

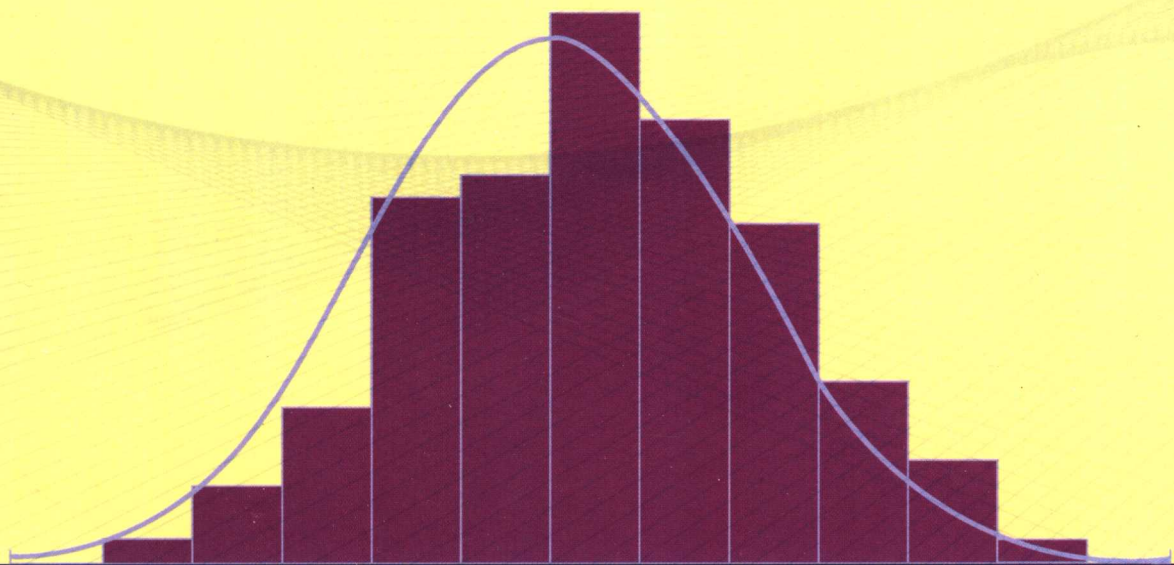


普通高等教育“十五”国家级规划教材配套参考书

医学统计学 习题全解指南

主 编 徐勇勇

副主编 孙振球 颜 虹 陈长生



高等教育出版社
HIGHER EDUCATION PRESS

普通高等教育“十五”国家级规划教材配套参考书

医学统计学习题全解指南

主 编 徐勇勇

副主编 孙振球 颜 虹 陈长生

编 委 (按章节顺序排列)

徐勇勇(第四军医大学)

孙振球(中南大学湘雅医学院)

颜 虹(西安交通大学医学院)

毕育学(西安交通大学医学院)

丁建生(兰州医学院)

尹 平(华中科技大学同济医学院)

颜 艳(中南大学湘雅医学院)

姜 潮(大连医学院)

刘启贵(大连医学院)

沈其君(宁波大学医学院)

陈平雁(南方医科大学)

贺 佳(第二军医大学)

梁维君(湖南师范大学医学院)

钟晓妮(重庆医科大学)

陈冠民(武汉大学公共卫生学院)

夏结来(第四军医大学)

陈长生(第四军医大学)

王 彤(山西医科大学)

高等教育出版社

内容提要

本书是普通高等教育“十五”国家级规划教材《医学统计学(第二版)》的配套教材。本书对《医学统计学(第二版)》中的习题作了详细的解答。题型有最佳选择题、问答题和计算题。对于最佳选择题,本书直接给出了解答;对于问答题,本书给出的解答侧重于对统计学的基本概念、理论和方法的正确解释;对于计算题,本书在解答时注重的是培养和提高学生联系实际、解决和分析问题的实际应用能力。学生在使用本配套教材时,要先仔细领会各题题意,认真作答,然后再核对其解答。

本书可作为高等医药类院校药学专业本科教材,也可供从事医药卫生工作的科技人员学习参考。

图书在版编目(CIP)数据

医学统计学习题全解指南/徐勇勇主编. —北京:高等教育出版社,2006.11
ISBN 7-04-020197-6

I. 医… II. 徐… III. 医学统计-医学院校-解
题 IV. R195.1-44

中国版本图书馆 CIP 数据核字(2006)第 121076 号

策划编辑	李蕊	责任编辑	董达英	封面设计	张楠	责任绘图	郝林
版式设计	马静如	责任校对	王效珍	责任印制	宋克学		

出版发行	高等教育出版社
社 址	北京市西城区德外大街 4 号
邮政编码	100011
总 机	010-58581000
经 销	蓝色畅想图书发行有限公司
印 刷	北京凌奇印刷有限责任公司
开 本	850×1168 1/16
印 张	6.5
字 数	160 000

购书热线	010-58581118
免费咨询	800-810-0598
网 址	http://www.hep.edu.cn http://www.hep.com.cn
网上订购	http://www.landraco.com http://www.landraco.com.cn
畅想教育	http://www.widedu.com
版 次	2006 年 11 月第 1 版
印 次	2006 年 11 月第 1 次印刷
定 价	8.80 元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换。

版权所有 侵权必究

物料号 20197-00

前 言

我们编写的这本《医学统计学习题全解指南》，是普通高等教育“十五”国家级规划教材《医学统计学》的配套教材。

医学统计学的教学必须通过教师课堂教学和学生自己动手练习两个环节完成，练习是使学生加深理解统计学的基本理论和方法、培养和提高联系实际、解决和分析问题能力的重要手段。本配套教材对《医学统计学》教材中的所有练习题都作了解答。教材第一章(绪论)中的题型只有问答题；第二章(研究设计)至第三章(观察与抽样)中的题型有最佳选择题和问答题；其余各章中的题型有最佳选择题、问答题和计算题。对于教材中的最佳选择题，本书直接给出了解答；各问答题的解答仅供参考；计算题的统计计算可用具有统计功能的计算器来完成。学生在使用本配套教材时，要先仔细领会各题题意，认真解答每题，然后再根据答案核对自己的解答，切忌直接看答案。

《医学统计学》教材的各章编委，为教材中各自编写部分的练习题提供了答案，第四军医大学卫生统计学教研室的杨鹏、吴昊、王锐三位同志核算了本配套教材练习题的计算结果，赵瑞松同志承担了全部稿件的打印工作。

由于编者水平所限，本配套教材不足之处在所难免，诚恳希望广大师生提出进一步修改的意见。

徐勇勇

2005年10月于西安

目 录

第一章 绪论	1	练习题	46
练习题	1	练习题参考答案	48
练习题参考答案	1	第十二章 线性回归分析	53
第二章 研究设计	2	练习题	53
练习题	2	练习题参考答案	54
练习题参考答案	3	第十三章 线性相关分析	60
第三章 观察与抽样	5	练习题	60
练习题	5	练习题参考答案	61
练习题参考答案	6	第十四章 基于秩次的统计方法	66
第四章 数据特征与统计描述	10	练习题	66
练习题	10	练习题参考答案	67
练习题参考答案	12	第十五章 临床测量的误差评价与诊断 试验	70
第五章 概率、概率分布与临床决策	18	练习题	70
练习题	18	练习题参考答案	72
练习题参考答案	19	第十六章 生命统计的常用指标	75
第六章 正态分布与医学参考值范围	22	练习题	75
练习题	22	练习题参考答案	76
练习题参考答案	23	第十七章 生存分析	79
第七章 参数估计	27	练习题	79
练习题	27	练习题参考答案	80
练习题参考答案	28	第十八章 Meta 分析	84
第八章 假设检验的基本概念	31	练习题	84
练习题	31	练习题参考答案	85
练习题参考答案	33	第十九章 统计方法的综合运用与统计 结果的表达	89
第九章 χ^2 检验	36	练习题	89
练习题	36	练习题参考答案	91
练习题参考答案	38	第二十章 样本量估计	95
第十章 t 检验	41	练习题	95
练习题	41	练习题参考答案	96
练习题参考答案	42		
第十一章 多个样本均数比较的方差 分析	46		

第一章 绪 论

练 习 题

1. 如果有人认为手机辐射可能导致脑癌,你会要求医学研究者提供什么“证据”?
2. 为什么生物医学数据的收集、整理和分析必须使用正确的统计方法?
3. 某年级甲班、乙班各有男生 50 人.从两个班各抽取 10 人测量身高,如果甲班的平均身高大于乙班,能否推断甲班所有同学的平均身高也大于乙班?为什么?
4. 根据英国发表了评价链霉素治疗肺结核疗效的随机对照的临床试验报告,试验组的生存率为 93%,对照组的生存率为 73%,这只是 55 例患者的试验结果,根据什么说“链霉素对所有 15~30 岁双肺活动性肺结核患者都有效”?

练习题参考答案

1. 答 可以用观察对比的方法,根据有无脑癌将研究对象分为病例组(脑癌组)和对照组(非脑癌组),通过调查他们以往手机的使用情况(是否用手机,用手机种类、时间),比较两组的暴露水平是否有差异,从而得出手机电波与脑癌有无关联的初步判断.

2. 答 因为在生物医学试验中,试验数据是随机变化的,即个体差异(变异)是生物医学研究领域里普遍存在的现象.在自然状态下,任何两个患者或两个研究群体间都存在差异,其表现为各种生理测量值的参差不齐,如身高、体质量、血压、胆固醇、血糖、免疫球蛋白、心功能、肺功能、肝功能等.即使同一个个体,其生理测量值也不是恒定的常数.

变异的存在,导致对实验结果判断的不确定性.因此,存在变异时,即使是同样条件下重复实验的结果也不会完全相同,医生和医学研究人员必须学会用概率的观点来看待生命现象中的因果联系,用正确的统计方法来进行生物医学数据的收集、整理和分析.

3. 答 不能.因为从甲、乙两班分别抽取的 10 人,测量其身高,得到的分别是甲、乙两班的一个样本.样本的平均身高只是甲、乙两班所有同学平均身高的一个点估计值.即使是按随机化原则进行抽样,由于存在抽样误差,样本均数与总体均数一般很难恰好相等.因此,如果仅凭两个样本均数孰高孰低判断全班身高的差别,不能估计出判断错误的概率.

4. 答 因为对于这个试验报告,采用了统计学方法进行了临床干预试验.在试验设计时采用了随机化分组的方法,尽可能地保证了两组患者的病情相同.在此基础上推论出试验结果代表的两组总体生存率有差别,并且满足 $P < 0.05$ (推论错误的概率小于 5%),才得出链霉素对目标总体的所有患者都有治疗效果的结论.所以,这里因为有统计理论的指导,虽然仅仅是 55 例患者的试验结果,但可得出对所有 15~30 岁双肺活动性肺结核患者都有效的推论.

第二章 研究设计

练习题

一、最佳选择题

1. 实验设计的基本原则是().
A. 随机化、盲法、设置对照
B. 重复、对照、随机化
C. 随机化、盲法、配对
D. 处理因素、受试对象、实验效应
E. 重复、均衡、随机化
2. 某医生欲研究加锌牛奶对儿童生长发育的影响,将某小学校 10 岁儿童随机分为两组,一组喝加锌牛奶,另一组喝普通牛奶,此属于().
A. 假干预对照
B. 空白对照
C. 安慰剂对照
D. 标准对照
E. 无对照
3. 非随机对照试验包括().
A. 历史对照、平行对照、准试验
B. 历史对照、前后对照、配对设计
C. 空白对照、前后对照、准试验
D. 历史对照、前后对照、准试验
E. 标准对照、实验对照、空白对照
4. 实验设计的三个基本要素是().
A. 对照、重复、随机化
B. 重复、随机化、设置对照
C. 处理因素、试验单位、处理效应
D. 均衡、重复、随机化
E. 处理因素、测量指标、设置对照
5. 计量资料、计数资料和等级资料的关系是().
A. 计量资料兼有计数资料和等级资料的一些性质
B. 计数资料兼有计量资料和等级资料的一些性质
C. 等级资料兼有计量资料和计数资料的一些性质
D. 三种资料不可以相互转化
E. 等级资料较计量资料和计数资料精确

二、问答题

1. 简述科学研究与日常临床工作的区别.
2. 简述研究设计的基本要素及主要的统计学问题.
3. 简述对照组常用的干预方式.
4. 简述随机化的意义及主要作用.
5. 简述非随机对照试验偏倚的种类.

6. 简述医学测量结果的分类(数据类型).
7. 简述“软”指标的定义及测量方法.
8. 试将 200 名患者用随机排列分段随机化的方法分为 A、B 两组.

练习题参考答案

一、最佳选择题解答

1. B 2. A 3. D 4. C 5. C

二、问答题解答

1. 答 在临床上,要区分科学研究与日常临床工作的关系. 科学研究预先有明确的研究目的,需要标准的试验条件和统一的数据采集方法,而日常临床工作只需要满足临床需要即可. 再者,在日常临床工作中,患者的病情记录、各种检查单、影像资料也不会像科学研究那样完整. 因此,临床疗效评价的结论性研究不能依赖日常临床工作采集的数据,应该按照正规的研究设计计划统一采集,而且数据的统计分析方法,也要事先在研究设计计划中明确,而不是到了资料分析阶段才临时选择.

2. 答 研究设计的三个要素分别为:

(1) 处理因素:取决于研究假设和主要研究问题,主要的统计学问题是如何设立对照组,选择合适的分组方法和实验设计模型,尽可能使非处理因素在各处理组中的分布均衡.

(2) 试验单位:是接受处理的基本单位,主要的统计学问题是如何选定一批具有代表性的试验单位,并且有足够多的数量.

(3) 处理效应:也称试验效应,指试验单位接受处理因素后的反应,用于反映处理效应的大小. 统计学主要考虑的问题是对测量结果的分类和进一步的统计处理,如判断测量结果的统计分布、计算合适的统计指标、推断统计指标的抽样误差.

3. 答 在实验性研究中,设立对照组也是一种处理. 对照组常用的干预方式有:

(1) 空白对照:不给任何干预的对照,通常用于无损伤、无疼痛的实验室研究、动物实验和临床试验.

(2) 实验对照:又称假干预对照,是指与试验组操作相同,但与处理效应无关的对照,通常用于有损伤、有刺激的动物实验.

(3) 安慰剂对照:也是一种假干预对照,主要用于临床试验.

(4) 标准对照:对照组的干预采用公认的“标准”方式,主要用于临床试验.

4. 答 随机化是指在实验分组时,每个观察对象具有相同的机会(概率)被分配到试验组(对照组). 随机化是一种实验设计的分组程序,不能理解为随心所欲的“随便”、“随意”和“非选择性”. 其主要作用有:

(1) 保证试验组和对照组各种已知或未知的特征(如患者的年龄、性别、病情、用药史等)都达到均衡. 或者说,在随机分组后、施加干预以前,所有观察对象都属于同一总体. 施加干预以后如果发现试验组和对照组有差异,则认为干预有效.

(2) 避免因研究者主观愿望破坏试验组和对照组的均衡性,如有意(或无意)将轻病人分配到试验组,将重病人分配到对照组. 因此,在临床新药试验中,随机化分组过程和分组结果必须作为档案保存,以便药物监督部门核查.

(3) 随机化分组是所有统计假设检验(如 t 检验、 χ^2 检验)推论因果关系的基础。

5. 答 非随机对照试验主要偏倚种类有:

- (1) 无对照研究的偏倚;
- (2) 历史对照的偏倚;
- (3) 前后对照的偏倚;
- (4) 非随机同期对照和准试验;
- (5) 依从性偏倚与失访偏倚;
- (6) 意向性分析偏倚。

6. 答 医学测量结果按其数据类型分为计量资料、计数资料和等级资料三类。在研究设计时,应尽量选择准确性高、客观性好的计量资料作为测量指标。

7. 答 在众多的医学测量指标中,可以精确测量的疗效评价指标,如长度、重量、体积、浓度、压力、时间等,称之为“硬”指标。有些疗效评价指标则不能精确测量,如患者的疼痛程度、生活能力改善情况等,称之为“软”指标。“软”指标的测量工具称“软”工具,如调查表、调查问卷。

8. 答 将 200 名患者用随机排列分段随机化的方法分为 A,B 两组。

令 $T=5$,区段长度为 $2 \times 5=10$,200 个受试者划分为 20 个区段。规定每个区段内 R 为 1~5 者为 A 组,6~10 者为 B 组,第 1 个区段的分组结果如下:

受试者编号	1	2	3	4	5	6	7	8	9	10
随机数	33	35	72	67	47	77	34	55	45	70
R	1	3	9	7	5	10	2	6	4	8
分配组别	A	A	B	B	A	B	A	B	A	B

重复以上步骤,即可得到另外 19 个区段的分组结果。

第三章 观察与抽样

练习题

一、最佳选择题

1. 观察性研究与实验性研究的区别为().
A. 是否以人为研究对象 B. 是否是抽样研究 C. 是否设立了对照
D. 是否对研究对象施加了干预 E. 是否进行了测量
2. 下列研究方法中,属于描述性研究的方法是().
A. 实验性研究 B. 病例对照研究 C. 前瞻性队列研究
D. 回顾性队列研究 E. 横断面研究
3. 下列抽样方法中属于非概率抽样的方法是().
A. 单纯随机抽样 B. 系统抽样 C. 分层抽样
D. 配额抽样 E. 整群抽样
4. 一般而言在下列抽样方法中抽样误差最小的是().
A. 整群抽样 B. 系统抽样 C. 配额抽样
D. 分层抽样 E. 单纯随机抽样
5. 队列研究和病例对照研究的主要区别为().
A. 是否以人为研究对象 B. 是否是抽样研究 C. 是否设立了对照
D. 是否对研究对象施加了干预 E. 是“由果推因”,还是“由因寻果”
6. 无应答偏倚属于().
A. 信息偏倚 B. 混杂偏倚 C. 选择偏倚
D. 易感性偏倚 E. 检出偏倚

二、问答题

1. 简述观察性研究的主要特征.
2. 什么是普查? 什么是抽样调查? 它们各有什么特点?
3. 什么是总体? 什么是样本? 它们有什么关系?
4. 什么是参数? 什么是统计量? 它们有什么关系?
5. 什么是随机抽样? 它有哪些抽样方法?
6. 什么是横断面研究? 它的主要目的和用途是什么?
7. 什么是队列研究? 它有哪些类型和优缺点?
8. 什么是病例对照研究? 它有哪些类型和优缺点?
9. 简述疾病的病因学研究.
10. 简述病因推断的基本过程.

11. 如何推断病因学研究的关联与因果?
12. 简述病因的判断标准.

练习题参考答案

一、最佳选择题解答

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

二、问答题解答

1. 答 与实验性研究相比,观察性研究的特点是对影响观察结果的因素通常不进行人为干预,往往采用配对的方式或在进行统计分析时使用统计模型来实现均衡.其研究成本比较小,伦理学问题也不突出,一般可通过分析判断逻辑推理来获得关联性的结论.

2. 答 普查是对由研究目的确定的所有观察单位逐一进行调查的过程,理论上讲只有普查才能取得总体参数,且没有抽样误差.一般用于了解总体某一特定“时点”的情况.当观察单位较多时,要花费大量人力、物力和财力.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.一般而言,抽样调查比普查涉及的观察单位数少,因而节省人力、财力和时间,在进行了严密的设计之后同样可获得较为深入细致和准确的资料.

3. 答 总体是指根据研究目的确定的、同质的全部研究单位的测量值.当总体中所有研究对象事先可确定且有限时,称为有限总体,反之,称为无限总体.来自总体的部分研究单位(或其测量值)称为样本.有限总体在获得抽样框架后,可以实现随机抽样,即总体中的每个观察单位都有同样的机会被选作样本.这样获得的样本称为随机样本.它们的关系是样本来自总体,通过对样本数据的统计分析,可对总体的特征做出推断.

4. 答 参数是反映总体统计学特征的数字.统计量是用样本观测值计算的、反映样本分布统计学特征且不依赖于未知参数的数字.参数来源于总体,重要的参数除总体均数外,还有总体方差、总体标准差、总体相关系数等.由于大多数医学研究得不到总体数据,所以参数通常是未知的.统计量来源于样本,是根据样本计算出来的数字.随机样本的统计量可作为总体参数的估计值.一类用于参数估计,如样本方差;另一类统计量称为检验统计量,用于描述样本的分布特征,如根据样本观测值计算的 u 值、 t 值、 χ^2 值、 F 值等.

5. 答 随机抽样指总体中的每个观察单位都有同样的机会被选入样本.这样获得的样本称为随机样本.常用的方法有 4 种:

(1) 单纯随机抽样:单纯随机抽样是在总体中以完全随机的方法抽取一部分个体组成样本的抽样方法,即总体中每一个观察单位都有同等机会被抽取进入样本.其优点是均数(或率)及标准误的计算较简便.但对有限总体和无限总体标准误的计算公式有所不同.具体公式可参考有关书籍.其缺点是当总体例数较多时,要对观察单位逐一编号较麻烦.

(2) 系统抽样:系统抽样又称等距抽样或机械抽样,是指随机地在所要抽样的名单中每间隔若干个个体抽取一个个体的抽样方法.这种方法的优点是简便易行,容易得到一个按比例分配的样本.如上例,样本中男女的比例约等于该校全部学生的性别比例.这种方法的优点是简便易行,容易得到一个按比例分配的样本.在一般情况下,系统抽样的抽样误差与单纯随机抽样相仿,但并没有专门的计算系统抽样的抽样误差公式,一般都借用单纯随机抽样时抽样误差的计算公式.

系统抽样的缺点是:当总体的已有的顺序号有周期趋势或规律性时,系统抽样将产生明显的

偏性。

(3) 分层抽样:按照与研究目的明显有关的因素(或某种特征),将总体分为若干类型或区域,统计上叫“层”,然后从每一层内按比例抽取一定数量的观察单位,将各层的观察单位合起来组成样本。这种抽样方法称为分层抽样。

(4) 整群抽样:首先将总体按照某种与研究目的无关的分布特征(如地区范围;不同的团体;病历、格子等)划分为若干个“群”组,每个群包括若干观察单位;然后根据需要随机抽取其中部分“群”,并调查被抽各“群”中的全部观察单位。

整群抽样的优点是便于操作,容易控制调查质量。单纯随机抽样需要对调查总体的全部观察单位进行统一编号。在实际工作中,尤其当总体范围较大时,做到这一点往往比较困难。而总体中的地区、集团等分布特征,则是范围清楚的可利用的调查“群”。故整群抽样颇为常用。其缺点是由于所得样本在总体中的分布并不均匀,所以,当样本例数一定时,其抽样误差一般大于单纯随机抽样的误差。抽取的“群”数越少,“群”间差异越大,抽样误差也越大。

6. 答 横断面研究是指按照事先设计的要求,在某一人群中以人为基本研究单位,采用普查或抽样调查的方法收集特定时间内有关疾病及其影响因素的描述性资料,用以描述所研究疾病的分布及影响因素与疾病之间的关联,是描述性研究中应用最为广泛的方法。从时间上看,这一研究的过程在特定时间内进行,即在某一时刻或在一段短暂时间内完成,为时间的一个横断面,故称为横断面研究。

横断面研究的主要目的和用途:

(1) 横断面研究多用于研究某地区在时点上某种疾病的流行强度和该病在该地区的分布特点,以便分析患病率与哪些环境因素、社会因素、人群特征以及防治措施等因素有关。这些资料可为公共卫生管理人员评估某地区人群健康状况和制定有效的健康促进措施提供参考依据。

(2) 横断面研究的结果可为进一步进行疾病病因分析性研究提供线索;横断面研究适宜对不会发生改变的控制因素如血型、肤色、种族、性别等的研究;也适宜对能发挥长期、慢性累积影响的因素的研究,如高血压与冠心病的关系,糖尿病与动脉粥样硬化的关系等。对这样一些因素的横断面研究可提供真实的影响因素与疾病联系的证据。

(3) 横断面研究常作为早期发现病人的手段。利用普查、筛检等手段,可以早期发现病人,开展早期治疗,有效地减轻疾病的危害。

(4) 横断面研究可用于评价疾病的防治效果。如果定期地在某一地区人群中进行横断面研究,收集有关影响因素与疾病的资料,通过将多次横断面研究的结果进行对比分析,可对疾病的防治效果作出评价。

7. 答 队列研究是一种“由因寻果”的纵向前瞻性观察研究。根据观察开始时有无暴露史,研究者将没有出现研究疾病或其他结局的研究人群分为暴露人群和非暴露人群,并随访观察一定时期,旨在比较两组人群的疾病“发病”率有无差异,从而得出暴露与结局有无关联的推断。

队列研究分为前瞻性队列研究和回顾性队列研究两种类型。前瞻性队列研究的特点是观察开始时无论暴露是否已经出现,但结果一定没有发生。优点为研究者能按自己的意愿定义研究因素、直接获取资料甚至包括混杂因素的信息。回顾性队列研究的特点是观察开始时暴露与结局均已出现。优点为节省人力、物力与时间。但与前瞻性队列研究比缺点也较突出,即研究者在确定暴露因素、研究人群、研究结局时受到已有记录的限制,对混杂因素亦无法实施控制,易发生偏倚。

队列研究的优点:

- (1) 研究者可直接观察或测量研究因素和结局,所以资料翔实.
- (2) 因结局发生在后,故对暴露资料的收集是无偏倚的;在前瞻性队列研究中,亦可收集已知混杂因素的信息.
- (3) 可直接计算发病率、相对危险度等度量疾病与病因关联的指标.
- (4) 由于病因出现在前,疾病发生在后,因果现象发生的时间顺序合理,可证实病因假说.当暴露为稀有事件时,研究结论更为可靠.
- (5) 可获得多种结局资料.

队列研究的缺点:

- (1) 研究耗费时间、人力、经费.
- (2) 当观察结局为罕见事件(如发病率很低的疾病)时,需要的样本量相当大,在实际工作中不易做到.
- (3) 研究常常需要追踪随访较长一段时间,研究对象不易保持依从性,有时可能因得知研究目的而中途改变态度,造成偏倚.
- (4) 在观察开始时,暴露已被严格定义,在观察期内研究者不易对暴露重新定义或增加有意义的新暴露因素.

8. 答 病例对照研究是一种“由果推因”的回顾性观察研究,故又称回顾性研究.根据有无研究疾病或其他结局,将研究人群分为病例组和对照组,追溯过去某些暴露情况,比较两组暴露水平有无差异,从而得出结局与暴露有无关联的推断.病例对照研究具有下列特点.

优点:

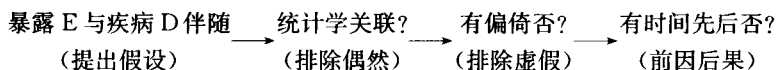
- (1) 与其他研究相比,病例对照研究相对节省时间、人力、经费,易组织实施.
- (2) 适于结局为罕见事件(如恶性肿瘤)的病因研究,当暴露因素较常见时,研究结论更为可靠.
- (3) 适于“潜伏期”较长疾病的研究.
- (4) 可获得与结局有关的多个病因因素的资料,往往是探索不明因素的有效途径.
- (5) 可计算比数比,以反映疾病与暴露的关联程度.当疾病罕见时,比数比与相对危险度近似.

缺点:

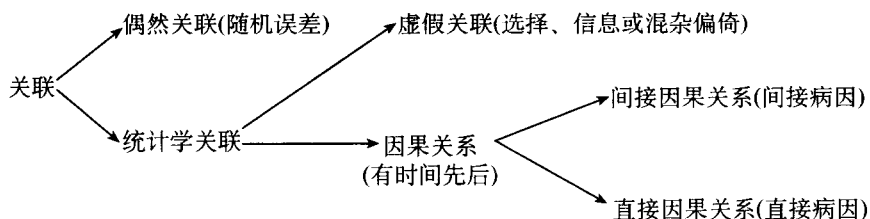
- (1) 不适于研究人群中暴露比例很低的因素,因这需要的样本量相当大,实际工作中不易做到.
- (2) 不能直接计算发病率.
- (3) 有时难以判断暴露与疾病之间的时间先后关系.
- (4) 易发生对照、病例的选择偏倚;确定暴露的测量偏倚,包括回忆偏倚、信息偏倚、观察者偏倚等等;混杂因素引起的影响“因果关系”推断的混杂偏倚.

9. 答 病因学研究旨在运用队列研究、病例对照研究等方法寻找可能的致病因素,分析它们之间的相互关系,及其对疾病发生、发展的影响,进而确定疾病的病因.

10. 答 病因推断过程如下:



联系的分类总结如下：



11. 答 病因学研究是一个去伪存真、由表面联系探究内在因果关系的过程。在这个过程中，人们首先见到的现象是不同事件伴随发生，这种相伴发生的关系称之为关联。有关联并非就是因果关系。在确定因果关系之前，先要确定关联是否可重复，是否稳定存在，而不是偶然事件或机会造成。这就需要一定数量的观察样本，通过统计学检验排除偶然关联。但是，由于实际观察样本的数量往往有限，因此要用统计学推断的方法估计犯错误概率（如 5% 或 1%）。排除了偶然关联，这时还只能称之为有统计学关联，因为研究中可能存在着各种偏倚而影响研究结果的真实性，得出虚假关联。所以在确定因果关系之前，还必须排除虚假关联。

12. 答 病因的判定标准有：

（1）联系的时间先后：病因一定要在疾病发生之前出现，并且两者之间应该有一个必要的潜伏期。

（2）联系的强度：研究因素与疾病联系的强度愈强，成为病因的可能性愈大。

（3）联系的一致性：研究因素与疾病的联系不同时间、不同地区、不同人群的研究结果一致；基础、临床及人群的研究从各个方面也都得到相同的结论。这种联系的一致性或恒定性说明研究的因素是危险因素。

（4）剂量效应关系：研究因素的强度增强或减弱，疾病发生的频率也应相应地增高或降低。一般只是在一定剂量范围内存在剂量效应关系。

（5）联系的特异性：某一因素只与某一特定的疾病发生有关，而与其他疾病无关或关系不大，这种现象称为联系的特异性。对于传染病或一些病因简单的非传染病，联系的特异性不难发现。但是，对于大多数非传染病，特别是病因不明疾病，这些疾病并往往无特异的致病因素，每一致病因素并非疾病发生的必要条件。

（6）联系的合理性。

第四章 数据特征与统计描述

练习题

一、最佳选择题

1. 编制频数表时,分组数目一般取().
A. 5~10 组 B. 8~15 组 C. 10~30 组 D. 15~20 组 E. 越多越好
2. 描述一组正态分布资料的离散程度,以()指标较好.
A. 极差 B. 离均差 C. 标准差 D. 离均差平方和 E. 变异系数
3. 描述一组正态分布资料的集中程度,以()指标较好.
A. 算术均数 B. 几何均数 C. 中位数 D. 四分位数 E. 百分位数
4. 对成倍增长的计量资料描绘其集中趋势,宜用().
A. 算术均数 B. 几何均数 C. 中位数 D. 方差 E. 百分位数
5. 若比较身高、体质量资料的变异度,宜用().
A. 标准差 B. 离均差 C. 四分位数间距 D. 变异系数 E. 极差
6. 调查某地中学生的近视情况,若描述近视学生的年龄分布可用().
A. 普通线图 B. 直方图 C. 半对数线图 D. 圆图 E. 条图
7. 比较某地区解放以来三种病的发病率在各年度的发展速度,宜绘制().
A. 普通线图 B. 百分条图 C. 半对数线图 D. 圆图 E. 条图
8. 欲表示某地区 2003 年 SARS 病人的职业构成,可绘制().
A. 单式条图 B. 圆图 C. 直方图 D. 线图 E. 散点图

二、问答题

1. 统计描述主要从哪几个方面发现和描述数据特征?
2. 频数表的主要用途有哪些?
3. 算术均数、几何均数和中位数各有什么适用条件?
4. 标准差有何用途?
5. 变异系数与标准差有何异同?
6. 应用相对数应注意些什么?
7. 简述统计表的主要结构.
8. 简述统计图的主要结构.

三、计算题

1. 某市 110 名健康女大学生血清总蛋白(g/L)测量资料如表 4-1.
(1) 编制频数分布表并绘制直方图,简述其分布特征;
(2) 计算均数与中位数;

(3) 计算标准差和变异系数.

表 4-1 110 名健康女大学生血清总蛋白含量($\text{g} \cdot \text{L}^{-1}$)

73.5	71.2	67.9	80.5	80.5	70.8	78.0	68.8	78.8	74.6	74.3
73.5	79.5	70.4	76.5	72.7	77.2	84.4	75.0	76.5	70.4	72.8
70.5	75.6	75.0	78.8	72.0	74.3	71.2	72.0	73.0	71.8	74.0
68.0	69.7	71.2	74.3	65.0	75.8	74.3	78.8	73.5	75.0	76.3
73.5	73.5	74.3	69.7	80.3	75.8	64.5	72.0	75.0	73.5	74.3
68.0	70.4	75.0	81.2	71.2	70.4	76.5	68.9	75.8	76.8	73.2
74.3	75.0	72.0	67.3	77.6	76.5	74.3	76.5	74.3	76.5	72.0
74.0	72.7	72.7	75.4	82.6	76.5	65.0	74.7	73.5	79.5	73.5
72.7	67.3	67.3	67.3	68.8	77.2	70.4	72.7	76.5	67.2	74.5
74.3	70.3	82.6	72.7	73.5	73.5	72.7	75.0	73.5	76.8	72.7

2. 某防疫站对 30 名麻疹易感儿童经气溶胶免疫一个月后,测得其血凝抑制抗体滴度资料如下,试计算其平均滴度.

抗体滴度	1:8	1:16	1:32	1:64	1:128	1:256	1:512	合计
例 数	2	6	5	10	4	2	1	30

3. 50 例链球菌咽峡炎患者的潜伏期如下,试计算均数、中位数、几何均数,并说明何者的代表性较好.

潜伏期(小时)	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

4. 分别将表 4-2 中两种疾病发病率的历年变动情况绘制成普通线图和半对数线图,并说明两种图形的不同意义.

表 4-2 某单位 1990—1998 年普通感冒与支气管炎发病率/%

年 份	感 冒	支 气 管 炎
1990	126.27	6.37
1991	92.19	6.37
1992	107.59	5.90
1993	101.93	5.69
1994	92.60	5.49
1995	73.20	4.32
1996	51.40	3.04
1997	42.39	2.42
1998	33.92	2.27

5. 将以表 4-3 和表 4-4 中的资料绘成适合的图形.

表 4-3 1998 年不同地区艾滋病流行情况

地区	成人受感染率/%	地区	成人受感染率/%
北非及中东	0.13	拉丁美洲	0.57
西欧	0.25	南亚及东南亚	0.69
北美洲	0.56		

表 4-4 某年某地性病传播途径分布情况

传播途径	病例数	构成比/%
非婚性接触	413303	72.1
配偶接触	103064	18.0
其他传播	57174	9.9

练习题参考答案

一、最佳选择题解答

1. B 2. C 3. A 4. B 5. D 6. B 7. C 8. B

二、问答题解答

1. 答 可以从以下几个方面考虑:

(1) 判断数据的分布类型. 对大样本资料可编制频数表或绘频数分布图.

(2) 根据数据的分布类型, 确定描述统计量.

(3) 集中趋势描述. 如对于正态分布资料, 首选算术均数; 对于偏态分布资料, 可选中位数; 对于几何级数资料, 则选几何均数.

(4) 离散趋势描述. 如对于正态分布资料, 首选标准差; 对于偏态分布资料, 则选四分位数间距.

(5) 对于正态分布资料, 基本的描述统计量为 n , $\bar{X}$ 和 S .

2. 答 频数分布表的用途是:

(1) 描述频数分布的类型;

(2) 描述频数分布的特征;

(3) 便于发现一些特大或特小的可疑值;

(4) 便于进一步作统计分析和处理.

3. 答 算术均数主要适合描述正态分布资料的集中位置, 几何均数适合描述当资料呈倍数关系或对数正态分布时的集中趋势; 当大部分观测值比较集中, 少数观测值偏向一侧时, 或资料分布情况不清楚时, 或数据的最大值(最小值)无准确测量数据时, 宜用中位数来表示其平均水平.

4. 答 标准差的用途归纳起来有:

(1) 表示变量分布的离散程度. 在两组资料(总体或样本)均数相近、度量单位相同的条件