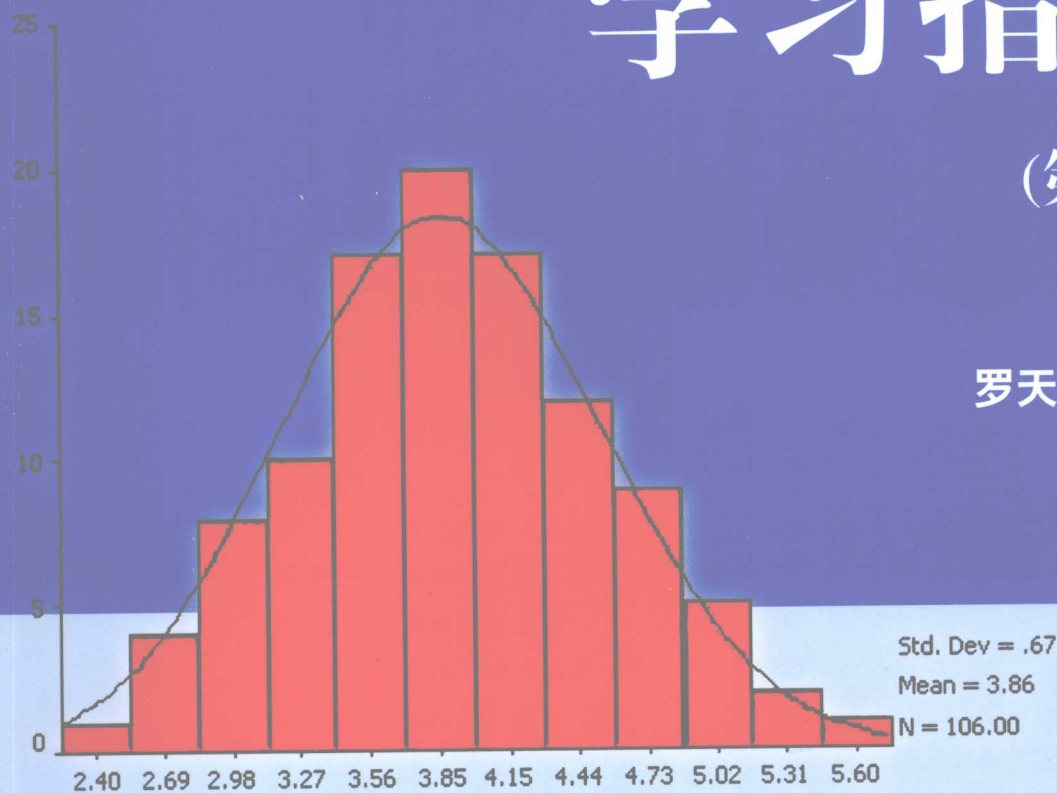


普通高等教育“十一五”国家级规划教材配套教材

医学统计学 学习指导

(第2版)

罗天娥 主编



X

中国协和医科大学出版社

● 普通高等教育“十一五”国家级规划教材配套教材

医学统计学学习指导

(第2版)

主 编 罗天娥

副主编 赵晋芳

编委 (以姓氏笔画为序)

于 浩	南京医科大学	肖 琳	中国疾病预防控制中心
仇丽霞	山西医科大学	陈长生	第四军医大学
尹 平	华中科技大学	陈平雁	南方医科大学
王立芹	河北医科大学	易 东	第三军医大学
田考聪	重庆医科大学	金水高	中国疾病预防控制中心
刘 艳	哈尔滨医科大学	罗天娥	山西医科大学
刘玉秀	南京军区南京总医院	赵晋芳	山西医科大学
刘桂芬	山西医科大学	郜艳晖	广东药学院
吕 桦	上海市浦东新区 CDC	凌 莉	中山大学
余红梅	山西医科大学	宿 庄	内蒙古医学院
吴艳乔	华西医科大学	阎玉霞	南方医科大学
张岩波	山西医科大学	黄高明	广西医科大学
李新华	贵阳医学院	韩少梅	中国协和医科大学

参编人员

萨 建 王晓芳 曾 平 原 静 尉 洁 田娇妮 刘志勇

中国协和医科大学出版社

图书在版编目 (CIP) 数据

医学统计学学习指导 / 罗天娥主编. —北京: 中国协和医科大学出版社, 2009. 2
普通高等教育“十一五”国家级规划配套教材
ISBN 978 - 7 - 81136 - 135 - 3

I. 医… II. 罗… III. 医学统计 - 高等学校 - 教学参考资料 IV. R195.1

中国版本图书馆 CIP 数据核字 (2008) 第 210814 号

● 普通高等教育“十一五”国家级规划教材配套教材

医学统计学学习指导 (第 2 版)

主 编: 罗天娥
责任编辑: 吴桂梅 姜淑惠

出版发行: 中国协和医科大学出版社
(北京东单三条九号 邮编 100730 电话 65260378)

网 址: www.pumcp.com
经 销: 新华书店总店北京发行所
印 刷: 三河市华晨印务有限公司

开 本: 787 × 1092 毫米 1/16 开
印 张: 11
字 数: 270 千字
版 次: 2009 年 3 月第二版 2009 年 3 月第一次印刷
印 数: 1—5000
定 价: 20.00 元

ISBN 978 - 7 - 81136 - 135 - 3/R · 135

(凡购本书, 如有缺页、倒页、脱页及其他质量问题, 由本社发行部调换)

前 言

作为医学教育的公共必修课医学统计学，其理论知识体系和学习方法技能等都有别于其他医学课程。学习医学统计学必须在理解原理和基本概念的基础上，通过亲自动手完成一定数量、不同形式的习题练习，融会贯通，才能更好地掌握本门课程。为结合计算机软件的发展与应用，保证医学统计学学习质量，进一步提高学生分析与解决医学实际问题的能力，我们采用分工编写、集体讨论、按各章知识体系及章节间联系，再次编写出版本书。

本书作为普通高等教育“十一五”国家级规划教材《医学统计学》配套教材，传承了我校已故著名医学统计学家何大卫教授 1997 年主编出版的《卫生统计学习题解答》的思想；在余红梅教授 2005 年主编出版的《卫生统计学学习指导》（中国协和医科大学出版社《卫生统计学》配套教材）基础上；进一步汇集《医学统计学》教材编委意见，承蒙众多国内同行和学生的诸多好评与鼓励，再次组织编写了本配套教材。全书共分六个单元，每个单元包括目的要求、内容精要、练习题与参考答案以及补充题及参考答案等。目的要求主要介绍每一章按教学大纲要求掌握、熟悉和了解的内容；内容精要以简明扼要的形式总结概括该章的主要内容，可作为读者实习或复习的参考；练习题包括判断题、单选题、计算分析题、SAS 结果解释与应用题、简答题和讨论题，书中给出了各类习题的参考答案。有别于国内同类参考书的是补充了大量的、不同题型的、因原教材篇幅所限未能编入的习题，供读者参考学习。

本书使用中，建议初学者切勿死记硬背或不求甚解，一定要结合教材讲授理论，联系医学实际，认真思考小结，亲自解析或集体讨论后再核对答案，以便更深入地理解医学统计学的基本原理和概念。

本书出版之际，真诚地感谢《医学统计学》教材编委的鼎力支持，山西医科大学卫生统计学教研室博士、硕士研究生参加的热烈讨论，教研室刘桂芬、仇丽霞、余红梅、张岩波等老师们核校答案无私奉献的辛勤劳动，并提出了许多宝贵的建议和意见。

鉴于编者能力所限，书中难免纰漏，诚恳读者批评指正。

罗天娥

2008 年 11 月 26 日

目 录

第一单元 医学资料的统计描述 (1~3 章)	(1)
目的要求	(1)
内容精要	(1)
教材练习题及参考答案	(5)
补充题及参考答案	(13)
第二单元 定量资料的统计推断 (4~5 章)	(21)
目的要求	(21)
内容精要	(21)
教材练习题及参考答案	(24)
补充题及参考答案	(45)
第三单元 分类资料的统计推断 (6~8 章)	(54)
目的要求	(54)
内容精要	(54)
教材练习题及参考答案	(58)
补充题及参考答案	(74)
第四单元 回归与相关 (9、14、15 章)	(83)
目的要求	(83)
内容精要	(83)
教材练习题及参考答案	(85)
补充题及参考答案	(104)
第五单元 医学统计设计 (10~12 章)	(108)
目的要求	(108)
内容精要	(108)
教材练习题及参考答案	(110)
补充题及参考答案	(121)
第六单元 医学统计的应用 (13、16~22 章)	(135)
目的要求	(135)
内容精要	(135)
教材练习题及参考答案	(148)
补充题及参考答案	(163)

第一单元 医学资料的统计描述 (1~3 章)

目的要求

1. 了解医学统计学的意义和基本内容;熟悉统计工作的基本步骤;掌握医学统计学中总体与样本、频率与概率及统计资料的分类。
2. 熟悉频数分布表的编制与频数分布图的绘制方法;掌握频数分布的特征和类型;掌握描述计量资料常用指标的计算与适用范围。
3. 掌握相对数(率、构成比、相对比)的概念,计算方法;掌握应用相对数时应注意的问题;掌握率的标准化及其计算方法;了解动态数列及常用指标。
4. 了解正态分布的概念与特征,掌握正态曲线下面积分布规律;了解标准正态分布的概念,掌握标准化变换;掌握医学参考值范围的意义与制定方法。

内容精要

第1章 绪 论

一、医学统计学的意义和基本内容

医学统计学是运用概率论和数理统计的基本原理和方法,研究数据的搜集、清理、分析和推断的一门应用学科。基本内容包括医学研究设计、统计学基本原理与方法及医学统计应用。

二、统计工作的基本步骤

1. 医学研究设计 包括对医学研究资料的收集、整理和分析全过程的总设想或安排。
2. 搜集资料 按科研设计要求和目的,获取准确、可靠、有意义的原始数据。
3. 清理资料 按设计要求将一些分散的、表现个体特征的原始数据系统化、条理化,以便更好地揭示所研究事物的规律和特征。
4. 统计分析 包括统计描述和统计推断。统计描述是按研究设计要求,计算相应的统计指标,选用适当的统计表或统计图来说明数据特征;统计推断是根据抽样原理,在概括样本信息特征的基础上,对所研究总体的特征进行推断。

三、统计学中的基本概念

1. 统计资料类型 根据变量的观察结果不同,将变量分为定量变量、定性变量和顺序变量。

(1) 定量变量:通常由仪器、工具或其他定量方法测定某项指标而得到,一般带有度量衡单位。由一组同质的定量变量值所组成的资料称为定量资料。其变量值 x 可以是实数轴

上的一个区间,表现为连续型随机变量,也可以是整数范围内的随机变量。同组变量值之间,没有质的不同,只有量的差异。

(2) 定性变量:其变量值表现为不同的属性。变量的观察结果只具有相互对立的两种情况,称作二分类变量;变量的观察结果表现为相互对立的多种情况,称作多分类变量。若按事物属性归组,清点各组观察单位数得到的资料称为定性资料。定性资料同类变量值间无量的区分,异类变量值间性质截然不同。

(3) 顺序变量:其变量的观察结果按不同的类别、等级或程度分组,且不同组之间有顺序存在,如检验结果 $-$ 、 $\pm$ 、 $+$ 、 $++$ 、 $+++$ 、 $++++$ 。若按等级分组,清点各组观察单位数而得到的资料,称为等级资料或半定量资料。这类资料变量值之间不仅存在类别的不同,且不同分组间也有顺序、等级或量的差别,而这种差别又难以精确量化。

2. 总体与样本 总体(population)指特定研究对象中所有观察单位的测量值。可分为有限总体和无限总体。总体中的所有单位都能够标识者为有限总体;反之,为无限总体。从总体中随机抽取部分观察单位,其测量结果的集合称为样本(sample)。样本应具有代表性。所谓有代表性的样本,是指用随机抽样方法获得的样本。

3. 频率与概率 频率(frequency)是指在相同条件下,进行有限 n 次重复试验,某随机事件 A 发生次数 X 与 n 次试验的比值,其值介于 $0\sim 1$ 之间。概率(probability)是描述某随机事件 A 发生可能性大小的度量,常记作 P ,可用小数或百分数表示,其值 $0\leq P\leq 1$ 。医学统计研究中,常把 $P\leq 0.05$ 的随机事件称作小概率事件,表明该随机事件发生的可能性很小。

第2章 医学资料的统计描述

一、定量资料的频数分布表

频数表(frequency table)通过资料清理,用来显示数据分布的范围、数据集中的区间和分布的形态。

1. 频数分布有两个重要特征 即集中趋势(central tendency)和离散趋势(tendency of dispersion)。

2. 频数分布的类型 对称分布和偏态分布(又分正偏态分布和负偏态分布)。

二、定量资料的频数分布图

根据频数表资料绘制频数分布直方图。以变量 X 大小为横坐标,频数或百分数为纵坐标,用矩形面积大小表示频数多少,即为频数分布直方图。它比频数表更直观、更形象地描述频数分布的情况。

三、定量资料集中趋势指标与离散趋势指标

表 1-1 常用平均数的意义、应用场合及计算

指 标	意 义	应用场合	计 算
均数 (mean, $\bar{x}$)	平均数量水平	对称分布，特别是 正态分布或近似正 态分布资料	直接法: $\bar{x} = \frac{\sum x}{n}$ 加权法: $\bar{x} = \frac{\sum fx}{\sum f}$
几何均数 (geometric mean, G)	平均增（减）倍数	对数正态分布资料	直接法: $G = \lg^{-1} \left(\frac{\sum \lg x}{n} \right)$ 加权法: $G = \lg^{-1} \left(\frac{\sum \lg x}{\sum f} \right)$
中位数 (median, M)	排序后位次居中的观 察值水平	偏态分布； 分布不明； 分布末端无确定值	直接法: n 为奇数 $M = x_{(\frac{n+1}{2})}$ n 为偶数 $M = [x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}] / 2$ 频数表法: $M = L + \frac{i}{f_x} \left(\frac{n}{2} - \sum f_L \right)$

表 1-2 常用变异指标的意义、应用场合及计算

指 标	意义	应用场合	计 算
极差（全距） (range, R)	个体差异 范围	说明传染病，食物 中毒等的最短、最长潜伏 期等	一组观察值中最大值与最小值之差
四分位数间距 (quartile range, QR)	个体变异 程度	描述偏态分布、分布不 明、分布末端无确定值资 料的离散程度	$QR = Q_U - Q_L = P_{75} - P_{25}$ $P_x = L_x + \frac{i_x}{f_x} (nx\% - \sum f_L)$
方差 (variance, σ^2, s^2) 标准差 (standard deviation, σ, s)	个体变异 程度	描述正态分布或近似正态 分布资料的离散程度	直接法: $s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} = \sqrt{\frac{\sum x^2 - (\sum x)^2 / n}{n - 1}}$ 加权法: $s = \sqrt{\frac{\sum fx^2 - (\sum fx)^2 / \sum f}{\sum f - 1}}$
变异系数 (coefficient of variation, CV)	相对变异 程度	比较度量衡单位不同或单 位相同但均数相差悬殊的 多组资料的变异度	$CV = s / \bar{x} \times 100\%$

四、常用的相对数及其计算方法

1. 率（rate） 又称频率指标，用来说明在一定条件下某现象发生的频率或强度。

$$\text{率} = \frac{\text{实际发生某现象的观察单位数}}{\text{可能发生该现象的观察单位总数}} \times K$$

2. 构成比（proportion） 用来说明某一事物内部各组成部分所占的比重或分布。

$$\text{构成比} = \frac{\text{事物内部某一组成部分的观察单位数}}{\text{事物内部各组成部分观察单位总数}} \times 100$$

3. 比 (ratio) 又称相对比, 表示两个有关指标之比, 可用倍数或百分数表示。

$$\text{比} = \frac{\text{甲指标}}{\text{乙指标}}$$

五、相对数应用时应注意的问题

1. 计算相对数时分母要适宜, 一般不宜太小。
2. 区分率与构成比的意义和应用 构成比只能说明事物内部各组成部分的比重或分布, 并不能反映某现象发生的频率或强度。最常见的错误之一就是利用构成比代替率来分析和解释某研究现象。应用相对数时, 需正确选择统计指标。

3. 正确计算总率。
4. 注意资料的可比性。
5. 对样本率 (或构成比) 的比较需作假设检验。

六、率的标准化及其计算方法

1. 标准化法 (standardization method) 的基本思想, 就是将各对比组采用统一的内部构成 (消除混杂因素影响), 分别计算标准化率后再作比较。

2. 使用标准化率时的注意事项

- (1) 选定标准的原则。
- (2) 若资料中各分组率出现交叉, 不宜采用标准化法。
- (3) 所得标准化率不反映某现象发生的实际水平。
- (4) 样本标准化率的比较仍需作假设检验。

七、动态数列及常用指标

1. 动态数列 将一系列统计指标 (绝对数、相对数、平均数等) 按时间 (时点或时期) 顺序排列起来所组成的数列称动态数列 (dynamic series), 用以表示事物现象在时间上的变化和发展趋势。

2. 动态数列的分析常用绝对增长量、发展速度、增长速度、平均发展速度和平均增长速度等指标。

第3章 正态分布及其应用

一、正态分布

正态分布 (normal distribution), 也称高斯分布, 是一种重要的连续型分布, 它是以均数为中心, 中间高、两边逐渐减少、左右对称分布。 μ 表示总体均数, 用以描述正态分布的集中位置, 又称位置参数; σ 表示总体标准差, 用以描述正态分布的离散程度, 又称为变异度参数。常用 $X \sim N(\mu, \sigma^2)$ 表示均数为 μ , 标准差为 σ 的正态分布。

正态曲线下与横轴间的总面积为 1, 以 μ 为中心左右两侧对称。越靠近 μ 处, 曲线下面积越集中, 两边逐渐减少。理论上 $\mu \pm \sigma$ 、 $\mu \pm 1.645\sigma$ 、 $\mu \pm 1.96\sigma$ 及 $\mu \pm 2.58\sigma$ 范围内曲线下的面积分别占总面积的 68.27%、90.00%、95.00% 及 99.00%。

二、标准正态分布

对任意服从 $N(\mu, \sigma^2)$ 分布的变量，经 u 变换都可转换为 $\mu = 0$ 、 $\sigma = 1$ 的分布，即标准正态分布，记作 $u \sim N(0, 1)$ 。

$$u = \frac{x - \mu}{\sigma}$$

该变换称变量的标准化变换，变换后的 u （有时用 Z 表示）称为标准化正态离差（standardized deviate）。实际应用中，经 u 变换后，均可把求解任意一个正态分布曲线下面积的问题，转化成标准正态分布曲线下相应的面积问题。当 u 从 $-\infty$ 到 0 时， $\Phi(-u)$ 为 u 左侧曲线下对应的面积，根据正态分布概率曲线的对称性， $\Phi(u) = 1 - \Phi(-u)$ 。我们可求出在任何一个区间内标准正态分布曲线下的面积，也就是 u 落在该区间内的概率。

三、正态分布的应用

1. 制定医学参考值范围 对正态（或近似正态）分布资料以及可以通过转换后服从正态分布的资料，可用正态近似法；对于偏态或未知分布型的资料可选用百分位数法。

表 1-3 参考值范围的估计方法

参考值范围 (%)	正态近似法			百分位数法		
	双侧	单侧下限	单侧上限	双侧	单侧下限	单侧上限
90	$\bar{x} \pm 1.645s$	$\bar{x} - 1.282s$	$\bar{x} + 1.282s$	$P_5 \sim P_{95}$	P_{10}	P_{90}
95	$\bar{x} \pm 1.96s$	$\bar{x} - 1.645s$	$\bar{x} + 1.645s$	$P_{2.5} \sim P_{97.5}$	P_5	P_{95}
99	$\bar{x} \pm 2.58s$	$\bar{x} - 2.33s$	$\bar{x} + 2.33s$	$P_{0.5} \sim P_{99.5}$	P_1	P_{99}

2. 质量控制 常以 $\bar{x} \pm 2s$ 作为上、下警戒线，以 $\bar{x} \pm 3s$ 作为上、下控制值。

教材练习题及参考答案

一、是非题

1. 要对比身高与体重的变异程度，既可用标准差，也可用变异系数。
2. 中位数指一组观察值中，位次居中观察值的大小，常用来说明该组数据的平均水平。
3. 用百分位数法确定双侧 95% 正常值范围的下限值是 P_5 ，上限值是 P_{95} 。
4. 测定 9 名男性新生儿的出生体重（kg）为 3.26、3.01、3.40、3.15、3.40、3.00、3.60、3.20 和 3.35，则体重中位数为 3.40kg。
5. 某病住院日的频数分布呈单峰分布，平均住院日均数为 10 日，中位数为 5 日，可以看出该病住院日呈正偏态分布。
6. 标准化病死亡率比实际病死率能更好地反映该现象发生的实际水平。
7. 已知甲乙两地肝癌死亡总数和各年龄组人口数，宜选用直接法计算标准化率。
8. 动态数列指标计算可以用实际数、平均数，但不能用率。
9. 某医师用传统中药配方治疗 3 例肝癌病人，其中 2 例存活 2 年以上，则该药治愈率

为 66.7%。

10. 比较甲乙两医院乳腺癌术后 5 年生存率, 已知两医院有无腋下淋巴结转移的病例构成明显不同, 适当的方法是对病情标准化后再作对比。

【参考答案】 1 错误 2 错误 3 错误 4 错误 5 正确 6 错误 7 错误 8 错误 9 错误 10 正确

二、单项选择题

1. 下列资料_____为定性资料。

- A. 体重 (kg)
- B. 血型 (A、B、O、AB)
- C. 血红蛋白 (g/L)
- D. 坐高指数 (% , 坐高/身高)
- E. 病人的病情分级 (轻、中、重)

2. 频数分布的两个重要特征是_____。

- A. 统计量与参数
- B. 集中趋势与离散趋势
- C. 样本均数与总体均
- D. 标准差与标准误
- E. 样本与总体

3. 表示一组对称分布资料变量值的平均水平, 宜选用_____。

- A. 算术均数 ($\bar{x}$)
- B. 方差
- C. 几何均数 (G)
- D. 标准差
- E. 变异系数

4. 下列有关中位数 (M) 描述不正确的有_____。

- A. 中位数 (M) 是一组观测值中最大值与最小值之差
- B. 中位数 (M) 是一组观测值排序后, 位次居中观测值的水平
- C. 一组观察值中, 比中位数小的观测值有 50%
- D. n 为偶数时, $M = [X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}] / 2$
- E. $M = P_{50}$

5. 医学研究中, 反映一组血清抗体效价资料的平均水平, 常选用_____。

- A. $\bar{x}$
- B. M
- C. G
- D. R
- E. CV

6. 下列有关四分位数间距描述不正确的是_____。

- A. 四分位数间距为 ($P_{75} - P_{25}$)

- B. 四分位数间距比极差稳定
 - C. 四分位数间距即居中 50% 观测值的极差
 - D. 四分位数间距主要用于描述正态分布资料的变异度
 - E. 四分位数间距越大, 表示变异度越大
7. 正态曲线下, 横轴上从 $-\infty$ 到均数的面积为_____。
- A. 50%
 - B. 95%
 - C. 97.5%
 - D. 99%
 - E. 不能确定
8. 若随机变量 X 服从 $N(\mu, \sigma^2)$ 的正态分布, 则 X 的第 97.5 百分位数等于_____。
- A. $\mu - 1.96\sigma$
 - B. $\mu - 1.645\sigma$
 - C. $\mu - 1\sigma$
 - D. $\mu + 1.645\sigma$
 - E. $\mu + 1.96\sigma$
9. 正态分布 $N(\mu, \sigma^2)$, 当 μ 恒定时, σ 越小_____。
- A. 曲线沿横轴越向左移动
 - B. 曲线沿横轴越向右移动
 - C. 曲线越“瘦高”
 - D. 曲线越“矮胖”
 - E. 曲线形状和位置不变
10. 健康男子收缩压的正常值范围一般指_____。
- A. 所有健康成年男子收缩压的波动范围
 - B. 绝大多数正常成年男子收缩压的波动范围
 - C. 所有正常成年男子收缩压的波动范围
 - D. 少部分正常成年男子收缩压的波动范围
 - E. 所有正常人收缩压的波动范围
11. 反映某事物现象实际发生严重程度的指标宜采用_____。
- A. 年龄发病构成比
 - B. 年龄别发病率
 - C. 性别发病比
 - D. 发展速度环比
 - E. 标化率
12. 麻疹疫苗接种后血清阳转率的分母应为_____。
- A. 麻疹患者人数
 - B. 麻疹易感人数
 - C. 获得麻疹终身免疫人数

- D. 麻疹疫苗接种人数
- E. 麻疹疫苗接种后的阳转人数

13. 经调查甲、乙两市的冠心病粗病死率为 45/10 万，按标准人群年龄构成标化后，甲市标化率为 52/10 万，乙市为 40/10 万，故可认为_____。

- A. 甲市年龄别人口构成较乙市年轻
 - B. 甲市年龄别人口构成较乙市老化
 - C. 甲市冠心病诊断较乙市准确
 - D. 乙市冠心病诊断较甲市准确
 - E. 甲乙市冠心病患病率差别有统计学意义
14. 欲比较两工厂同工种工人某职业病患者率的高低，需假设_____。

- A. 两厂该职业病的患病年龄构成相同
- B. 两厂工人数相同
- C. 两厂患病人数相同
- D. 两厂该工种工人的工龄构成相同
- E. 两厂工人工种构成相同

15. 用某新药治疗急性腹泻患者 31 例，1 周后痊愈 25 例，由此可认为_____。
- A. 新药疗效好
 - B. 该新药疗效一般
 - C. 该新药只有近期疗效
 - D. 此治疗例数少，可以用可信区间推论治疗情况
 - E. 无法说明该新药疗效是否有意义

【参考答案】 1 B 2 B 3 A 4 A 5 C 6 D 7 A 8 E 9 C 10 B 11 B 12 D

13 A 14D 15E

三、计算分析题

1. 某研究组测得 120 名女大学生的收缩压（mmHg）并编制频数表如下：

表 1-4 120 名女大学生的收缩压（mmHg）频数分布表

组 段	组中值 (x)	频数 (f)	频率 (%)	累计频数	累计频率 (%)
80 ~	82.5	3	2.50	3	2.50
85 ~	87.5	7	5.83	10	8.33
90 ~	92.5	15	12.50	25	20.83
95 ~	97.5	19	15.83	44	36.67
100 ~	102.5	38	31.67	82	68.33
105 ~	107.5	16	13.33	98	81.67
110 ~	112.5	12	10.00	110	91.67
115 ~	117.5	7	5.83	117	97.50
120 ~125	122.5	3	2.50	120	100.00
合计	—	120	100.00	—	—

①计算均数、标准差和变异系数；②计算中位数和四分位数间距；③试说明采用哪些指标描述其分布特征比较合适，并说明理由；④估计该地健康女大学生的收缩压在 90 mmHg 以下者所占的比例；⑤估计该地健康女大学生收缩压 95% 参考值范围。

①计算均数、标准差和变异系数：进入计算器统计状态，将各组段的组中值与相应的频数输入，得

$$\bar{x} = \sum fx / \sum f = 102.125 \text{ (mmHg)}$$

$$s = \sqrt{\frac{\sum fx^2 - \frac{(\sum fx)^2}{\sum f}}{\sum f - 1}} = 8.749 \text{ (mmHg)}$$

$$CV = \frac{s}{\bar{x}} \times 100\% = \frac{8.749}{102.125} \times 100\% = 8.567\%$$

②求中位数和四分位数间距

$$M = L + \frac{i}{f_m} \left(\frac{n}{2} - \sum f_L \right) = 100 + \frac{5}{38} \times \left(\frac{120}{2} - 44 \right) = 102.11 \text{ (mmHg)}$$

$$QR = Q_U - Q_L = P_{75} - P_{25} = \left[105 + \frac{5}{16} (120 \times 75\% - 82) \right] - \left[95 + \frac{5}{19} (120 \times 25\% - 25) \right] \\ = 107.5 - 96.32 = 11.18 \text{ (mmHg)}$$

③选择适宜指标描述其分布特征：频数分布表集中在 95 ~ 105 组段内，中间高，两侧频数减少，左右对称。由频数表可以看出女大学生收缩压近似服从正态分布。宜选用算术均数和标准差来描述该组资料的集中趋势和离散趋势。

④估计该地健康女大学生的收缩压在 90 mmHg 以下者所占的比例：90 mmHg 以下者占总数的 $10/120 \times 100\% = 8.33\%$ 。

⑤估计该地健康女大学生收缩压 95% 参考值范围：

$$\text{下限: } \bar{x} - u_{\alpha/2}s = 102.125 - 1.96 \times 8.749 = 84.98 \text{ (mmHg)}$$

$$\text{上限: } \bar{x} + u_{\alpha/2}s = 102.125 + 1.96 \times 8.749 = 119.27 \text{ (mmHg)}$$

该地健康女大学生收缩压 95% 参考值范围为 (84.98, 119.27) (mmHg)。

2. 某地 10 人接种某种疫苗后，测定其抗体效价如下：1:2, 1:2, 1:4, 1:4, 1:4, 1:4, 1:8, 1:8, 1:16, 1:32，求该疫苗接种后平均抗体效价。

该疫苗接种后平均抗体效价用几何均数表示：

$$G = \lg^{-1} \left(\frac{\lg x_1 + \lg x_2 + \cdots + \lg x_n}{n} \right) \\ = \lg^{-1} \left(\frac{\lg 2 + \lg 2 + \lg 4 + \lg 4 + \lg 4 + \lg 4 + \lg 8 + \lg 8 + \lg 16 + \lg 32}{10} \right) \\ = 5.657$$

该疫苗接种后平均抗体效价为 1:5657。

3. 某地 220 例正常成年人血铅含量 ($\mu\text{mol/L}$) 的频数分布如下，试估计该地正常成年人血铅含量的 95% 参考值范围。

表 1-5 220 名正常成年人血铅含量 (μmol/L) 的频数分布表

组段	频数 (f)	频率 (%)	累计频数	累计频率 (%)
0 ~	12	5.45	12	5.45
0.24 ~	55	25.00	67	30.45
0.48 ~	52	23.64	119	54.09
0.72 ~	36	16.36	155	70.45
0.96 ~	28	12.73	183	83.18
1.20 ~	11	5.00	194	88.18
1.44 ~	13	5.91	207	94.09
1.68 ~	4	1.82	211	95.91
1.92 ~	5	2.27	216	98.18
2.16 ~	1	0.45	217	98.64
2.40 ~	2	0.91	219	99.55
2.64 ~	1	0.45	220	100.00
合计	220	100.00	-	-

由表可知, 本资料呈偏态分布, 应选用百分位数法估计血铅参考值范围合适, 在 0.24 ~ 内频数分布最高, 随血铅增高, 频数分布函数减少, 拖尾向右, 据专业知识知血铅含量越低越好, 应计算单侧参考值范围 P_{95} 。

$$P_{95} = L_x + \frac{i_x}{f_x}(nx\% - \sum f_L) = 1.68 + \frac{0.24}{4} (220 \times 95\% - 207) = 1.80 (\mu\text{mol/L})$$

该地正常成年人 95% 血铅含量参考值范围不得高于 1.80μmol/L。

4. 某医师在甲地检查 40 岁以下居民 7000 人, 其中发病 315 人, 40 岁及以上 3000 人, 发病 30 人; 乙地 40 岁以下居民 3000 人, 发病人数 150 人, 40 岁及以上 7000 人, 发病人数 140 人。欲比较甲、乙两地该病发病率, 问应如何平衡年龄因素的影响, 列表计算并解释标准化发病率与实际发病率有何不同?

表 1-6 两地某病患病情况

年龄 (岁)	甲 地		乙 地	
	检查人数	发病人数	检查人数	发病人数
40 岁以下	7000	315	3000	150
40 岁及以上	3000	30	7000	140
合计	10000	345	10000	290

表 1-7 甲、乙两地用合并的“标准人口数”计算标准化发病率(‰)

年龄组	标准 人口数	甲 地		乙 地	
		原死亡率 (‰)	预期死亡 数	原死亡率 (‰)	预期死亡 数
40 岁以下	10000	45	450	50	500
40 岁及以上	10000	10	100	20	200
合计	20000	34.5	550	29	700

甲、乙两地人口的年龄构成不同，欲比较甲、乙两地某病发病率应该采用标准化的方法平衡年龄因素的影响。

甲地标准化发病率 = $550/20000 \times 1000‰ = 27.5‰$

乙地标准化发病率 = $700/20000 \times 1000‰ = 35‰$

甲地各年龄组发病率低于乙地，但总发病率却高于乙地。这是因为甲地 40 岁以下人口构成较乙地高，而 40 岁以下死亡率高于 40 岁及以上，导致甲地发病人数增多，故总发病率增高。若直接根据两地区总率作比较，则会得出甲地比乙地某病的发病状况严重的错误结论。我们采用标准化法计算甲、乙两地的标准化发病率，可以看出乙地高于甲地。

5. 某地 1990 年 6 月 30 日有男性人口 13697600 人，男性人口中因心血管疾病死亡 18311 人，其中冠心病死亡 654 人，欲反映该地男性人口冠心病死亡强度与冠心病死亡在心血管疾病死亡中所占比重，应如何计算？

该地男性人口冠心病死亡强度应该选用该地男性人口冠心病死亡率：

冠心病死亡率 = $654/13697600 \times 100000/10 \text{ 万} = 4.77456/10 \text{ 万}$

冠心病死亡在心血管疾病死亡中所占比重用构成比：

冠心病死亡在心血管疾病死亡中所占比重 = $654/18311 \times 100\% = 3.57\%$

6. 某化工厂慢性气管炎患病与专业工龄的关系。①试计算构成比和患病率并填充在表中；②简要分析哪一工龄组气管炎患病最严重。

(1) 构成比和患病率如下表所示：

表 1-8 某化工厂慢性气管炎患病与专业工龄的关系

工龄 (年)	检查人数	患者数	构成比 (%)	患病率 (%)
1 ~	340	17	11.56	5.00
5 ~	254	30	20.41	11.81
10 ~	432	73	49.66	16.90
15 ~	136	27	18.37	19.85
合计	1162	147	100	12.65

(2) 描述一定条件下某现象发生强度的指标应选用频率。由表中患病率可见，随年龄增大，工龄越长，慢性气管炎患病率有逐渐增高的趋势，以 15 年以上工龄组气管炎患病

较重。

四、简述题

1. 试结合医学研究实例，阐述统计资料的类型。

医学统计资料可分为定量资料和分类资料，分类资料包括定性资料和等级资料。①定量资料是通过仪器或方法测量而得到的一组定量变量值所组成，如某医师测得 10 名 3 岁儿童血红蛋白含量 (g/L) 为 108、110、116、95、109、87、92、113、120、116；②分类资料是指按事物属性或类别分组，清点各组观察单位数而得到的资料。若分类特征是相互对应的两种属性，如性别变量表现为男性、女性，检验结果表达为阳性和阴性，成为二项分类变量；若分类特征表现为两种以上属性，称为多分类变量，如血型表现为 A 型、B 型、O 型、AB 型等，职业归属为工人、干部、农民、学生、自由职业者；③若分类特征表现为两种以上类别，且类别间有顺序性，称为等级资料，如职工体检眼底动脉硬化级别检查结果：正常 326 例，轻度硬化 18 例，中度硬化 18 例，重度硬化 3 例。

2. 举例说明 2 型糖尿病患者总体与样本的关系，采用随机抽样研究的意义何在？

总体 (population) 是根据研究目的确定的同质的所有观察单位某种变量值的集合。根据研究目的从研究总体中随机抽取反映总体特征的部分观察单位，其实测值组成样本 (sample)。为了解 2 型糖尿病分布特征及发病率，研究者可从该地随机抽取 500 名 2 型糖尿病患者。则该地所有 2 型糖尿病患者为目标总体，随机抽取的 500 名患者构成样本。

随机抽样研究的意义：①可节省大量的人力、物力、财力和时间，便于实施；②可估计抽样误差大小，便于进行统计推断；③可避免和减小人为干扰所产生的误差，达到事先给定的精度要求，保证样本的代表性；④有利于保证和提高调查质量，便于进行较深入的研究；⑤不宜进行普查的实际研究工作。

3. 试述医学参考值范围的意义及其确定的步骤。

医学参考值范围是指同质总体中某项指标大多数实测值的波动范围，也称为正常值范围，标准值范围，参考值范围。

医学参考值范围的估计方法有两种：正态分布法和百分位数法。其分析步骤分为：①根据研究指标的同质性要求，对研究指标分性别、年龄、地区等指标进行正态性检验；②在保证观察单位数量较多的前提下，选单、双侧界值；③对资料进行探索性分析，揭示分布特征及异常值；④根据专业要求选恰当的百分界值；⑤根据资料的分布类型、研究目的等，选择参考值范围的估计方法。资料服从正态分布，计算均数 $\bar{x}$ 与标准差 s ，按 $\bar{x} \pm u_{\alpha/2}s$ (双侧)， $\bar{x} + u_{\alpha}s$ 或 $\bar{x} - u_{\alpha}s$ (单侧) 估计参考值范围；当资料不满足正态分布要求或分布类型未知时，可用百分位数法估计参考值范围。

4. 正常成年人的白细胞计数近似正态分布，其 95% 参考值范围为 $4.0 \times 10^9/L \sim 10.0 \times 10^9/L$ 。若有一名成年男子测得其白细胞计数为 $3.8 \times 10^9/L$ ，可否判断该男子一定有病？

该判断不正确。医学参考值范围是指绝大多数“正常”人某指标的波动范围，是临床上判定正常与异常的参考标准，并不表示超出该范围的就一定是病人。该男子白细胞计数低于参考值范围的下限仅提示按大多数人该指标分布来讲，他可能有病，但也有可能是个体变异较大所致。

5. 何谓相对数？常用的有哪几种？增长速度与发展速度各有何用途？