

## · 药学进修教育 ·

## 药学人员统计学入门(上)

Eucharia Nnadi Okolo (美国佛罗里达A及M大学药学院副教授)

陈·新译 张紫洞校

统计学一词在许多不同的领域已有广泛的应用和不同的意义。统计学可以仅仅是一种数据的收集,例如公共卫生部所用的生命统计学,它归纳了一定人口的出生、死亡和发病数字的资料。更完整地说,统计学能够用来分析和解释数据,使得用者据此作出预测。在卫生保健方面,生物统计学将统计学应用于生物的状况,诸如药物研究的评价或者关于某一特定病人总体数据的解释。本文阐述了统计学的基本原理,叙述了程序和术语,以便有可能对数据进行归纳、分析和解释。

## 统计学的作用和统计人员

当正确使用,统计学在研究数据的分析和解释上起着重要作用。假定把药物A给予一组病人样本,药物B给予另一组类似病人样本。当治疗期结束时,可以观察到对药物的不同反应。统计检验可以帮助确定已观察到的差异是否能用随机性来解释,或这种差异是否显著。因此使用正确,统计学就可帮助把偶然巧合从有意义的模式区别出来。

统计检验的误用可能得出错误的结论而贻误读者。将数据硬性去适合不正确的统计检验可造成错误地支持研究者最初假定的结论。这样就能有意地滥用统计学去说服别人。

统计学可用于分析和解释数据,能使用户做出预测。

基本上有两种类型的统计学:描述统计学和推理统计学。顾名思义,描述统计学是描述已收集的数据,它可包括列表的定量的或定性的数据。定量数据包括那些可以计量的项目,如一种特殊药物在血液或尿中的数

量、尿排出量、身高和体重。定性数据则与个体据以分类的属性有关,如种族、性别或个性类别。描述统计学概括大量的数据,因此使数据易于处理。概括的数据可以用频率、百分率或同时用两者的形式表达(表1)。描述统计学通常与从给定数据中得出结论无关,它仅仅是表述情报。

表1 药学计算班学生的吸烟行为\*

	男性(%)	女性(%)	合计(%)
吸烟者	15 (13.3)	13 (11.5)	28 (24.8)
不吸烟者	35 (31)	50 (44.2)	85 (75.2)
合计	50 (44.3)	63 (55.7)	

\*n = 113

推理统计学试图得到有关样本总体的结论,它是用来回答研究的问题。如果一位研究者对比较阿司匹林和扑热息痛在降低婴儿发烧的效果方面有兴趣,那么所有的婴儿就构成了总体。然而,由于不可能去研究所有的婴儿,所以研究者就选择一个能进行研究的代表性样本。在收集数据后,研究者再应用推理统计学去确定这两种药物对研究中婴儿的作用是否有统计学上的显著差异。描述统计学组织数据,使其集中于焦点;推理统计学则对组织好的数据进行分析,并从中得出结论。

统计人员执行的职能概括如下:

1. 辅助研究设计,即在数据收集前计划和选择适合的统计检验。
2. 对数据进行分类和整理,通常采用表格或图形。
3. 归纳收集到的资料,尽可能采用百分数或平均数的形式。

4. 发现变量之间的关系。

5. 作出推论。研究者应用推理统计学就能够从一个合适的样本中归纳出一般性结论推论到抽取样本的那个总体。

6. 分析和解释数据。

7. 交流研究的成果。

根据其统计显著性, 统计检验有可能交流研究结果, 因而消除前面指出的偶然性。事实上, 多数科学家对未经统计检验和分析的研究结果都是怀疑的。

### 测量水平

大多数药学研究类型都涉及到进行某种形式的计量。一个对研究血压有兴趣的人要测量血压读数。有人研究药物清除率可以测量药物在血清或尿中浓度。这些测量的形式(有时称为测量尺度)有四种主要类型: 名义测量、有序测量、区间测量和比率测量。

#### 名义测量

名义测量是对客体进行分类的一种方法以便使具有某些相同属性或性质的所有客体归纳为一组。例如, 在分析科研数据的过程中, 研究者可以对性别编码, 指定所有女性的性别编码为0, 而所有男性编码为1。这就告诉你, 只有0组与1组是不同的, 而在同一组的所有个体的性别都是相同的。名义测量并不说明一组优于或高于另一组, 只是说明有差异。其它名义尺度包括种族、宗教信仰和血型。

#### 有序测量

在有序测量中, 属性或客体可以排序(如从最低到最高)。假定你想对药房的三名雇员(x、y、z)按工作量从轻到重排序; 那么他们可以排成y、z、x。在有序测量中数值记为1、2和3可以指定给雇员y、z和x。也可以分别记分为50、59和80, 因为在有序测量中, 任何两个相邻数目之间的距离是无意义的。有序尺度把观察值按大小顺序排列, 它们之间并无距离的考虑。它们只能告诉你一个数高点、大点或较另一个好点, 但

不能说明有多少。有序测量包括疲劳、疼痛程度, 易与别人相处和困难程度。

#### 区间测量

第三类尺度是区间测量。这种形式具有有序测量的所有性质, 此外相邻两个计数或数目之间的差值是有意义的。指定给客体的数值是该客体具有的测量单位数。区间尺度包括日历年份和温度度数。一种区间尺度没有绝对零; 相反, 零点是任意选择的。例如在日历年份中, 某年是任意设置的, 但零年并不意味着没有这一年。同样, 摄氏零度并不表示没有温度。由于区间测量的数值的差异是有意义的, 所以关于数值从区间测量要比从有序测量作出更有力的推论。

#### 比率测量

比率测量具有区间尺度的所有属性, 也有一个绝对零。因此, 零点不是任意的, 而确实说明不存在正被测量的性质。比率测量的例子是身高、体重、身长、年龄、尿排出量、药品数、事件发生后的时间、开氏温标测量的温度和反应时间等。关于测量的更专门资料可查阅文献。

### 变 量

变量是研究的个体或客体的特征, 例如性别、年龄、反应时间、血药浓度或血压。大多数研究者在统计分析中分析一种或多种变量, 用不同的变量取不同的值。

二元型变量是二值变量, 如男性—女性、生存—死亡、药学人员—非药学人员、联号药房—非联号药房。有些变量可以是多值的, 如宗教信仰、婚姻状况、药物种类或年龄组。变量也可以是离散的或连续的。离散型变量通常表示为整数。离散型变量的例子包括兄弟姐妹数、服用药品数及药学人员数。连续型变量可以是分数, 如年龄、身高或消逝时间。一个人可以是65 $\frac{1}{2}$ 岁, 但不能有1 $\frac{1}{2}$ 姐妹。

### 数据的显示

#### 图 形

数据分析的一种重要工具是图形，它是表示各种类型定量数据的有效方法。图形有助于阐明变量之间存在的关系。药房实践中最常用的图形是直方图、条形图、多边形图和圆形图。假定在一个医学术语班有10名学生，在一次测验中的成绩如下：100、60、99、80、83、85、81、95、100和50。这个资料可用直方图（图1）或多边形图（图2）来显示。如果这10名学生是由4个男性和6个女性组成，那么这新资料可用条形图（图3）或圆形图（图4）显示。图5是1985、1986和1987年医学术语学生按性别分组的复式条形图，它给读者清楚而迅速地显示了较大的资料。

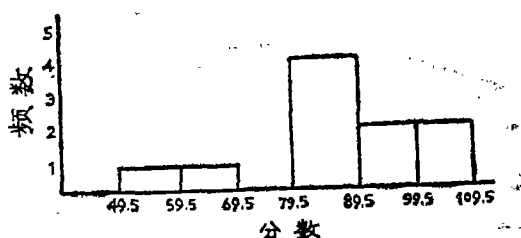


图1 医学术语班10名学生成绩的直方图

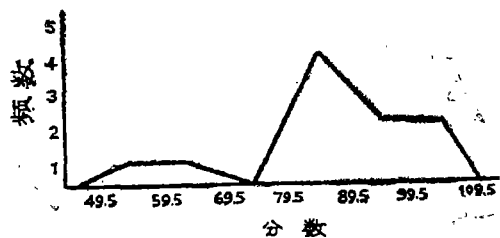


图2 医学术语班10名学生成绩的多边形图

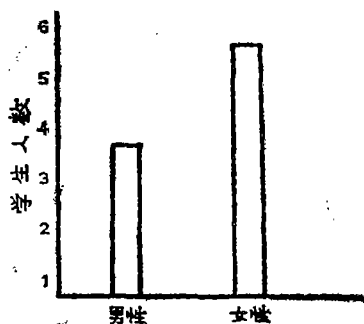


图3 医学术语班学生数的条形图

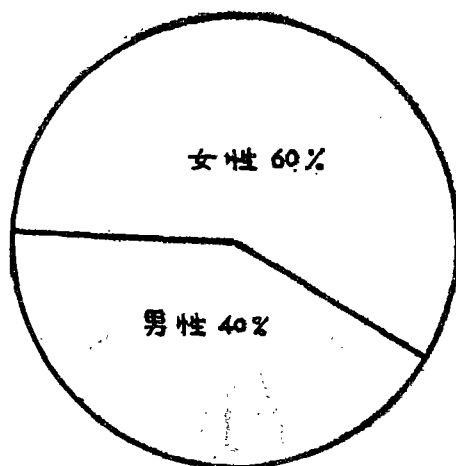


图4 医学术语班学生数的圆形图

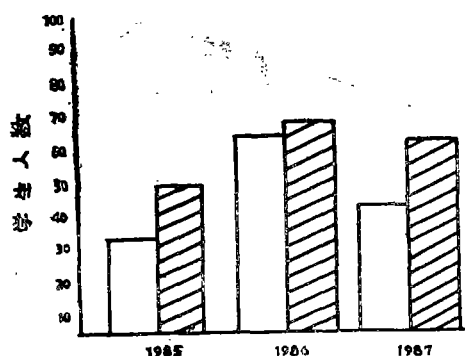


图5 医学术语班三年学生的复式条形图

#### 表格

表格是概括和阐明研究结果的另一种方法（表1），像图形一样，表格应向读者提供简明和完整的情报。表格应当仔细的标清不要包含过多的资料，否则可掩盖数据，使其难以理解。

中心趋势和离差的测量数通常用其它数据在表格中表述。这些测量数是数据分析中常见的统计资料。（待续）

[ Journal of pharmacy Technology

《药学工艺学杂志》，(11/12): 222~224, 1988  
(英文) ]