

军事医学科学院 2004 年攻读硕士学位研究生入学考试

卫生统计学试题

考生注意: 本试题的答案必须写在答题纸上, 写在试题上不给分。

一、选择题 (每题 3 分, 共 30 分)

- 某市收集了自 1949 年到 1987 年 5 种传染性疾病发病率的资料, 想绘一张统计图直观地比较它们的变化趋势, 为了不使读者产生误解, 请选用适当的统计图。
A. 圆图 B. 半对数线图 C. 直方图 D. 复式条图
- 已知 A 药对某病有效, 现发明一种增效剂 B, 可提高 A 药的疗效, 想通过临床试验了解 A+B 的疗效是否显著地优于单用 A 药的疗效, 应选用:
A. t 检验 B. χ^2 检验 C. 双侧检验 D. 单侧检验
- 在进行成组设计资料的 t 检验中, H_0 为 $\mu_1 = \mu_2$, H_1 为 $\mu_1 \neq \mu_2$, $\alpha = 0.05$ 。若检验的结果为接受 H_0 , 拒绝 H_1 , 问此时可能会犯什么类型的错误?
A. 严重错误 B. 第 II 类错误 C. 第 I 类错误 D. 5% 的错误
- 某市为了了解正常成人血铅含量的水平, 按职业进行分层随机抽样, 调查了 1000 例正常成人, 经检验发现血铅含量 x ($\mu\text{g}/100\text{g}$) 服从对数正态分布。已知血铅含量过高为异常。请选用适当公式计算血铅含量的正常值范围 (令 $y = \lg x$, $\alpha = 0.01$, 设 S_x 、 S_y 分别代表 x 和 y 的标准差, $S_{\bar{x}}$ 和 $S_{\bar{y}}$ 分别代表 \bar{x} 和 \bar{y} 的标准误)。
A. $\lg^{-1}(\bar{y} \pm 2.576 S_{\bar{y}})$ B. $\lg^{-1}(\bar{x} \pm 1.960 S_{\bar{x}})$
C. $\lg^{-1}(\bar{y} + 2.326 S_{\bar{y}})$ D. $\lg^{-1}(\bar{x} - 2.326 S_{\bar{x}})$
- 掌握方差分析的基本思想, 即使记不住具体公式, 一般也能做出结果, 得出结论, 其基本思想可简述为:
A. 组间方差大于组内方差 B. 误差的方差必然小于组间方差
C. 总离均差平方和可以被合理地分解 D. 两方差之比服从 χ^2 分布
- 为了直观地比较化疗后相同时间点上一组乳癌患者血清肌酐和血液尿素氮两项指标观测值的变异程度的大小, 可选用的变异指标是下列选项中的哪一个?
A. 标准差 B. 标准误 C. 相关系数 D. 变异系数
- 某医生测得男、女各 50 人淋巴细胞计数 L (%) (即占白细胞总数的百分比), 想用成组设计资料的 t 检验分析男、女两性之间的差别有无统计学意义, 在作前提条件的检验前最好对数据作什么样的变量变换为宜?
A. 平方根变换 B. 对数变换 C. 平方根反正弦变换 D. 倒数变换

8. 试验设计中强调必须遵守“随机、对照、重复、均衡”四大原则,其根本目的是为了什么?
- A.减少过失误差、降低随机误差、消除系统误差;
B.纯化“信号”、降低“噪音”、多快好省;
C.便于收集资料、便于统计处理、便于撰写论文;
D.仅用一、两次,最多十几次试验,就可得到可靠结果。
9. 在某项试验研究中,考察的指标有一百多项,样本含量有 1000 多个。尚无办法将它们明确地划分开。指标彼此之间并无因变量和自变量之分,但从专业上来看,希望通过统计分析方法,对指标进行分组,以便发现哪些指标关系比较密切,为专业研究提供线索。为此,可选用下列哪种统计分析方法有利于实现前述的分析目的?
- A.聚类分析 B.多元回归分析 C.典型相关分析 D.判别分析
10. 已知 A、B、C、D、E 都是三水平因素,且根据预试验结果得知: $A \times B$ 、 $B \times C$ 不可忽视。由于经费紧张,研究者希望试验次数尽可能少一些,还希望数据分析比较方便,结论比较可靠。问最好选用下列哪种试验设计类型?
- A.析因设计 B.交叉设计 C.正交设计 D.分割设计

二、是非题(正确填“√”;错误填“×”。每题 2 分,共 14 分)

1. 何为标准误?标准误就是平均数的标准差。 ()
2. 2003 年年初, SARS 病毒肆虐北京,在某段时间内,感染上 SARS 病毒的 100 人中约有 30 人是医护人员,故可以说这段时间医护人员的 SARS 病毒感染率约为 30%。()
3. 编制得最差的统计表应当是未将表中数据的含义表达清楚的统计表。 ()
4. 当某定量的观测指标的取值服从偏态分布时,若希望用简略形式表达资料,最好采用“ $\bar{x} \pm s_x$ ”的形式,不宜采用“ $\bar{x} \pm s$ ”的形式。 ()
5. 某年某单位一千人中有 30 人患了肝炎,可以认为该年该单位肝炎发病率为 3%,这个 3%就是一个定性的数据;为了反映这一千人细胞的免疫机能,需要测定他们的“E-玫瑰结形成率(%)”,可从每个人身上测出一个具体的数值来,由于它们的总称都是“E-玫瑰结形成率(%)”,故这些数据也都应叫做定性数据。 ()
6. 同时测了 10 人身高(cm)和体重(kg)的数据,希望计算变异指标的数值,并通过其数值的大小初步说明哪个指标的离散度稍大一些,宜选用变异系数。 ()
7. 当两个定量变量之间的直线相关系数 r 的平方(即决定系数)大于 0.5 时,用其中一个定量变量的值去预测另一个定量变量的数值就有实用价值了。 ()

三、统计学应用错误辨析题(每题 6 分,共 30 分)

1. 某研究者将 24 只大鼠随机分成 3 组,每组 8 只。一组:正常对照组(用生理盐水灌胃),二组:激素组(氢化可的松灌胃),三组:补骨 1 号合用激素组(氢化可的松灌胃的同时加用补骨 1 号)。实验一段时间后,测定骨小梁面积等定量指标,经分析,认为补骨 1 号有防治类固醇性骨质疏松的作用。试问:从实验设计角度看,该研究者犯

了什么错误? 正确的做法是什么?

2. 为了观察甲紫注入小型猪正常腮腺后组织病理变化情况, 有人选择6个月龄、体重20~25kg的中国实验用小型猪15只, 雄性9只、雌性6只。每只动物任选一侧腮腺为实验侧, 另一侧作为正常对照, 以消除个体差异及增龄对实验结果的影响。按注入甲紫后1周、2周、1个月、3个月及6个月将15只动物随机分为5组, 每组3只(每个组的3只动物分别随机注入0.6ml、1.0ml及4.0ml 1%甲紫溶液), 然后观察组织病理变化情况。试问: 此项实验研究中违背了实验设计的什么原则? 其后果是什么?
3. 原文题目: 《600例小儿烧伤休克期治疗分析》, 原作者在文中指出: “600例烧伤患儿, 210例早期有休克症状, 其中3岁以下者110例, 占52%, 3岁以上者100例, 占48%, 年龄越小, 休克发生率越高”。试问: 从统计描述角度看, 原作者犯了什么错误?
4. 原文题目: 《准分子激光原位角膜磨镶术治疗近视散光的初步研究》, 原作者将B、C两组患者手术前后散光轴夹角度数的变化绘图如下, 见图1。试问: 原作者绘制的这幅图犯了什么错误?

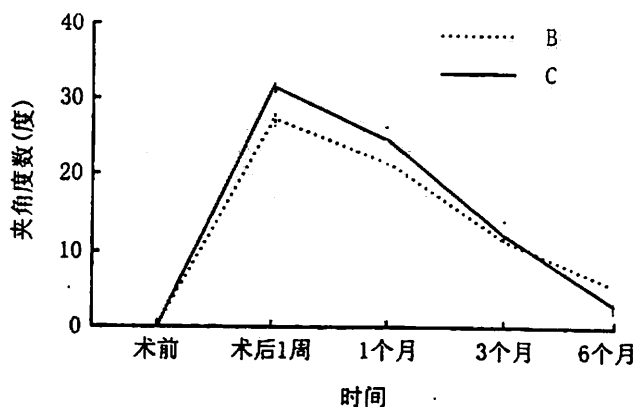


图1 B、C两组患者手术前、后散光轴夹角度数的变化(原图)

5. 原文题目: 《格列本脲对糖尿病及正常大鼠心肌磺胺类药物受体mRNA的影响》, 为研究长期格列本脲治疗对正常与糖尿病大鼠心肌SUR1、SUR2和kir6.2的影响, 将大鼠随机分为4组, 结果见表1, 原作者采用t检验进行了统计分析。

表1 各组心肌 SUR1、SUR2 和 kir6.2 的 mRNA 杂交信号光密度积分的比较 ($\bar{x} \pm s$)

组别	鼠数	SUR1	SUR2	Kir6.2
糖尿病组	8	35 ± 10	152 ± 35	35 ± 6
糖尿病格列本脲治疗组	14	34 ± 14	153 ± 51	36 ± 7
非糖尿病组	9	31 ± 8	119 ± 41	37 ± 8
非糖尿病格列本脲治疗组	13	43 ± 17*	162 ± 51**	41 ± 9

注: 与非糖尿病组比较, $*t=2.2$, $P=0.039$; $**t=2.1$, $P=0.045$

试问: 原作者所采用的统计分析方法是否正确, 为什么? 假定资料满足参数检验的前提条件, 希望同时考察三项定量指标的平均值在不同条件下的差别是否具有统计学意义时, 正确的统计分析方法是什么?

四、实验设计题 (每题 10 分, 共 20 分)

1. 某研究者拟用 4 只家兔做试验, 希望了解注射 4 种不同剂量 (0.32, 0.47, 0.62, 0.77U/kg) 的胰岛素后, 家兔血糖下降值的改变有无统计学意义。由专业知识得知, 若前后两次试验的间隔超过一天, 即使用同一只兔做试验, 结果也不会受到影响。根据已知条件, 选择一个最合适的实验设计类型进行安排, 并将其以表格的形式表达出来。
2. 现有两个研究小组, 需要测定受试者的血压值。现有两台血压计, 要分别放在两个研究小组用于测定血压值, 为了检查两台血压计测定的结果是否一致, 需要先做试验进行考察。希望每台血压计都要检测每个人各一次, 现有 10 个受试者, 请选用一个最合适的实验设计类型进行安排, 并将其以表格的形式表达出来。

五、统计分析方法合理选用题 (每题 6 分, 共 24 分)

资料与问题: 患者年龄不同, 检查出品状体混浊程度不同, 现有资料如表 2 所示, 试根据分析的目的选用合适的统计分析方法。

表 2 矽肺期次与肺门密度级别之关系的观测结果

矽肺 期次	患者例数			
	肺门密度级别:	+	++	+++
I期		43	188	14
II期		1	96	72
III期		6	17	55

1. 若将“肺门密度级别”视为观测结果, 并用秩和检验进行了数据处理, 得“ $P<0.001$ ”的统计计算结果, 问应当如何陈述专业结论?
2. 若研究的目的是考察“矽肺期次”与“肺门密度级别”之间的相关关系是否具有统计学意义, 问应选择什么统计分析方法处理此资料合适? 为什么?
3. 若研究的目的是考察“矽肺期次”与“肺门密度级别”之间是呈直线还是呈曲线变化关系, 问应选择什么统计分析方法处理此资料合适? 为什么?
4. 若用一般 χ^2 检验处理此资料, 其 H_0 是什么? 若假设检验的结果为拒绝 H_0 , 如何下专业结论?

六、统计计算与假设检验题（每题 8 分，共 32 分）

1. 某人测得 9 名正常成人非蛋白氮 (mg/100ml) 为：
20.1, 22.3, 23.4, 24.8, 25.7, 26.9, 28.2, 31.4, 34.3
假定资料近似服从正态分布，试根据这组资料求正常人非蛋白氮 (mg/100ml) 的 95% 置信区间。
2. 对未成年大白鼠经 3 种不同剂量雌激素注射一定时间后的子宫重量，以及取 4 窝不同种系的大白鼠，每窝 3 只，随机分配到三种剂量组内进行实验，资料见表 3。假定资料已满足正态性和方差齐性，问剂量和窝别的各自水平下子宫重量平均值之间的差别有无统计学意义？

表3 未成年大白鼠注射不同剂量雌激素后的子宫重量

窝别	子宫重量(g)			
	剂量 (μg/100g):0.2	0.4	0.8	合计
1	106	116	145	367
2	42	68	115	225
3	70	111	133	314
4	42	63	87	192
合计	260	358	480	1098

注： $\sum_{i=1}^4 \sum_{j=1}^3 x = 1098$, $\sum_{i=1}^4 \sum_{j=1}^3 x_{ij}^2 = 113542$, 校正数 $C=100467$

3. 某病患者 13 人，检测其两项定量指标 MNC ($\times 10^8/\text{kg}$) 和 CD34⁺ ($\times 10^6/\text{kg}$)，由专业知识可知，这两项指标在专业上有一定的联系；若绘出反映这两个变量同时变化的散布图也可看出两变量之间呈直线变化趋势，希望通过测定 CD34⁺ 的值去推测 MNC 的值。数据如下，试根据已知条件拟合直线回归方程，并检验该直线回归方程在统计学上是否成立。

患者编号	1	2	3	4	5	6	7
y: MNC 的值	4.42 [*]	2.67	3.23	2.54	2.37	2.38	2.58
x: CD34 ⁺ 的值	7.07	1.39	1.58	1.09	1.42	0.48	1.55
患者编号	8	9	10	11	12	13	
y: MNC 的值	1.24	2.43	2.16	3.49	3.06	2.65	
x: CD34 ⁺ 的值	0.31	3.43	1.19	4.36	5.51	2.41	

已知：n=13, $\sum x=31.79$, $\bar{x}=2.445385$, $\sum x^2=128.7057$
 $\sum y=35.22$, $\bar{y}=2.709231$, $\sum y^2=102.1618$, $\sum xy=101.0927$

$$l_{xx} = \sum (x - \bar{x})^2 = \sum x^2 - \frac{1}{n}(\sum x)^2 = 50.96692308$$

$$l_{yy} = \sum (y - \bar{y})^2 = \sum y^2 - \frac{1}{n}(\sum y)^2 = 6.742692308$$

$$l_{xy} = \sum (x - \bar{x})(y - \bar{y}) = \sum xy - \frac{1}{n}(\sum x)(\sum y) = 14.96625385$$

4. 两组医生对一批 X 线片判读的结果如表 4 所示，试评价两组医生评判的结果是否一致。

表 5 年资不同的两组医生对一批 x 线片判读的结果

低年资		X 线片张数					
医生组	高年资医生组:	I	II	III	IV	V	合计
I		23	11	3	2	0	39
II		10	32	3	3	0	48
III		1	29	10	7	3	50
IV		4	7	5	22	19	57
V		2	2	3	16	86	109
合计		40	81	24	50	108	303

注意：分析此资料时有一个统计量的标准误比较复杂，设该统计量为 K，其标准误为 S_K，现已知 S_K=0.0346。请在此基础上完成对该资料统计分析的全过程。

附录: 计算题中可能用到的临界值简表

自由度	t 的临界值	
	$\alpha=0.05$	$\alpha=0.01$
6	2.447	3.707
7	2.365	3.499
8	2.306	3.355
9	2.262	3.250
10	2.228	3.169
11	2.201	3.106
12	2.179	3.055
13	2.160	3.012

		方差分析用的 F 临界值	
分子自由度	分母自由度	$\alpha=0.05$	$\alpha=0.01$
2	5	5.79	13.27
2	6	5.14	10.92
2	7	4.74	9.55
2	8	4.46	8.65
2	9	4.26	8.02
3	5	5.41	12.06
3	6	4.76	9.78
3	7	4.35	8.45
3	8	4.07	7.59
3	9	3.86	6.99

自由度	相关系数 r 的临界值	
	$\alpha=0.05$	$\alpha=0.01$
10	0.576	0.708
11	0.553	0.684
12	0.532	0.661
13	0.514	0.641
14	0.497	0.623
15	0.482	0.606