关回归系数的叙述，正确的是（ ）。

单选题 (2.0 分) (难度度:中)

- A. 回归系数越大，两变量的关系越密切
- B. 回归系数不能反映两变量间的关系密切与否
- C. 回归系数越小，两变量的关系越密切
- D. 回归系数越大，两变量的关系越不密切

正确答案：B

答案解析：暂无

21. 样本相关系数r=0，说明（ ）。

单选题 (2.0 分) (难度度:中)

- A. 两变量间不可能存在直线关系
- B. 两变量间的关系不确定
- C. 两变量不存在任何关系
- D. 两变量必然存在曲线关系

正确答案：A

答案解析：暂无

22. 简述决定系数的意义。

简答题 (10.0 分) (难度度:中)

答案解析：

答：决定系数是回归平方和与总平方和的比值，用 $R^2$ 表示，计算公式为： $R^2 = \frac{SS_{\text{回}}}{SS_{\text{总}}}$ （4分）。决定系数 $R^2$ 无度量衡单位，取值

在0到1之间，数值的大小反映了Y的总变异中由X或线性模型所能解释的比例（4分）。决定系数 $R^2$ 等于Pearson线性相关系数r的平方（2分）。

23. 简述SPSS变量名的命名规则。

简答题 (10.0 分) (难度度:中)

答案解析：

答：SPSS的变量名可为英文亦可为中文（2分）。使用英文作为变量名时，首字母要求为英文字符（3分）。SPSS的变量名不区分大小写（1分）、长度在64个字符以内（1分）、变量之间不允许重名（1分）、不允许使用SPSS的保留用词如ALL、AND、B Y、EQ、GE、GT、LE、LT、NE、NOT、OR、TO、WITH等（2分）。

24. 计算题

综合题 (10.0 分)

(1) 某地10名健康儿童的头发与血液中的硒浓度见下表，试估计由发硒预测血硒的线性回归模型，并作假设检验。

10 名健康儿童的发硒与血硒含量		
编号	发硒(1： 1000)	血硒(1： 1000)
1	73	14
2	96	16
3	87	13
4	65	11
5	70	12
6	76	13
7	92	15
8	67	9
9	62	6
10	74	10

简答题 (10.0 分) （难度度:中）

答案解析：

答：（1） 由原始数据作散点图

通过散点图（略）初步判断儿童的血硒与发硒有直线趋势，因此可以进行线性回归分析。

（1分）

（2） 计算得：

$\sum X=762, \sum Y=119, \sum X^2=59288, \sum Y^2=1497, \sum XY=9327$ ，每一个个体的计算结果见下表。

10名健康儿童的发硒与血硒含量					
编号	发硒 X(cm)	血硒 Y(cm)	$X^2$	$Y^2$	XY
1	73	14	5329	196	1022
2	96	16	9216	256	1536
3	87	13	7569	169	1131
4	65	11	4225	121	715
5	70	12	4900	144	840
6	76	13	5776	169	988
7	92	15	8464	225	1380
8	67	9	4489	81	603
9	62	6	3844	36	372
10	74	10	5476	100	740
合计	762	119	59288	1497	9327

（1分）

（3） 计算  $x$  的均数  $\bar{X}$  和  $y$  的均数  $\bar{Y}$ ，以及离均差平方和  $l_{xx}$ 、 $l_{yy}$  与离均差积和  $l_{xy}$

$$\bar{X} = \frac{\sum X}{n} = 762 / 10 = 76.2$$

$$\bar{Y} = \frac{\sum Y}{n} = 119 / 10 = 11.9$$

$$l_{xx} = \sum (X - \bar{X})^2 = \sum X^2 - (\sum X)^2 / n = 59288 - 762^2 / 10 = 1223.6$$

$$l_{yy} = \sum (Y - \bar{Y})^2 = \sum Y^2 - (\sum Y)^2 / n = 1497 - 119^2 / 10 = 80.9$$

$$l_{xy} = \sum (X - \bar{X})(Y - \bar{Y}) = \sum XY - (\sum X)(\sum Y) / n = 9327 - 762 \times 119 / 10 = 259.2$$

(2分)

(4) 计算回归系数  $b$  和截距  $a$

$$b = \frac{l_{xy}}{l_{xx}} = \frac{259.2}{1223.6} = 0.212$$

$$a = \bar{Y} - b\bar{X} = 11.9 - 0.212 \times 76.2 = -4.254$$

(2分)

(5) 样本线性回归模型为:

$$\hat{Y} = -4.254 + 0.212X$$

(1分)

(6) 简单线性回归模型的假设检验可以用  $t$  检验, 也可以用方差分析。本例利用方差分析进行假设检验。  
(3分)

①建立检验假设:

$H_0: \beta_1=0$ , 即发硒与血硒之间不存在直线回归关系

$H_1: \beta_1 \neq 0$ , 即发硒与血硒之间存在直线回归关系

$\alpha=0.05$

②计算统计量  $F$  值

计算结果见下表:

回归分析的方差分析计算表

变异来源	SS	$\nu$	MS	F
回归	54.907	1	54.907	16.90
剩余	25.993	8	3.249	
总	80.900	9		

③查  $F$  界值表, 确定概率值  $P$  并做统计推断

本例自由度  $\nu_1=1$ ,  $\nu_2=n-2=10-2=8$ , 查  $F$  界值表得  $F_{0.05(1, 8)}=5.32$ 。样本  $F=16.90>5.32$ , 故  $P<0.05$ 。按照  $\alpha=0.05$  的检验水准, 拒绝  $H_0$ , 儿童发硒与儿童血硒之间的线性回归关系有统计学意义。

## 25. 计算题

综合题 (10.0 分)

(1) 测量10名健康成年男子基础代谢24小时产热量(千卡)及体表面积( $\text{cm}^2$ ), 数据见下表, 试回答:

(1) 下表资料有几个变量, 分别是什么类型的变量?

(2) 适宜进行Pearson相关分析还是Spearman相关分析?

(3) 假定成年男子体表面积与产热量服从双变量正态分布, 本例相关系数  $r=0.624$ , 据此认为两个变量之间有相关关系是否妥当?

10 名健康成年男子体表面积与产热量

编号	体表面积 $X(\text{cm}^2)$	产热量 $Y(\text{千卡})$
1	1.36	0.115
2	1.46	0.137
3	1.53	0.145
4	1.57	0.153
5	1.59	0.157
6	1.32	0.161
7	1.64	0.163
8	1.66	0.159
9	1.73	0.162
10	1.74	0.164

简答题 (10.0 分) (难易度:中)

答案解析：

答：（1）本例有2个变量，均为数值型变量。（4分）

（2）绘制产热量与体表面积散点图，如有共变趋势，而且产热量与体表面积服从双变量正态分布，可以进行*Pearson*相关分析；否则宜进行*Spearman*相关分析。（3分）

（3）不妥当，应该进行相关系数的假设检验，根据检验结果作出推断结论。（3分）