



ALTH
RE

MEDICAL

安徽省高等学校“十二五”省级规划教材

医药数理统计技能实训

(案 例 版)

王万荣 高美华 · 主编



东南大学出版社
SOUTHEAST UNIVERSITY PRESS

安徽省高等学校“十二五”省级规划教材

高职高专药学类专业实训教材

医药数理统计技能 实训(案例版)



主 编 王万荣 高美华
副主编 李济平 吴道华

编 者(以姓氏笔画为序)

王万荣(安徽医学高等专科学校)

朱海燕(安徽医学高等专科学校)

吴道华(安徽中医药高等专科学校)

李济平(安庆医药高等专科学校)

杨冬梅(安徽医学高等专科学校)

周建庆(安徽医学高等专科学校)

姚 霖(安徽省第二人民医院)

高美华(合肥立方制药股份有限公司)

SE 东南大学出版社
SOUTHEAST UNIVERSITY PRESS

· 南京 ·

图书在版编目(CIP)数据

医药数理统计技能实训(案例版) / 王万荣, 高美华
主编. —南京: 东南大学出版社, 2016. 10
ISBN 978-7-5641-6765-3

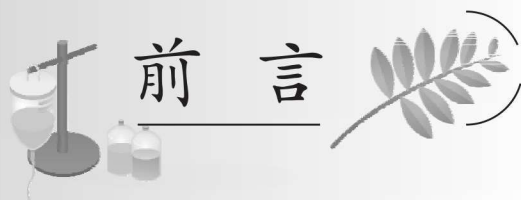
I. ①医… II. ①王… ②高… III. ①医用数学—数理统计—高等职业教育—教材 IV. ①R311

中国版本图书馆 CIP 数据核字(2016)第 231774 号

医药数理统计技能实训(案例版)

出版发行 东南大学出版社
出版人 江建中
社址 南京市四牌楼 2 号
邮编 210096
经销 江苏省新华书店
印刷
开本 787 mm×1 092 mm 1/16
印张 10
字数 250 千字
版次 2016 年 10 月第 1 版第 1 次印刷
书号 ISBN 978-7-5641-6765-3
定价 28.00 元

* 本社图书若有印装质量问题, 请直接与营销部联系, 电话: 025—83791830。



前言

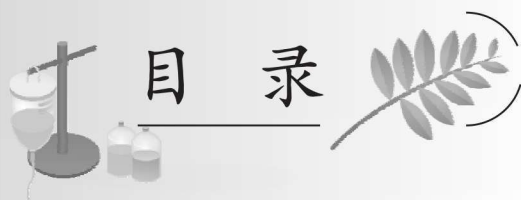
《医药数理统计技能实训》(案例版)是根据教育部制定的最新高等职业院校药学专业标准(2012版)配合专业平台基础课程《医药数理统计》而编写的一门实训课程,其培养目标是:严谨的治学态度、精细的治学作风、独立的自学能力、基本的创新意识。本教材不但适用于高职高专院校药学专业实训教学,而且可供临床医学、护理、医学相关类专业学习医学统计类课程时选用。

本实训教材以概率论与数理统计理论为基础,突出其在医药领域实际应用,以数据收集、处理、分析与推断作为主线,从计量资料的统计描述、计数资料的统计描述、参数估计、统计图表的制作、均数的假设检验、比率的假设检验、等级资料的比较、线性相关与回归分析、制药生产条件优化设计、医药科研论文撰写指导、医药文献资料评阅与常见错误识别等十二个方面,编写了医药领域具有代表性的与数理统计密切相关且综合性比较强的典型工作任务,作为技能实训案例及实训素材。教学中可结合实际灵活选用。本书也适合参加自考助学者学习《医药数理统计》配套使用。

本实训教材具有以下三个特点:一是突出“校企合作,工学结合”,编写团队来自高校具有丰富教学经验的高资历教师和来自医院、药企具有一线药品生产研发应用实践经验的专家组成,项目内容经过认真地集体筛选与认定;二是理实一体化,注重实际应用,所选择的项目尽可能贴近药学职业岗位研发、生产、应用实际,是从事药学相关服务工作必须掌握的、基本的职业素养;三是引入现代软件 Excel 与 SPSS,简化数理统计中大量繁琐的数据处理过程,发挥统计工具的高效作用,突出掌握数理统计的思想、应用条件和创新性应用。通过理论联系实际的实训,本书将使学习者在医药卫生领域“认识问题更本质,数据分析更准确,实验设计更合理,调研思路更开阔,论文发表更高档”。

本教材编写分工分别为:实训项目一、三、六、十一、十二由王万荣负责,实训项目四由姚霖负责,实训项目二由李济平负责,实训项目五由吴道华负责,实训项目七由杨冬梅负责,实训项目八由朱海燕负责,实训项目九由周建庆负责,实训项目十由高美华负责。教材编写过程中得到各参编院校及东南大学出版社、安徽省示范性高等职业院校合作委员会医药卫生类专业协作组的大力支持,并广泛参考了相关教材、著作、论著与网络资源,在此一并表示感谢。由于同类应用性实训教材鲜见,加之时间原因和编者水平,教材编写中可能还有一些疏漏之处,敬请广大师生提出宝贵意见。

王万荣
2016年6月



实训项目一	医药计量资料的统计描述	(1)
实训项目二	医药参考值范围与均数区间估计	(16)
实训项目三	医药计量资料的假设检验	(28)
实训项目四	医药计数资料的统计描述	(56)
实训项目五	医药计数资料的假设检验	(66)
实训项目六	医药等级资料数据处理	(77)
实训项目七	药品检测线性分析	(87)
实训项目八	医药统计图表的制作	(102)
实训项目九	药品质量稳定性分析	(115)
实训项目十	制药生产条件优化设计	(122)
实训项目十一	医学生毕业设计(论文)撰写指导	(137)
实训项目十二	医药文献常见统计学错误识别	(144)
附录		(153)
主要参考文献		(154)



实训项目一 医药计量资料的统计描述



实训目标

1. 能按照统计工作步骤开展医药数理统计工作。
2. 能正确制作频数表与直方图。
3. 能熟练选用有关指标对医药计量资料进行统计描述。
4. 能够应用计算器完成统计基本处理。
5. 学会利用统计软件完成基本统计分析。



实训内容

1. 统计工作基本步骤。
2. 频数表制作。
3. 集中趋势与离散趋势指标计算。
4. 误差识别与异常值判断。
5. 统计计算器与常用统计软件的使用。



实训指导

一、实训基础

(一) 误差分类

医药数理统计工作中的误差主要有三种类型：

(1) 系统误差(又称偏倚)：由于仪器不准确、标准试剂未经校准、评判标准偏高或偏低以及心理因素等原因，可使观察结果呈倾向性的偏大或偏小而造成的误差。可通过标准化、扣除空



白值、培训、盲法等方法予以校正。

(2) 随机误差:由于目前尚不能被人所控制的许多作用微小的因素(如气压或气温变化、肌肉抖动、电流波动)而造成测量结果的差异。随机误差虽不可避免、无法消除也无法校正,但服从正态分布规律。人们常用标准方法通过多次重复测定,所求出的算术平均值作为真实值。

(3) 抽样误差:在抽样研究中,即使消除了系统误差和过失误差,样本统计量与总体参数之间以及各样本统计量之间仍会存在差异,这种差异称为抽样误差。就其本质而言,抽样误差属于随机误差。

(二) 反映数据分布特征的指标

集中趋势即平均数(average)是描述一组观察值集中位置或平均水平的统计指标。它常作为一组数据的代表值用于分析和进行组间的比较。常用平均数包括算术平均数、几何平均数和中位数。

离散趋势指标又称变异程度指标,它反映各观察值之间参差不齐的程度。频数分布有集中趋势和离散趋势两个重要特征,只有把二者结合起来才能对事物有全面的认识。离散趋势指标有:极差、四分位数间距、方差、标准差和变异系数(又称相对标准误差 RSD),其中最常用的是方差和标准差。

(三) 计量资料异常值取舍

在一组平行测定的实验数据中,有时会出现个别过大或过小、远离平均数的测量值,这类数值被称为异常值、极端值、可疑值或逸出值。异常值有两种可能:一种可能是测量值随机波动的极度表现即极值,属于总体的局内值;另一种可能是与其余数据不属于同一总体的异常值,是局外值。局外的极端值应当舍弃,局内的极端值不应当舍弃。对于一组数据中的异常值,不能为了获得精密度高的分析结果,而随意地舍弃。应按统计学方法进行处理,决定其取舍。判断计量资料异常值是否属于局外值,常用 $\bar{x} \pm 3s$ 法、Q 检验法、G 检验法(格拉布斯法)和间距法。

(1) $\bar{x} \pm 3s$ 法

$\bar{x} \pm 3s$ 法适用于正态分布资料,且样本含量较大($n \geq 60$)。以 x_j 代表极端值,按正态分布理论, $(x_i - \bar{x})/s$ 的绝对值大于 2 的概率为 1/20,大于 3 的概率仅约为 1/370。按小概率原理,小概率事件在一次测量中实际上是不可能发生的,2 与 3 可认为统计上允许的合理误差范围,而超出此范围的数据则为极端值。因此,可以根据 $\bar{x} - 3s \sim \bar{x} + 3s$ 范围内是否包括 x_i 作出判断:当 x_i 在 $\bar{x} - 3s \sim \bar{x} + 3s$ 范围之外时可舍弃,当 x_i 在此范围之内时保留。

(2) Q 检验法

Q 检验法不要求正态分布资料。数据从小到大排列为 $x_1, x_2, x_3, x_4, \dots, x_n$ 极差 $R = x_n - x_1$,最小值 x_1 或最大值 x_n 为极端值时,计算统计量 Q 的公式分别为

$$Q = \frac{x_2 - x_1}{R} \text{ 或 } Q = \frac{x_n - x_{n-1}}{R}$$

若 $Q > 1/3$ 或查 Q 临界值表得 $Q_{表}$,则异常值是局外值,应舍弃。 $Q_{表}$ 大小取决于测定次数和置信度。



(3) G 检验法

格拉布斯法(Grubbs)适用于正态分布资料。以 x_i 代表极端值,计算包括极端值 x_i 在内的测量值 $\bar{x}$ 与 s ,总体均数 μ 及标准差 σ 已知或未知时计算统计量 T 的绝对值公式分别为

$$|T| = \frac{|x_i - \bar{x}|}{s} \text{ 或 } |T| = \frac{|x_i - \mu|}{\sigma}$$

根据第一类错误概率 α ,样本含量 n 值,查如表 1-1 所示的格拉布斯 $T_{\alpha,n}$ 界值表,与 T 的绝对值比较。若 $|T| \leq$ 界值 $T_{\alpha,n}$,则不能判极端值 x_i 为局外值;若 $|T| > T_{\alpha,n}$,则可判 x_i 为局外值,应舍去。

表 1-1 检验正态分布资料极端值用格拉布斯 $T_{\alpha,n}$ 界值表

μ, σ 未知时						μ, σ 已知时					
n	$\alpha=0.05$	$\alpha=0.01$	n	$\alpha=0.05$	$\alpha=0.01$	n	$\alpha=0.05$	$\alpha=0.01$	n	$\alpha=0.05$	$\alpha=0.01$
10	2.18	2.41	20	2.56	2.54	10	2.57	3.06	20	2.80	3.29
12	2.29	2.55	30	2.55	2.87	12	2.63	3.14	30	2.93	3.40
14	2.37	2.66	40	2.87	2.96	14	2.68	3.19	40	3.02	3.48
15	2.41	2.41	50	2.96	3.03	15	2.51	3.21	50	3.08	3.54
16	2.44	2.44	60	3.03	3.21	16	2.53	3.23	60	3.14	3.59
18	2.50	2.50				18	2.57	3.26			

(三) 频数分布及其规律

频数分布就是变量在其取值范围内各组段分布情况。通过实验或临床观察等各种方式得到的原始资料,如果是计量资料并且观察的例数较多,可以对数据进行分组,然后制作频数分布表(简称频数表 frequency table)或绘制直方图,用以显示数据的分布规律。

实际工作中,经常要了解正态曲线下横轴上某一区间的面积占总面积的百分数,或估计该区间的例数占总例数的百分数(频数分布)。正态曲线下一定区间的面积可以通过查标准正态分布曲线下左侧尾部面积表求得。当 μ, σ 未知且样本含量 n 足够大时,可用样本均数 $\bar{x}$ 和标准差 s 分别代替 μ 和 σ ,按 $u = \frac{(x - \bar{x})}{s}$ 式求得 u 值,再查表。应用 Excel 软件,快速获得正态分布曲线下左侧累积面积 P 和分界点 x 值:

① 区间 $(-\infty, x)$ 的面积 $P = \text{NORMDIST}(x, \text{均数}, \text{标准差}, \text{TRUE})$ 。

② 区间 $(-\infty, x)$ 的面积为 P 时,分界点 $x = \text{NORMINV}(P, \text{均数}, \text{标准差})$ 。

对于正态分布或近似正态分布的资料,已知均数和标准差,就可以估计其频数分布。

(四) 基层常用统计软件

基层工作中常用的数据处理软件,有 Excel、EpiData、EPI Calc、SPSS、Stata 等。针对不同类型数据处理要求,可分别选用相应的软件建立数据库并进行相应格式转换或导出。其中,



SPSS 软件和 Excel 软件尤其重要。

SPSS 由美国 SPSS(Statistical Package for the Social Science)公司研制,采用窗口菜单方式选择各种统计方法,图形界面良好,能方便地与 Excel 等数据库软件交换数据。该软件的特点是:统计方法较齐全,运行速度快,容量大;有较强的数据管理能力;运算结果可直接编辑后输出。该软件的主要功能包括:数据的探索性分析、描述性统计、 t 检验、方差分析、非参数统计、协方差分析、列联表分析、相关分析、多元回归分析和图表制作等。该软件能够满足常规医学数据处理和分析的一般需要。

二、实训用物

CASIO 82-TL 计算器或同系列计算器;计算机;SPSS 及 Excel 软件。

三、实训要点

(一) 统计计算器的使用方法

根据所使用的计算器的说明书进行教学。

CASIO 82-TL 统计功能的使用

基本统计操作

1. 首先按 **MODE 2** 进入 SD 模式。统计前要先清除上次内存中已有的统计记录: **SHIFT AC=**。

2. 接着输入统计数据(一个一个输,每个数据输完后按 **M+** 录入);多个相同数据可用 **;** 来输入(例如:输入 10 个 2,可按 **2;10 M+**, **;** 的输入是 **SHIFT,**);发现输入错误立即删除的方法是按 **SHIFT AC**。

3. 输完数据后,求统计结果:

SHIFT 1= 是平均数;

SHIFT 2= 是总体标准差;

SHIFT 3= 是样本标准差;

RCL(-) 是 $\sum x^2$;

RCL.,.,. 是 $\sum x$;

RCL hyp 是 n (输入的数据的个数)。



线性回归计算和相关系数 r

1. 首先进入线性回归状态: $\boxed{\text{MODE } 3}$ (REG 模式)1(函数类型 $Y=A+BX$)

2. 然后输入对应的(两组以上)X 和 Y:

X_1, Y_1 M+

X_2, Y_2 M+

...

X_n, Y_n M+

3. 输出结果:

$\boxed{\text{SHIFT } 7=}$ 和 $\boxed{\text{SHIFT } 8=}$ 可分别求出关系式中的 A(截距)、B(斜率)。

$\boxed{\text{SHIFT}(=)}$ 相关系数 r 。

REG 模式里同样可以统计

输入 X, Y M+..., 然后 $\text{SHIFT}+1\sim 3$ 分别是 X 的平均数、标准差、方差; $\text{SHIFT}+4\sim 6$ 分别是 Y 的平均数、方差、标准差。另外:

$\boxed{\text{RCL}(-)}$ 是 $\sum x^2$;

$\boxed{\text{RCL}_{\circ,.,.}}$ 是 $\sum x$;

$\boxed{\text{RCL hyp}}$ 是样本数;

$\boxed{\text{RCL sin}}$ 是 $\sum y^2$;

$\boxed{\text{RCL cos}}$ 是 $\sum y$;

$\boxed{\text{RCL tan}}$ 是 $\sum xy$ 。

(二) 实训操作案例

【案例 1-1】 测某种药片 50 片, 其重量(mg)如表 1-2:

表 1-2 50 片药片重量(mg)

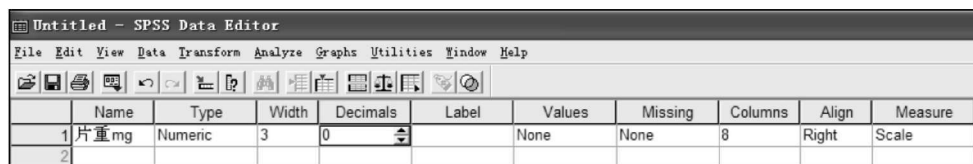
94	96	100	95	97	92	98	87	95	96
98	92	101	93	94	97	96	94	100	94
100	99	97	95	92	93	95	97	99	90
89	92	92	91	96	96	94	97	101	87
102	96	101	99	95	97	98	86	100	97

要求: 利用 SPSS 软件求该药片重量的均数、中位数、全距、标准差、标准误, 并绘制频数表、直方图。



①建立数据文件

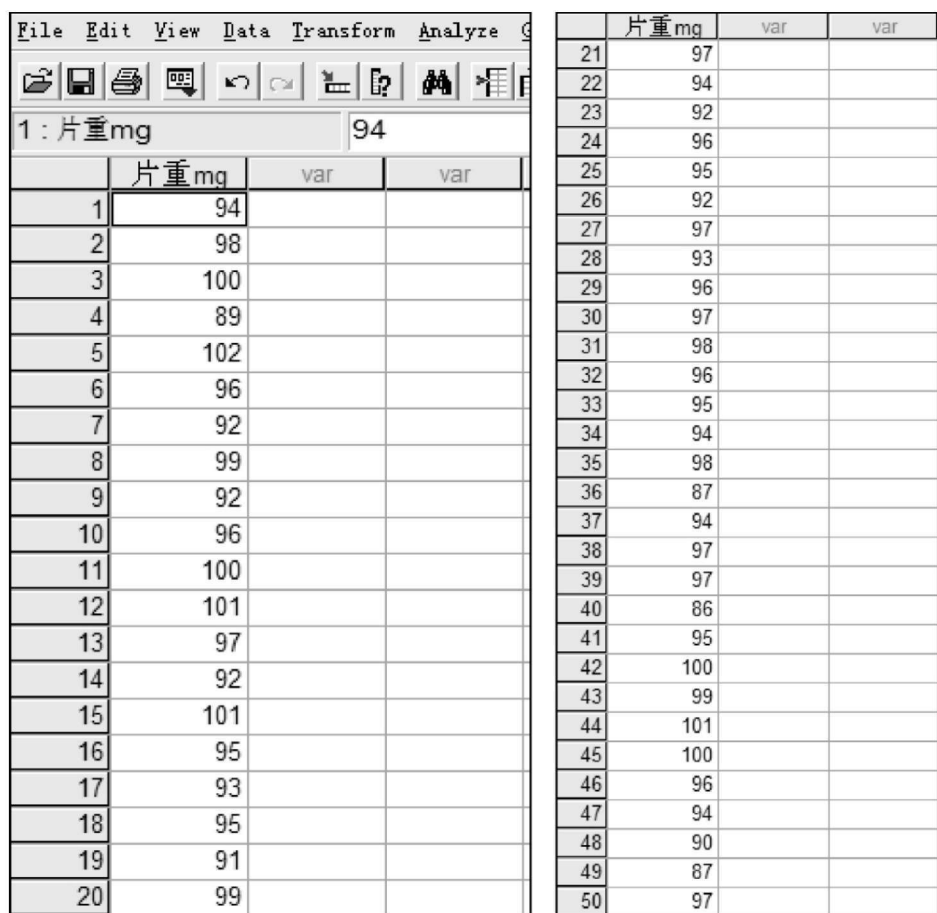
本例中变量设置如图 1-1 所示：



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	片重mg	Numeric	3	0		None	None	8	Right	Scale
2										

图 1-1 SPSS 变量视图

建立数据文件,依次调用菜单 File→Save,以文件名. sav 保存数据文件。



	片重mg	var	var
1	94		
2	98		
3	100		
4	89		
5	102		
6	96		
7	92		
8	99		
9	92		
10	96		
11	100		
12	101		
13	97		
14	92		
15	101		
16	95		
17	93		
18	95		
19	91		
20	99		

	片重mg	var	var
21	97		
22	94		
23	92		
24	96		
25	95		
26	92		
27	97		
28	93		
29	96		
30	97		
31	98		
32	96		
33	95		
34	94		
35	98		
36	87		
37	94		
38	97		
39	97		
40	86		
41	95		
42	100		
43	99		
44	101		
45	100		
46	96		
47	94		
48	90		
49	87		
50	97		

图 1-2 SPSS 变量视图

②依次调用菜单 Analyze→Descriptive Statistics→Frequencies 打开频数对话框。将左边的源变量“片重 mg”调入右边中部的“Variable”下的矩形框内,并选中 Display frequency tables 选项。

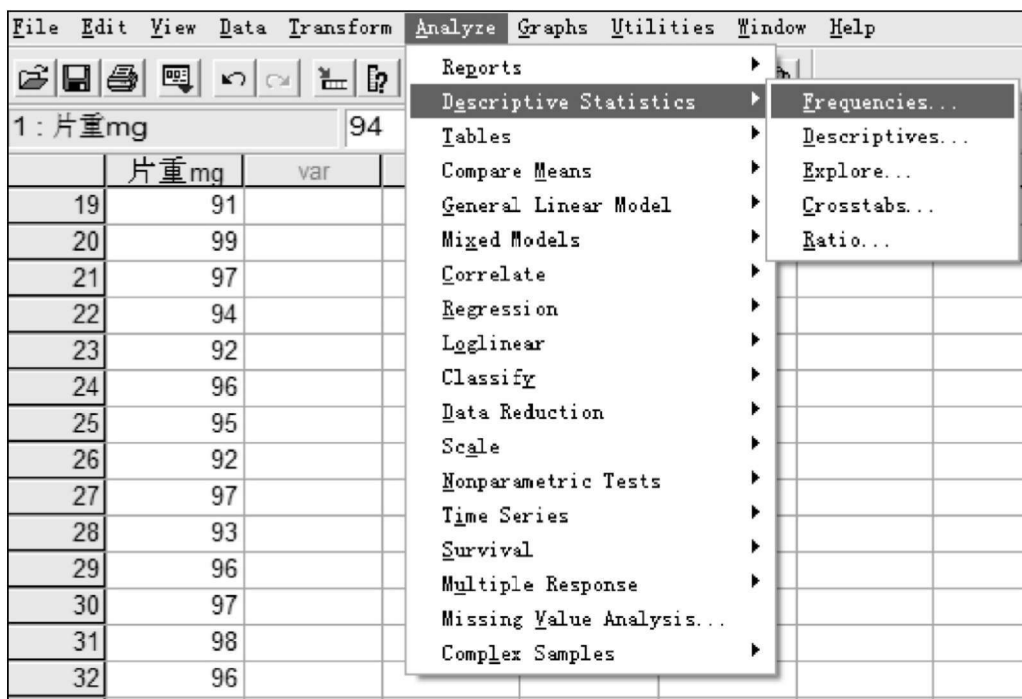


图 1-3 调用命令 Frequencies

③单击 Statistics 按钮,选中 Percentile 选项,向其后的空中填入 95,单击 Add 按钮确认;点选对话框中的平均数(Mean)、中位数(Median)、方差(Variance)、标准差(Std. deviation)、全距(Range)、最大值(Maximum)、最小值(Minimum)和均值的标准误(S. E. mean);点选分布中的偏度(Skewness)和峰度(Kurtosis),按 Continue 按钮确认(图 1-4)。

④单击 Charts 按钮,打开 Frequencies:Charts 对话框,选择条图(Histograms),并选择带正态曲线(With normal curve)选项。

⑤单击 Format 按钮,使用其默认值。

⑥最后单击“OK”按钮,将结果保存,并进行简要分析。

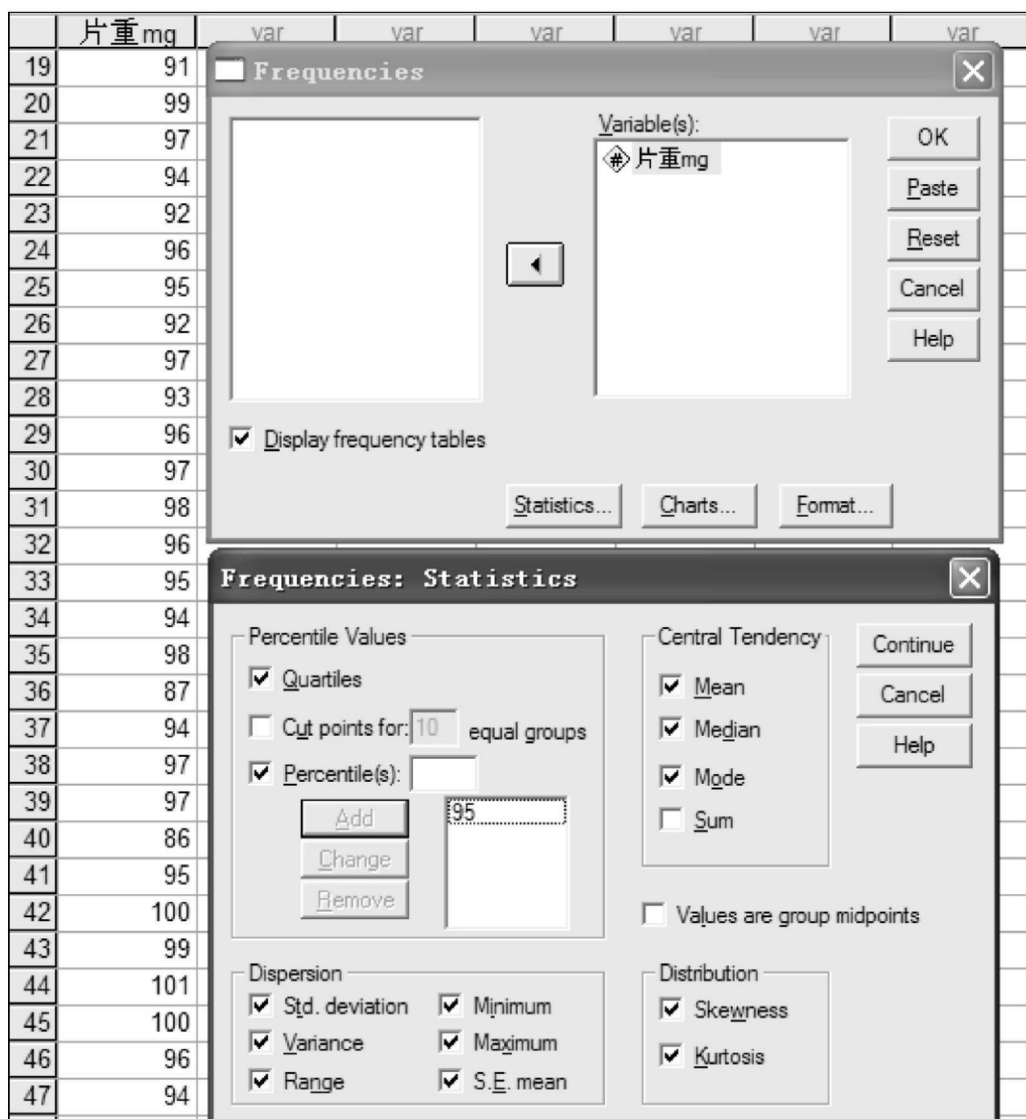


图 1-4 调用命令 Statistics 并设置

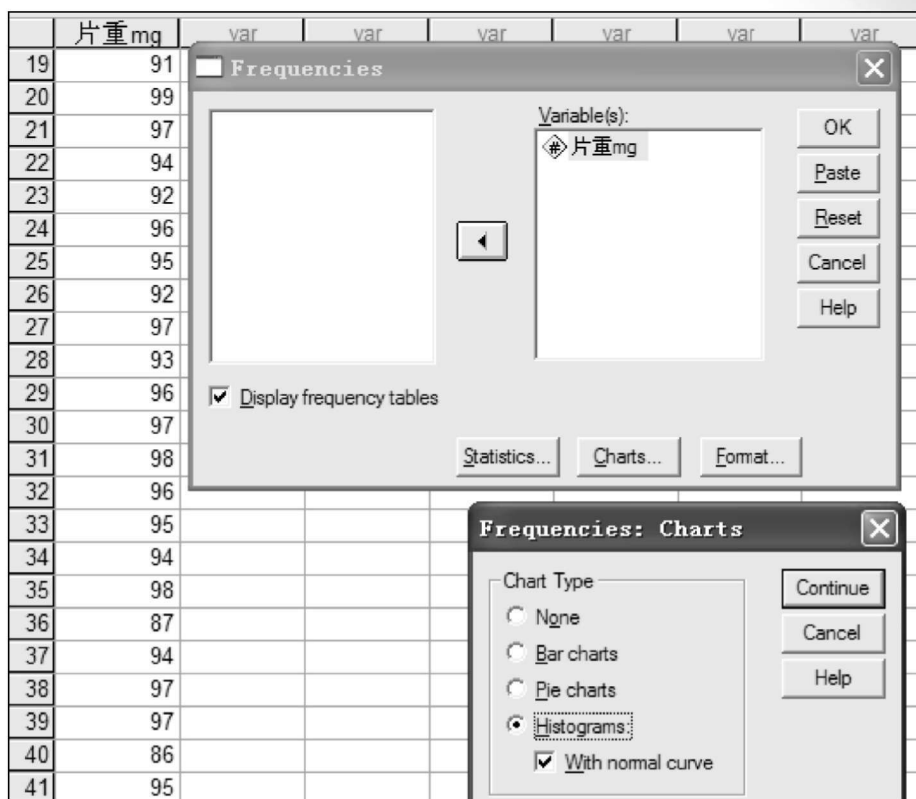


图 1-5 直方图制作设置

片重 mg		
N	Valid	50
	Missing	0
Mean		95.44
Std. Error of Mean		.535
Median		96.00
Mode		97
Std. Deviation		3.781
Variance		14.292
Skewness		-.557
Std. Error of Skewness		.337
Kurtosis		.078
Std. Error of Kurtosis		.662
Range		16
Minimum		86
Maximum		102
Percentiles	25	93.00
	50	96.00
	75	98.00
	95	101.00

图 1-6 片重选择性统计结果显示



片重 mg

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	86	1	2.0	2.0	2.0
	87	2	4.0	4.0	6.0
	89	1	2.0	2.0	8.0
	90	1	2.0	2.0	10.0
	91	1	2.0	2.0	12.0
	92	5	10.0	10.0	22.0
	93	2	4.0	4.0	26.0
	94	5	10.0	10.0	36.0
	95	5	10.0	10.0	46.0
	96	6	12.0	12.0	58.0
	97	7	14.0	14.0	72.0
	98	3	6.0	6.0	78.0
	99	3	6.0	6.0	84.0
	100	4	8.0	8.0	92.0
	101	3	6.0	6.0	98.0
	102	1	2.0	2.0	100.0
Total		50	100.0	100.0	

图 1-7 片重数据频数构成

偏度(Skewness)是描述变量取值分布形态对称性的统计量,其值为 0,说明数据为对称分布;大于 0 表示变量取值右偏,在直方图中有一条长尾拖在右边;小于 0,表示变量取值左偏,在直方图中有一条长尾拖在左边。峰度(Kurtosis)是用来描述变量取值分布形态陡缓程度的统计量,当数据分布和标准正态分布的陡缓程度相同时,该值为 0;大于 0,说明曲线更陡峭,为尖峰分布;小于 0 为平峰分布。

【案例 1-2】采用 Excel 软件对案例 1-1 数据进行集中趋势与离散趋势描述。

首先建立数据库,数据处于一列 A1:A100;然后按照下图操作:工具→加载宏→分析工具→描述统计→输出数据至 J1 列。

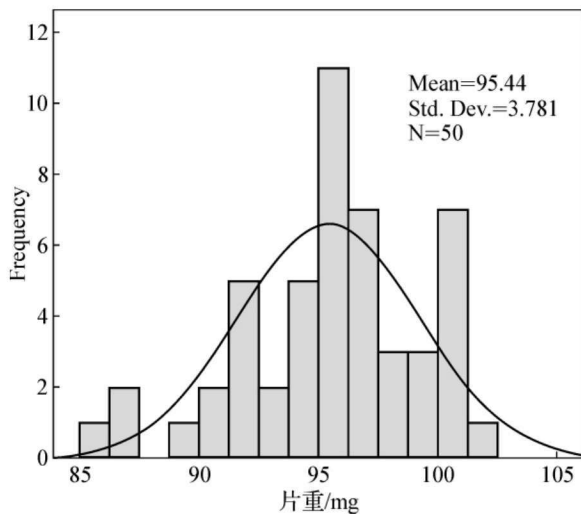


图 1-8 片重直方图



图 1-9 建立数据库并加载分析工具库

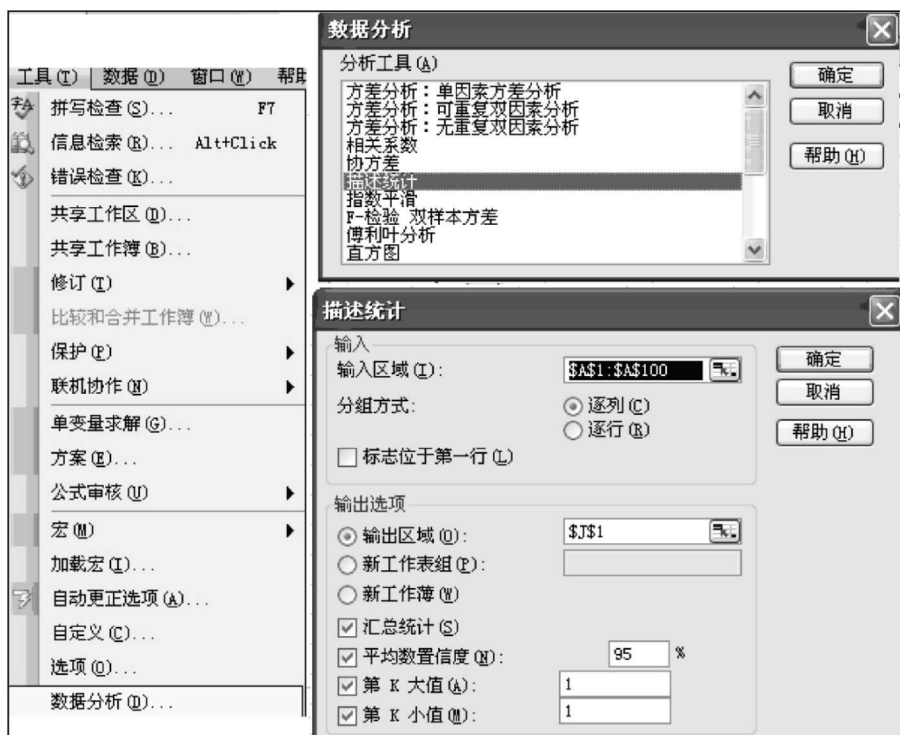


图 1-10 调出描述分析选择性输出统计结果

