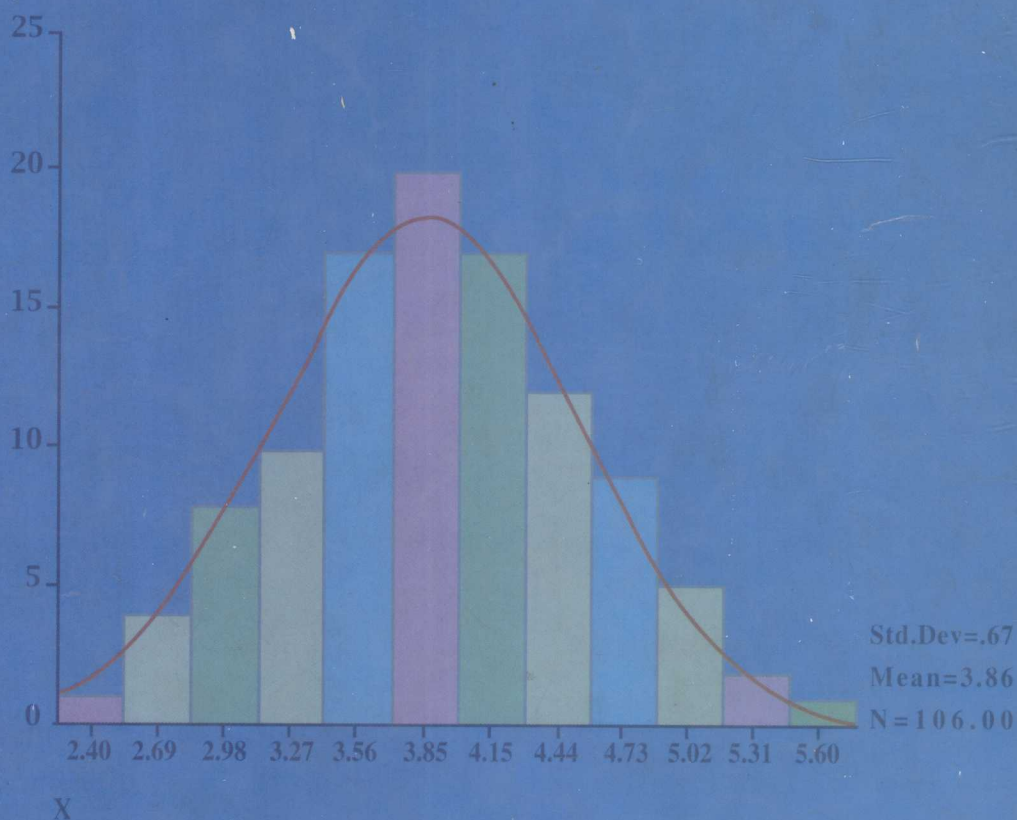


● 全国高等医药院校教材学习指导丛书

卫生统计学学习指导

余红梅 / 主编



中国协和医科大学出版社

第 2 版 全国高等医药院校教材配套用书

卫生统计学学习指导

第 2 版 主编



中国协和医科大学出版社

R 195.1-42

R 195.1-42

3

交教材学习指导丛书

卫生统计学学习指导

余红梅 主 编

张岩波 副主编

编委 (按姓氏笔画排序)

仇丽霞 王琳娜 张岩波 刘桂芬

余红梅 肖 琳 赵晋芳



096225

中国协和医科大学出版社

图书在版编目(CIP)数据

卫生统计学学习指导 / 余红梅编著. —北京: 中国协和医科大学出版社, 2005.6

ISBN 7-81072-688-9

I. 卫… II. 余… III. 卫生学: 统计学—医学院校—教学参考资料 IV. R195.1

中国版本图书馆 CIP 数据核字 (2005) 第 045467 号

全国高等医药院校教材学习指导丛书

卫生统计学学习指导

主 编: 余红梅

策划编辑: 张俊敏

责任编辑: 吴桂梅 赵玲秀

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)

经 销: 新华书店总店北京发行所

印 刷: 北京丽源印刷厂

开 本: 787×1092 毫米 1/16 开

印 张: 8.75

字 数: 260 千字

版 次: 2005 年 6 月第一版 2005 年 6 月第一次印刷

印 数: 1—3000

定 价: 15.00 元

ISBN 7-81072-688-9/R·681

(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

卫生统计学是一门应用性和实践性较强的学科。学习卫生统计学必须通过实习,亲自动手完成一定数量、不同形式的习题,才能更好地掌握卫生统计学的基本概念、原理和方法,培养正确的统计思维方法,以提高分析问题与解决问题的能力,保证卫生统计学学习质量。为此,我们采取分工编写、集体讨论定稿方式形成本书。值得一提的是,我校已故著名卫生统计学家何大卫教授于1997年主编出版了我国第一本《卫生统计学习题解答》(人民卫生出版社《卫生统计学》第三版配套教材),该书的出版使用赢得了国内同行和学生的诸多好评,本书的形成与此不无传承。

本书作为中国协和医科大学出版社2003年出版的《卫生统计学》配套教材,全书共分四个单元,每一单元包括目的要求、内容精要、练习题与参考答案以及补充题与参考答案。其中目的要求介绍了每一章按照教学大纲要求掌握、熟悉和了解的内容;内容精要以简明扼要的形式总结概括了每一章的主要内容,可作为读者实习或复习时的参考;练习题的题型包括判断题、单项选择题、计算分析题、SAS结果分析题、简答题和讨论题,书中给出了全部练习题的参考答案;本书有别于国内同类书籍的一个重要之处是补充了大量的、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。

使用本书时,建议初学者切勿死记硬背或不求甚解,一定要结合理论,联系实际,在认真思考、亲自解题或集体讨论的基础上再核对答案,这样才能深入理解卫生统计学的基本原理、概念和方法。

《卫生统计学》教材的编委对本书给予了大力的支持,山西医科大学卫生统计教研室的博士、硕士研究生参加了集体讨论,核对了部分答案,并提出了许多宝贵意见,在此一并致谢。

本书编者力求提高质量,但限于我们的学识,书中难免纰漏,恳请读者批评指正。

编 者

2005年4月6日

目 录

第一单元 定量资料统计描述与统计推断 (1~6 章)	t u F (1)
目的要求.....	(1)
内容精要.....	(1)
练习题与参考答案.....	(8)
补充题与参考答案.....	(31)
第二单元 分类资料统计描述与统计推断 (7~10 章)	χ^2 (36)
目的要求.....	(36)
内容精要.....	(36)
练习题与参考答案.....	(41)
补充题与参考答案.....	(59)
第三单元 回归与相关 (11~12 章)	 (66)
目的要求.....	(66)
内容精要.....	(66)
练习题与参考答案.....	(69)
补充题与参考答案.....	(80)
第四单元 卫生统计应用 (13~18 章)	 (85)
目的要求.....	(85)
内容精要.....	(85)
练习题与参考答案.....	(93)
补充题与参考答案.....	(122)

第一单元 定量资料统计描述与统计推断 (1~6章)

目的要求

1. 了解卫生统计学的意义和基本内容;熟悉统计工作的基本步骤;掌握卫生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;掌握描述定量资料常用指标的计算与适用范围。
3. 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。
4. 掌握抽样误差、标准误的基本概念及标准差与标准误的区别和联系;了解 t 分布与 u 分布的概念;掌握 t 界值表的使用,掌握 t 与 P 的关系;正确理解总体均数可信区间的意义,掌握其计算方法。
5. 正确理解假设检验的意义,掌握其基本步骤;掌握 t 检验应用条件和计算方法,并对结果作出合理解释;熟悉 I 型错误和 II 型错误的概念及关系;了解正态性判定、方差齐性检验的方法。
6. 掌握方差分析的基本思想;掌握方差分析用于多组均数比较时的整体分析思路;了解析因设计和重复测量设计资料的分析。

内容概要

第一章 绪 论

一、卫生统计学的意义和基本内容

卫生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、清理和分析的一门应用学科。基本内容包括医学研究设计(十四、十五章)、统计学基本原理与方法(一至十三章,十七章)及卫生统计应用(十六、十八章)。

二、统计工作的基本步骤

设计 → 收集资料 → 整理资料 → 分析资料

1. 医学研究设计 包括对医学研究资料的搜集、清理和分析全过程的总设想或安排。
2. 搜集资料 即按研究设计要求,获取准确、可靠、有意义的原始数据。
3. 清理资料 是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. 分析资料 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。

等级 (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) 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。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 $0 \leq P \leq 1$ 。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。对于某一次试验, 可认为该事件几乎不可能发生。

第二章 定量资料的统计描述

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、中心位置及离散特征。

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)。集中趋势反映变量分布的中心位置, 离散趋势反映变量值围绕中心位置的波动情况。

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)。频数分布以集中位置所在组段为中心, 左右两侧对称性逐渐减少, 称为对称分布。反之, 中心位置偏向一侧, 各组段的频数分布不对称, 称为偏态分布。若频数分布的高峰位置偏于左侧, 频数拖尾向右侧延伸, 称为正偏态分布; 若频数分布的高峰位置偏于右侧, 频数拖尾向左侧延伸, 称为负偏态分布。

二、定量资料的频数分布图

根据频数表资料, 以变量 X 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。它比频数表能更直观、更形象地描述频数分布的情况。

三、集中趋势指标与离散趋势指标

见表 1-1 和表 1-2

表 1-1 常用平均数指标的意义、应用场合及计算

指标	意义	应用场合	计算
均数 = $\bar{x}$ (mean, $\bar{x}$)	平均水平	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: $\bar{x} = \frac{\sum xf}{\sum f}$
几何均数 (geometric mean, G)	平均增(减)倍数	对数正态分布资料	直接法: $G = 1g^{-1} \left(\frac{\sum 1gx}{n} \right)$ 加权法: $G = 1g^{-1} \left(\frac{\sum f_1gx}{\sum f} \right)$
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: $M = L + \frac{i}{f_x} \left(\frac{n}{2} - \sum f_l \right)$

L 中位数所在组段下限 i 组距 中位数所在组 中位数前一组的累积频数

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	一组观测值中最大值与最小值之差

续 表

指 标	意 义	应用场合	计 算
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 varia- tion, 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	$CV = s/\bar{x} \times 100\%$

第三章 正 态 分 布

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, 它以均数为中心, 中间高, 两侧逐渐减少, 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%、95% 及 99%。

二、标准正态分布

对任意服从 $N(\mu, \sigma^2)$ 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, 根据正态分布的对称性, $\Phi(u) = 1 - \Phi(-u)$, 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, 见表 1-3。

表 1-3 医学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。

第四章 总体均数的估计

可信区间 $\bar{x} \pm 1.96S/\sqrt{n}$
波动范围是否正常 参考值范围 $\bar{x} \pm 1.96S$

一、均数的抽样误差与标准误

在抽样研究中, 由抽样引起的样本均数与总体均数间的差别称为均数的抽样误差 (sampling error)。常用标准误 (standard error) $\sigma_{\bar{x}}$ 来描述均数抽样误差的大小, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ 。当 σ 未知

时可用 s 来代替, 用 $s_{\bar{x}} = \frac{s}{\sqrt{n}}$ 作为 $\sigma_{\bar{x}}$ 的估计值。抽样研究中, 可适当增加样本含量, 控制和减小抽样误差。

二、 t 分布

在正态总体 $N(\mu, \sigma^2)$ 中作随机抽样, 样本均数的分布呈正态分布 $N(\mu, \sigma_{\bar{x}}^2)$; 对样本均数进行标准正态变换, $u = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$, 将 $N(\mu, \sigma_{\bar{x}}^2)$ 变换为标准正态分布 $N(0, 1)$, 即 u 分布。若用 $s_{\bar{x}}$ 作为 $\sigma_{\bar{x}}$ 的估计值, 其结果服从 t 分布。 t 分布与自由度 ν 有关, 当 ν 不同时, 曲线的形状不同。其特点为: ①以 0 为中心, 左右对称, 呈单峰分布; ② ν 越小, 偏离 0 较远

的 t 值相对较多, 此时 t 分布曲线的峰较低而尾部翘得较高; 随着 ν 增大, t 值的分布逐渐集中, 峰升高而尾部面积减小, 当 $\nu \rightarrow \infty$ 时, t 分布曲线逼近于标准正态分布曲线。

三、总体均数的估计

总体均数的估计方法有点值估计 (point estimation) 和区间估计 (interval estimation) 两种。点值估计把样本统计量直接作为总体参数的估计值, 这种方法虽然简单, 但未涉及抽样误差。区间估计是指按预先设定的可信度 (confidence level) $(1 - \alpha)$, 来确定总体均数的所在范围。它考虑了抽样误差的大小, 比点值估计更为常用。总体均数的 $(1 - \alpha)$ 可信区间 (confidence interval) 用 $\bar{x} \pm t_{\alpha/2, \nu} s_{\bar{x}}$ 估计, 如 $\alpha = 0.05$, 求得总体均数的 95% 可信区间, 其意义是指按该区间估计总体均数时可信度为 95%, 即用这种方法估计, 犯错误的可能性最大不应超过 5%。

第五章 假设检验原理和 t 检验

一、假设检验的基本思想与步骤

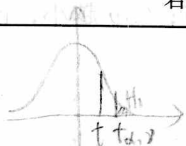
两组或多组指标比较时, 要检验比较组间的差别是由抽样误差引起的, 还是由于处理因素所产生的差别, 就需按小概率、反证法的思维进行假设检验。其步骤有: ①建立假设、确定检验水准 (level of a test) 及单双侧检验; ②根据资料的性质和研究目的, 选定检验方法, 计算检验统计量; ③确定 P 值, 作出推断结论。

二、 t 检验的应用条件、类型和方法

当样本例数 n 较小时, 要求样本取自正态总体; 作两样本均数比较时还要求两总体方差齐。本章主要有样本均数与已知总体均数的比较、配对设计两样本均数的比较和完全随机设计两样本均数的比较, 见表 1-4。

表 1-4 t 检验类型及步骤

假设检验步骤	样本均数与总体均数比较	配对设计两均数比较	完全随机设计两均数比较
① 建立假设, 确定检验水准	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$ (单侧时 $\mu > \mu_0$ 或 $\mu < \mu_0$) $\alpha = 0.05$	$H_0: \mu_d = 0$ $H_1: \mu_d \neq 0$ $\alpha = 0.05$	$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$ (单侧时 $\mu_1 > \mu_2$ 或 $\mu_1 < \mu_2$) $\alpha = 0.05$
② 计算检验统计量	$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$ $\nu = n - 1$	$t = \frac{\bar{d} - 0}{s_d/\sqrt{n}}$ $\nu = n - 1$	$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}}$ $\nu = n_1 + n_2 - 2$
③ 确定 P 值, 做出推断结论	查 t 界值表, 若 $P > \alpha$, 按 $\alpha = 0.05$ 水准, 不拒绝 H_0 ; 若 $P \leq \alpha$, 按 $\alpha = 0.05$ 水准, 拒绝 H_0 , 接受 H_1 。		



$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2(n_1-1) + S_2^2(n_2-1)}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

三、两类错误

假设检验结论是具有概率性的，下结论时一定要慎重。拒绝实际上成立的 H_0 所犯的误差称 I 型错误 (type I error) ^{弃真} 亦称第一类错误，用 α 表示。与之相应，推断正确的可能性为 $1 - \alpha$ ， $1 - \alpha$ 又称为可信度。不拒绝实际上不成立的 H_0 所犯的误差称 II 型错误 (type II error) ^{存伪}，亦称第二类错误，用 β 表示。 $1 - \beta$ 又称为检验效能 (power of a test)，其意义为当两总体确有差别，按规定的假设检验水准 α 能发现该差别的能力。有关 α 、 β 的取值，必须按具体问题事先做出选择。样本含量一定时，减小 α 会引起 β 增大，减小 β 会引起 α 增大。若要同时减少 α 和 β ，只有增加样本含量。

四、假设检验应注意的问题

基于严密的研究设计；注意假设检验方法的条件；正确表述统计推断结论；慎重解释“阴性”结果；充分利用可信区间与假设检验所提供的信息。

第六章 方差分析

一、完全随机设计方差分析的基本思想

$$SS_{\text{总}} = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2 \begin{cases} SS_{\text{组间}} = \sum_{i=1}^k n_i (\bar{x}_i - \bar{x})^2 & \nu_{\text{组间}} = k - 1 \\ SS_{\text{组内}} = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 & \nu_{\text{组内}} = n - k \end{cases}$$

统计量 $F = \frac{MS_{\text{组间}}}{MS_{\text{组内}}} = \frac{SS_{\text{组间}}/\nu_{\text{组间}}}{SS_{\text{组内}}/\nu_{\text{组内}}}$ ，若处理无作用时，组间变异与组内变异均由随机误差引起，故 F 值理论上等于 1；若处理有作用，则 F 值明显大于 1。当 $F \geq F_{\alpha, (\nu_{\text{组间}}, \nu_{\text{组内}})}$ 时， $P \leq \alpha$ ，可认为处理因素效应间差别有统计学意义；当 $F < F_{\alpha, (\nu_{\text{组间}}, \nu_{\text{组内}})}$ 时，则 $P > \alpha$ ，尚不能认为处理因素效应间差别有统计学意义。

二、随机区组设计方差分析的基本思想

$$SS_{\text{总}} = \sum_{i=1}^k \sum_{j=1}^b (x_{ij} - \bar{x})^2 \begin{cases} SS_{\text{处理}} = \sum_{i=1}^k b (\bar{x}_i - \bar{x})^2 & \nu_{\text{处理}} = k - 1 \\ SS_{\text{区组}} = \sum_{j=1}^b k (\bar{x}_j - \bar{x})^2 & \nu_{\text{区组}} = b - 1 \\ SS_{\text{误差}} = SS_{\text{总}} - SS_{\text{处理}} - SS_{\text{区组}} & \nu_{\text{误差}} = n - k - b + 1 \\ & = (k-1)(b-1) \end{cases}$$

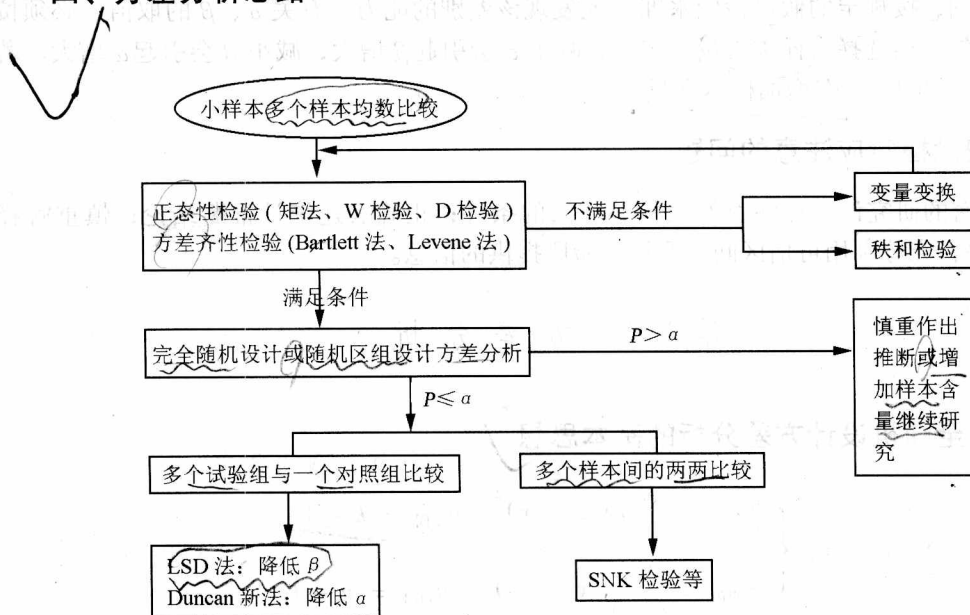
$$F_{\text{处理}} = \frac{MS_{\text{处理}}}{MS_{\text{误差}}} = \frac{SS_{\text{处理}}/\nu_{\text{处理}}}{SS_{\text{误差}}/\nu_{\text{误差}}}$$

$$F_{\text{区组}} = \frac{MS_{\text{区组}}}{MS_{\text{误差}}} = \frac{SS_{\text{区组}}/\nu_{\text{区组}}}{SS_{\text{误差}}/\nu_{\text{误差}}}$$

三、方差分析的应用条件

1. 各对比组资料服从均数为 μ_i 的正态分布。
2. 各总体方差相等，即方差齐。
3. 各样本相互独立。

四、方差分析思路



练习题与参考答案

一、判断题

1. 四组均数比较，若 $P < P_{0.05(3,28)}$ ，则可认为各总体均数间差别均有统计学意义。
2. 均数的抽样研究中，抽样例数越少，均数的标准误就越小。
3. 在均数的假设检验中，适当增加观察单位数，不仅可减小抽样误差，且可减小 I 型错误和 II 型错误。
4. 一组有变异的数据，无论它是否服从正态分布，其几何均数总小于均数。
5. 两均数比较的 t 检验中， P 值越小，说明两总体均数间相差越大。
6. 在多个均数的比较中，若不采用方差分析，而用 t 检验对每两个均数作比较，会加大 I 型错误的概率。
7. 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。
8. 用百分位数法确定双侧 95% 正常值范围的下限值 P_5 ，上限值是 P_{95} 。

参考答案

1. 错误 2. 错误 3. 正确 4. 正确 5. 错误 6. 正确 7. 错误 8. 错误

二、单项选择题

1. 频数分布的两个重要特征是 B。

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

2. 表示一组对称分布资料变量值的平均水平, 宜选用 A。

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

3. 下列有关中位数 (M) 描述不正确的有 A。

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. $M = P_{50}$

4. 医学研究中, 反映一组血清抗体效价 (滴度) 资料的平均水平, 常选用 C。

A. $\bar{x}$
B. M
C. G (几何均数)
D. R
E. CV

5. 下列有关四分位数间距描述中不正确的是 D。

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

6. 健康男子收缩压的正常值范围一般指 B。

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

7. 正态分布曲线下, 横轴上从均数 μ 到 $\mu + 1.645\sigma$ 的面积为_____。

- A. 95%
- ☒ B. 45%
- C. 90%
- D. 不能确定
- E. 1

8. 若随机变量 X 服从正态分布 (μ, σ^2) , 则 X 的第 95 百分位数等于_____。

- A. $\mu - 1.645\sigma$
- ☒ B. $\mu + 1.645\sigma$
- ☒ C. $\mu + 1.96\sigma$
- D. $\mu + 2.58\sigma$
- E. $\mu - 1.96\sigma$

9. 抽样研究中, 标准误越大, 表示_____越大。

- A. 系统误差
- B. 样本含量
- C. I 型错误
- D. 把握度
- ☒ E. 抽样误差

10. 抽样研究中, 适当增加观察单位数_____。

- A. 可减小 I 型错误
- ☒ B. 可减小 II 型错误
- C. 可减小抽样误差
- D. 可提高检验效能
- E. 以上均正确

11. 三组以上某实验室指标观测数据服从正态分布且满足参数检验的应用条件。任两组分别进行多次 t 检验代替方差分析, 将会_____。

- A. 明显增大犯 I 型错误的概率
- B. 使结论更加具体
- C. 明显增大犯 II 型错误的概率
- D. 使均数相差更为显著
- E. 使均数的代表性更好

12. 在完全随机设计的方差分析中, 必然有_____。

- A. $SS_{\text{组内}} < SS_{\text{组间}}$
- B. $MS_{\text{组间}} < MS_{\text{组内}}$
- C. $MS_{\text{总}} = MS_{\text{组间}} + MS_{\text{组内}}$

$$D. SS_{\text{总}} = SS_{\text{组间}} + SS_{\text{组内}}$$

$$E. MS_{\text{组间}} > MS_{\text{组内}}$$

13. 完全随机设计方差分析中, 组间均方主要反映 C。

- A. 抽样误差大小
- B. n 个数据的离散程度
- C. 处理因素的作用
- D. 随机误差的影响
- E. 系统误差的影响

参考答案

1. B 2. A 3. A 4. C 5. D 6. B 7. B 8. B 9. E 10. E 11. A 12. D
13. C

三、计算分析题

1. 某研究组测得 120 名女大学生的收缩压 (mmHg) 资料如下:

106	102	102	88✓	94	104	120	98	98	100	112	102	106	100	108
102	100	108	98	128	102	106	102	94	106	100	88✓	106	100	98
98	94	112	106	96	104	118	100	98	102	102	102	98	106	86✓
96	100	98	100	100	100	106	100	86✓	100	102	104	102	100	96
86✓	102	102	90	94	100	106	110	96	104	100	118	106	106	94
84	96	92	94	80	104	84	94	90	88✓	102	98	96	94	100
86✓	114	96	108	108	110	110	94	110	110	100	110	94	106	118
96	96	118	104	110	118	98	120	118	110	100	90	112	90	118

①试编制频数表并描述其分布特征; ②计算均数、标准差和变异系数; ③求其中位数和四分位数间距; ④估计该地健康女大学生收缩压 95% 参考值范围。

(1) 编制频数表, 描述其分布特征, 见表 1-5。

1) 求极差: $R = 128 - 80 = 48$ (mmHg)

2) 求组距: $i = R/10 = 48/10 = 4.8 \approx 5$ (mmHg)

3) 写出组段, 将表中数据逐一划记到各组段, 统计频数, 求组中值、频数、累计频数、频率、累计频率, 完成频数表。

由频数表可以看出, 该资料大部分观察值集中在频数表的中间, 即 95~110 之间, 两侧频数逐渐减少, 左右基本对称, 近似正态分布。

(2) 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:

$$\bar{x} = 102.125 \text{ (mmHg)}$$

$$s = 8.749 \text{ (mmHg)}$$

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