第二章　统计描述及其常见错误

统计描述是对数据的直接处理和分析，目的在于通过一定方式的描述与整理，计算统计数据的特征值，进而发现其数量的规律性，为用样本统计量（即样本数据的特征值）推断未知总体的参数提供充分的依据。因而，作为统计工作的重要组成部分，它是统计推断和统计预测的基础，并在一定程度上决定了统计工作的成败。由于第一章在统计工作的步骤和过程一节中已经对数据的收集与整理等统计描述的环节作了简单介绍，在此只对一些常用的概念和数据描述的方法作进一步的讨论和分析。

第一节　统计描述中的基本概念

一、集中趋势的测度

分布集中趋势的测度值反映数据一般水平的代表值或数据分布的中心值。从不同的角度考虑，集中趋势的测度值有很多，下面

介绍几个重要的代表值。

1. 众数

众数是将数据按大小顺序排列形成次数分配后，在统计分布中具有明显集中趋势点的数值。简单地说就是现象总体中出现次数最多的那个标志值。如果没有明显的集中趋势，众数可以不存在，当然，如果有两个集中趋势，也可以有两个众数。对于分组数据，计算众数的公式为

$$M_0 \approx L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i$$

其中：M_0 为众数；L 表示众数所在组的下组限；Δ_1 表示众数组次数与下一组次数之差；Δ_2 表示众数组次数与上一组次数之差；i 表示众数组的组距，即每组所包含的数据范围。

2. 中位数

把总体单位的某一标志值按大小顺序排列，居于中间位置的那个值就是中位数。中位数是一种位置性质的平均数，当某些现象不能进行严格的数学分析或不便于用算术平均数测定时，往往使用中位数。当原始资料未分组时，中位数的位置可由$\frac{N+1}{2}$来确定，其中 N 是数据的个数。若总体单位数是偶数时，中位数是处于中间位置的两个标志值的算术平均数。当数据是已分组的资料时，此时原始数据已被隐去，不能直接对其排队求其准确的中位数值，可由以下的近似公式计算：

$$M_e \approx L + \frac{\frac{N}{2} - S_{m-1}}{f_m} \times i$$

式中：$\frac{N}{2}$ 表示中位数所在的位置；L 表示中位数所在组的下组限；S_{m-1} 表示中位数所在组以下各组的累积次数；f_m 表示中位数所在组的次数；i 表示中位数所在组的组距。

3. 平均数

平均数是统计学中常用的概念，数学上有4种平均数，分别描述如下：

设 $x_1, x_2, \cdots, x_n$ 是未经整理的样本数据（不妨设 $x_i > 0$），则 $\sqrt{\frac{x_1^2 + x_2^2 + \cdots + x_n^2}{n}}$ 称为平方平均数；$\frac{x_1 + x_2 + \cdots + x_n}{n}$ 称为算术平均数；$\sqrt[n]{x_1 x_2 \cdots x_n}$ 称为几何平均数；$\frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}}$ 称为调和平均数。它们之间的关系为

$$\sqrt{\frac{x_1^2 + x_2^2 + \cdots + x_n^2}{n}} \geqslant \frac{x_1 + x_2 + \cdots + x_n}{n} \geqslant \sqrt[n]{x_1 x_2 \cdots x_n} \geqslant \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}}$$

上式等号成立的条件是 $x_1 = x_2 = \cdots = x_n$。统计学中最常用的是算术平均数和几何平均数，分别用 $\bar{x}$ 和 G 表示。一般情况下，凡个体标志值之和等于总体的标志总量的，可用算术平均数来反映变量值的一般水平；凡变量值乘积等于总比率或总速度的现象都可以用几何平均数来计算平均比率或平均速度。

二、离散程度的测度

在对统计分布或次数分配数据规律性的研究中，集中趋势表示的是分布的中心位置或一般水平的代表值，离散程度反映的则是分布离散和变异程度。变异指标的应用亦根据资料的不同而选取不同指标进行描述。常用的变异指标有极差、四分位数间距、方差、标准差和离散系数，尤其是方差和标准差更为常用。

1. 极差

极差又称全距，是数据最大值与最小值之差，它是数据离散或变异程度的最简单测度值，即 $R = \max\{x_i\} - \min\{x_i\}$。全距用于资料的粗略分析，其计算简便，但由于它只利用了数据两端的信息，故稳定性较差。

2. 百分位数与四分位数间距

百分位数是将 n 个观察值从小到大依次排列，再把它们的位次依次转化为百分位。在医学中，百分位数的另一个重要用途是确定医学正常参考值范围。百分位数用 P_x 表示，$0 < x < 100$，如25%位数表示为 P_{25}。四分位数间距又称内距，是由第3 四分位数（$Q_3 = P_{75}$）和第1 四分位数（$Q_1 = P_{25}$）相减计算而得，即

$$\text{内距} = \text{上四分位数} - \text{下四分位数} = Q_3 - Q_1$$

内距常与中位数一起使用，描述偏态分布资料的分布特征，比极差稳定。

3. 样本方差和标准差

样本方差是离差平方的平均数，即 $s^2 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$。它表示一组样本数据的平均离散情况。而标准差是方差的正平方根，即 $s = \sqrt{\dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$，使用的量纲与原量纲相同，适用于近似正态分布的资料，大样本、小样本均可，最为常用。

4. 离散系数

离散系数又称变异系数，用于观察指标单位不同或均数相差较大时两组资料变异程度的比较。因为在比较相关的两组数据的差异程度时，方差和标准差是以均值为中心计算出来的，因而有时直接比较方差并不准确，需要剔除均值大小不等的影响，计算并比较离散系数，其计算公式为 $V = \dfrac{\sigma}{\bar{x}}$，或 $V = \dfrac{s}{\bar{x}}$。前者为总体离散系

数,后者为样本离散系数。离散系数是从相对的角度观察差异和离散程度的,在比较相关事物的差异程度时,较之直接比较标准差要好些。

平均指标和变异指标分别反映资料的不同特征,作为资料的总结性统计量,两类指标要求一起使用。

三、偏态与峰度的测度

1. 偏态及其测度

偏态是对分布偏斜方向及程度的测度。利用众数、中位数和平均数之间的关系,判别偏态的方向并不难,但要测度偏态的程度则需要计算偏态系数,比较常用的计算公式是

$$\text{偏态系数} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3 f_i}{ns^3}$$

式中:s^3 是样本标准差的三次方;f_i 是权数。当分布对称时,正负离差可以相互抵消,因而偏态系数为0;当分布不对称时,正负离差不能抵消,就形成了正或负的偏态系数。偏态系数的数值越大,表示偏斜的程度越大。

2. 峰度及其测度

峰度(记为 K)是对数据分布平峰或尖峰程度的测度。峰度通常是与标准正态分布相比较而言的。如果一组数据服从标准正态分布,则峰度系数的值为0;若峰度系数的值明显不同于0,表明分布比正态分布更平或更尖,通常称为平峰分布或尖峰分布。峰度系数是用离差的四次方的平均再除以标准差的四次方,计算公式为

$$K = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4 f_i}{ns^4} - 3$$

用峰度系数说明分布的尖峰和扁平程度,是通过与标准正态分布

的峰度系数进行比较来实现的。由于正态分布的峰度系数为0，所以当$K>0$时为尖峰分布，当$K<0$时为平峰分布。如果上式不减3，此时比较的标准是3，当$K>3$时为尖峰分布，当$K<3$时为平峰分布。

四、分类资料的统计描述

1. 绝对数

绝对数是各分类结果的合计频数，用以反映总量和规模。如某地的人口数、发病人数、死亡人数等。绝对数通常不能相互比较，如两地人口数不等时，不能比较两地的发病人数，而应比较两地的发病率。

2. 常用相对数

相对数是两个有联系的指标之比，是分类变量常用的描述性统计指标，常用两个分类的绝对数之比表示相对数大小，如率、构成比、比等。

率又称频率指标，说明一定时期内某现象发生的频率或强度。计算公式为

$$率 = \frac{发生某现象的观察单位数}{可能发生某现象的观察单位总数} \times 100\%$$

率的表示方式有百分率(%)、千分率(‰)等。

构成比又称构成指标，说明某一事物内部各组成部分所占的比重或分布。计算公式为

$$构成比 = \frac{某一组成部分的观察单位数}{同一事物各组成部分的观察单位总数} \times 100\%$$

构成比的表示方式有百分数等。

比又称相对比，是A、B两个有关指标之比，说明A是B的若干倍或百分之几。计算公式为：比 $= \frac{A}{B}$。比的表示方式有倍数或分数等。

3. 标准化法

标准化法是常用于内部构成不同的两个或多个率比较的一种方法。标准化法的基本思想就是指定一个统一“标准”(如,标准人口构成比或标准人口数),按指定“标准”计算调整率,使之具备可比性以后再比较,以消除由于内部构成不同对总率比较带来的影响。标准化法只用于组间比较,不能替代实际率。

第二节　统计表与统计图的构造

统计表和统计图是显示统计数据的两种方式。统计表把杂乱的数据有条理地组织在一张简明的表格内,统计图则把数据形象地显示出来。一般情况下,用统计表和统计图描述统计数据比用文字表达更加形象直观。正确地使用统计表与统计图是做好统计工作的基本技能。

一、统　计　表

统计表是显示统计数据的基本工具,是表达统计资料的一种直观的表示形式,它既可用作调查表,又可用来进行整理和分析。统计表的形式有多种,根据使用者不同的要求和统计数据本身的特点,统计表可以绘制成不同的形式。表 2.1 就是一张比较常见的统计表。

从表 2.1 可以看出,常用的统计表一般由四个主要部分组成,即表头、行标题、列标题和数据资料。表头应放在表的上方,它说明的是统计表的主要内容;行标题和列标题通常安排在统计表的第一列和第一行,它所表示的是所研究问题的类别名称和指标名称。在编制统计表时,还应遵循以下规则:

(1) 内容力求简明扼要,主题突出,一目了然;

(2) 表题要简单地概括表的基本内容和资料所属的时间、地

点,标目要反映出横行纵栏的含义,并标明计量单位;

(3) 表内分组和指标的排列顺序要符合内容的逻辑关系;

(4) 字迹要清楚、规范,数字要排列整齐,同栏数据要有相同的精确度;

(5) 栏数较多的表,各栏要加编号,必要时还要注明各栏间的相互关系;

(6) 应尽量使用三线表,表的上、下两端划粗线,表头用细线分开,左右两边不封口,横行之间不必划线;

(7) 表中尽量少用横线,不用斜线;

(8) 表中的数据一般是右对齐,有小数时应以小数点对齐;

(9) 对于没有数据的单元格,一般用"—"表示,一张填好的统计表不应出现空白单元格;

(10) 必要时表下可加注释,说明资料的来源、制表人或单位、制表日期以及个别需要说明的指标或数据[3]。

表 2.1　某纺织系统已婚妇女对理想婚龄意愿的调查表[4]

企业编号	已婚女职工人数/人	平均理想婚龄/岁
1	495	24.1
2	1020	22.8
3	844	25.5
4	1518	24.6
5	635	25.8
6	394	23.7
7	2346	24.5

二、统　计　图

统计图是将统计指标以点的位置、线段的升降、直条的长短或面积的大小等几何图形直观地表示事物间的数量关系。与统计表一样,统计图可以从数量方面显示出研究对象的规模、水平、结构、发展趋势和比例关系,是表现统计资料的另一种重要的形式。它比统计表在形式上更加简化和形象,"生动活泼、鲜明醒目"是统计

图的突出的特点。熟练绘制和运用统计图是统计工作的基本要求。统计图中最常用的是频数图。根据应用的形式，统计图又可分为线图、散点图、条形图、圆形图、环形图、直方图、雷达图等。图 2.1 就是常见的线图的表现形式，通常由标题、标目、刻度和图例四部分组成。

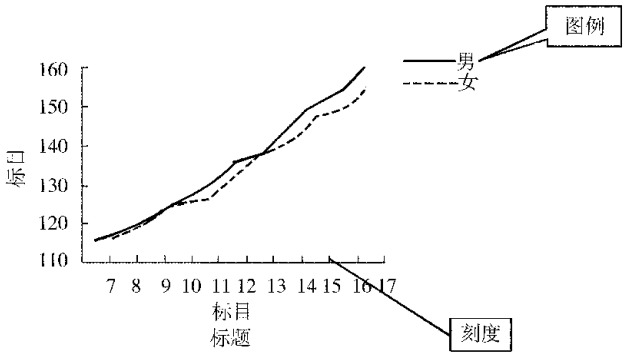

图 2.1　常用的线图的图形形式

（一）绘制统计图的基本要求

（1）根据资料的性质和分析目的，选择合适的图形；

（2）标题扼要说明图的主要内容，位于图的下方，必要时注明时间和地点；

（3）建立在直角坐标系上的统计图，其纵轴尺度自下而上，横轴尺度从左到右，数字一律由小到大，某些图还要求纵轴尺度从 0 开始（如直条图、直方图），纵、横两轴一般应有标目，注明单位；

（4）图的长宽比例（圆图除外）一般以 7∶5 或 5∶7 为宜；

（5）可用不同的线条或颜色表示不同的事物，但需用图例说明，一般放在图的右上角或图的下方。

（二）常用统计图的定义和制图要求

统计图的绘制相对于统计表要复杂，为简便直观起见，将常用

统计图的定义和制图要求列表如下(见表2.2)。

表2.2 常用统计图的定义和制图要求

名称	定义	制图要求
条形图	用等宽直条的长短来表示相互独立的各统计指标的数值大小	起点为0的等宽直条,条间距相等,按高低顺序排列
普通线图	适用于连续性资料。用线段的升降来表示一事物随另一事物变化的趋势	纵横两轴均为算术尺度,相邻两点应以折线相连。图内线条不宜超过3条
半对数线图	用线段的升降来表示一事物随另一事物变化的速度	横轴为算术尺度,纵轴为对数尺度。余同普通线图
圆形图	以圆面积表示事物的全部,用扇形面积表示各部分的比重	以圆面积为100%,将各构成比分别乘以3.6°得圆心角度数后再绘扇形面积。通常以12点为始边依次绘图
直方图	用矩形的面积来表示某个连续型变量的频数分布	常以横轴表示连续型变量的组段(要求等距),纵轴表示频数或频率,其尺度从"0"开始,各直条间不留空隙
散点图	以点的密集程度和趋势表示两种事物间的相关关系	绘制方法同线图,只是点与点之间不连接

第三节 统计描述中的注意事项及常见错误辨析

统计描述在整体上分为数据收集、数据整理和数据的汇总等环节,由于每个环节在具体操作上都因统计工作的性质和特点而有不同的形式,因而实际应用中必然存在很多问题,下面将针对每一环节相应问题的形式进行分类分析。

一、数据收集过程中的常见问题及改进方法

数据收集的方法之一就是统计调查，其组织形式和方式多种多样，普查是最基本的调查手段，具有涉及面广、工作量大、时间性强等特点。在具体的组织形式上，采访和问卷调查是目前较流行的收集数据的形式，它是通过既定的方案或设计的调查问卷，直接对单位或个人进行调查的一种方法。调查是否能够搜集到准确而全面的资料很大程度上取决于能否设计出一份高质量的调查方案或调查问卷表，其中经常出现的错误可归为以下几类：

（一）问题定义不准确

一个问题对于每个被调查者而言，应该代表同一主题，只有一种解释。定义不清的问题会产生很多歧义，使被调查者无所适从。例如，“您使用哪个牌子的洗发液?”这个问题表面上有一个清楚的主题，但仔细分析会发现很多地方含糊不清，假如被调查者使用过一个以上的洗发液品牌，则他对此可能会有四种不同的理解或回答：① 回答最喜欢用的洗发液品牌；② 回答最常用的洗发液品牌（最常但并不一定是最喜欢用的，例如受支付能力的影响）；③ 回答最近在用的洗发液品牌；④ 回答此刻最先想到的洗发液品牌。另外，在使用时间上也不明确：上一次?上一周?上一月?上一年甚至更长时间?都可由被调查者随意理解，这样的问题显然无法搜集到准确的资料。因此明确定义所要调查问题极其重要，下面的做法可以对这些问题进行改进。

（1）采取六要素明确法，即在问题中尽量明确什么人，什么时间，什么地点，做什么，为什么做，如何做。如上述的问题明确几个要素后改为：“在过去的一个月中，你在家中使用什么牌子的洗发液?如果超过一个，请列出其他的品牌名称”。这样定义的问题显然明确多了。

(2) 避免使用含糊的形容词、副词,特别是在描述时间、数量、频率、价格等情况的时候。如有时、经常、偶尔、很少、很多、相当多、几乎这样的词,对于不同的人有不同的理解。因此,这些词应用定量描述代替,以做到统一标准。

(3) 避免问题中含有隐藏的选择和选择后果,使隐藏的选择和后果明晰化。无论是是非式问题还是选择式问题,都是在几个备选选项中做出选择,因此,必须使被调查者清楚所有的备选选项及其后果,否则不能全面的搜集信息。

(二) 问题形式不妥当

问题的形式多种多样,大的可分为开放式、是非式、选择式、排序式、评分式、联想式等等;小的则涉及到一些语言技巧的运用和处理。问题形式的选择具有相当的艺术性,合理的形式选择与处理应使被调查者愿意,并且以最小的努力就能提供客观真实的答案。不恰当的形式选择会导致被调查者不愿意或不能够提供问题所要求的信息。例如,①"请问你家每人平均每年的食品支出是多少"?②"请问你个人每月的工资收入是多少"?③"人们都说 A 牌电视机比 B 牌电视机好,您是不是也这样认为的"?这三个问题都存在形式运用不当的问题。第一个问题要求被调查者付出额外的努力,进行复杂的计算,这可能使被调查者单方面结束访问。第二个问题涉及敏感的个人隐私,直接的提问容易遭到拒绝。第三个问题则带有引导性倾向,会影响被调查者的选择。问题形式的选择应注意以下几点:

(1) 基本信息应安排最前,分类信息居中,鉴别性信息放在最后。一般来说,应将最主要问题(基本信息) 置于最前面,然后列举后两类问题,只要前面的问题得到回答,那么,后面的问题如果被调查者不愿回答或因事中止也就无关大局了。

(2) 先易后难。容易、直观、清楚的问题置前,困难、复杂、敏感、窘迫的问题置后。

（3）总括性问题应先于特定性问题。总括性问题指对某个事物总体特征的提问；特定性问题指对事物某个要素或某个方面的提问。

（三）问题取舍不合理

问题的数量必须合理，应该既能保证搜集到全面的资料，又尽量保持问卷的简短，同时也尽力使问卷整体连贯、和谐、生动，能调动被调查者的积极性。有的问卷过于冗长，其中充斥着一些与调查主题毫无相关的问题；有的虽然短小，却不能全面搜集所需资料，而且过于严肃、死板，全文贯穿一问一答的形式，压抑被调查者的主动性。问题的取舍应注意以下几点：

（1）按调查主题组织问题，每个问题都应有益于调查信息的取得。应要明确调查的主题是什么，这是整个调查的基础，也是问卷设计的灵魂和核心所在。应避免为节省费用而附带调查主题之外的问题。问题东拉西扯，会使被调查者产生调查组织不严密的印象，影响他们的答卷态度。

（2）为了融洽调查气氛，不至于过于严肃、呆板，可以设置一些表面上与调查主题无关，但实质上有益于调查的问题。当问卷的调查主题较为敏感时，这点尤其有效。如在问卷开始，可以设置一些轻松的开放式问题，请被调查者畅述自己的看法，有利于调动被调查者的积极性；在各类信息的连接处，可以设置一些过渡性问题，顺畅被调查者的思维。

（3）为节省调查时间，保证被调查者符合调查对象的标准，可以在问卷开始设置一个“过滤性”问题，检查被调查者的合格性。如想调查现有掌上电脑的不足之处，则必然要调查掌上电脑的使用者。可以在问卷开始提问“您使用过掌上电脑吗?”这样就可检查被调查者是否合格，及时“过滤”不合格者了。

二、统计指标应用中存在的问题

（一）应用相对性指标的注意事项

（1）正确区分率和构成比。率和构成比所说明的问题不同，绝不能以构成比代率。构成比只能说明各组成部分的比重或分布，而不能说明某现象发生的频率或强度。例如：以男性各年龄组高血压分布为例，50 ~ 60 岁年龄组的高血压病例占 52.24%，所占比重最大，60 岁以上年龄组则只占到6.74%。这是因为60 以上受检人数少，造成患病数低于 50 ~ 60 岁组，因而构成比相对较低。但不能认为年龄在 50 ~ 60 岁组的高血压患病率最严重，而 60 岁以上反而有所减轻。若要比较高血压的患病率，应该计算患病率指标。

（2）计数资料相对数的分母不宜过小。一般来说，样本数量较多，计算的相对数可靠性也较大，否则相对数的计算没有说服力。

（3）用率或构成比进行组间比较时，要注意资料间是否有可比性。

（4）分组资料计算合并率时，不能用各个率简单相加，而应当用有关的合计数进行计算。

（二）平均数指标的应用范围和条件

（1）算术平均数适合于分布均匀的小样本数据或近似正态分布的大样本数据；几何平均数适合于等比级数的资料，尤其是平均增长速度、传染病发病的平均潜伏期等习惯上都用几何平均数表示；中位数适合于各种类型的资料，但尤其适合于大样本偏态分布的资料。

（2）对于同一资料，可能同时选用几个平均数指标，如对于某些偏态分布的资料，几何均数和中位数比较接近，对于这种情况，统计上的处理原则是，如果算术均数与中位数接近、几何均数与中位数接近，应采用算术均数或几何均数作为平均数指标，否则用中位数作为平均数指标。

（3）在医学方面，要了解各专业平均数的习惯用法。如儿童龋齿个数虽然呈偏态分布，但在口腔预防保健系统中，仍习惯上计算算术平均数。

（三）应用变异指标的注意事项

（1）方差与标准差属于同类指标，但标准差与均数的单位相同，作为变异指标多采用标准差。

（2）变异（离散）系数主要用于不同类型观察指标或同类型观察指标，但均数差悬殊时变异程度的比较。也常用于评价仪器测量精度和稳定性方面。

（3）四分位间距适合于任何分布的资料，计算结果比极差稳定，对于大样本偏态资料尤其适用。

（四）统计指标的误用举例

例1　为考察中西医结合法治疗滴虫性阴道炎和霉菌性阴道炎的疗效，对400名患者随机分两组，两组年龄、病程等方面的情况见表2.3，两组的疗效见表2.4。

表2.3　两组400名阴道炎患者基本情况

分组	观察人数	病型例数		年龄均数／岁	病程均数／天
		滴虫性	霉菌性		
中西医组	200	100	100	34（22 ~ 54）	29（3天 ~ 2.8年）
对照组	200	100	100	33（21 ~ 55）	32（4天 ~ 2.5年）

表2.4　两组400名阴道炎患者的治疗结果

分组	例数			百分比%		
	痊愈	有效	无效	痊愈	有效	无效
中西医组	191	9	0	95.5	4.5	0
对照组	147	25	28	73.5	12.5	14

该例中作者计算了中西医组和对照组患者病程的算术平均数以描述两组患者病程的平均水平，但是这两组患者病程分布都可能不是对称分布，对于非对称分布资料，算术平均数不能正确反

映数据的平均水平，应用中位数描述。

例2　为探讨青年士兵消化性溃疡病的相关因素，对436例收治的消化性溃疡士兵进行回顾性分析，考察士兵的军龄、籍贯、发病季节、吸烟情况等与发病的关系。结果，436例中有吸烟嗜好者326人，占74.8%，其中吸烟每天少于10支的占25.8%，10～20支的占60.1%，多于20支的占14.1%。由此认为“发病与吸烟及吸烟量有相关性”。

例2作者根据436例士兵吸烟情况的构成比大小认为吸烟是青年士兵消化性溃疡病的相关因素是不合适的，因为，若青年士兵实际吸烟比例较高，那么，即使吸烟与消化性溃疡病无关，青年士兵消化性溃疡病比例中，吸烟者的比例仍会较高。构成比只能反映事物的内部构成，不能说明事物的发生强度。可说明吸烟与消化性溃疡病的关系，正确的做法是对不吸烟及不同吸烟量的青年士兵进行前瞻性调查，统计各组的发病人数、计算发病率，根据各组发病率的大小来说明其间的关系。

三、统计表与统计图的常见错误辨识

统计表与统计图中的错误非常多见，根据编辑规范和各种类型错误出现的频率，只列举较典型的错误。

（一）统计表编制的常见错误

1. 不符合列表的原则和编制结构要求

这方面的错误其实是违背了编排规范的一种结果，大多是由于作者对编排规范掌握不足而导致的，也有是理解上的错误。表2.5是对119例宫颈糜烂冷冻治疗效果的列表，可以看出该表的主要目的在于表达冷冻治疗宫颈糜烂的近期疗效。存在的问题是：标题未突出“近期疗效”这一主要内容；主谓词安排不当且标目重复，如例数和%多处出现；总计意义不明确；线条过多，以致数据隔离，不便比较。改正后见表2.6。

表 2.5　119 例宫颈糜烂冷冻治疗结果　（原表）

效果	轻度糜烂		中度糜烂		重度糜烂		总计	
	例数	%	例数	%	例数	%	例数	%
治愈	39	32.77	11	9.24	2	1.68	52	43.70
好转	2	1.68	19	15.97	14	11.76	35	29.41
无效	8	6.72	7	5.88	17	14.29	32	26.89
合计	49		37		33		119	

表 2.6　冷冻治疗宫颈糜烂患者的近期疗效

糜烂程度	例数	疗效			疗效构成比/%		
		治愈	好转	无效	治愈	好转	无效
轻度	49	39	2	8	79.6	4.1	16.3
中度	37	11	19	7	29.7	51.4	18.9
重度	83	2	14	17	6.1	42.4	51.5
合计	119	52	35	32	43.7	29.4	26.9

2. 主辞和宾辞倒置，横、纵标目不明确

此类的错误也比较多见，多数是由作者自己的主观偏好或列表习惯导致的。表2.7就是此种错误的代表，修改后的结果见表2.8。

表 2.7　FDP 对幼兔离体心脏模型再灌注期肌酸肌酶的影响（$\bar{x} \pm s$，U/L，$n = 10$）

测定时间(min)	FDP 组	对照组
缺血前	6.73 ± 1.42	6.41 ± 1.26
5	12.50 ± 1.42^{bc}	39.33 ± 1.99^{c}
10	15.53 ± 1.09^{bc}	44.17 ± 1.28^{c}
30	20.20 ± 1.96^{bc}	48.10 ± 2.04^{c}

$^{b}P < 0.01$ vs 对照组；$^{c}P < 0.01$ vs 缺血前。

表 2.8　FDP 对幼兔离体心脏模型再灌注期肌酸肌酶的影响

分组	例数	肌酸激酶含量($\bar{x} \pm s$)/(U·L^{-1})			
		缺血前	再灌注期 5min	再灌注期 10min	再灌注期 30min
FDP 组	10	6.73 ± 1.42	12.50 ± 1.42^{bc}	15.53 ± 1.09^{bc}	20.20 ± 1.96^{bc}
对照组	10	6.41 ± 1.26	39.33 ± 1.99^{c}	44.17 ± 1.28^{c}	48.10 ± 2.04^{c}

$^{b}P < 0.01$ vs 对照组；$^{c}P < 0.01$ vs 缺血前。

（二）统计图的误用实例[5]

例 1　某行业季度销售额数据如图 2.2 所示。

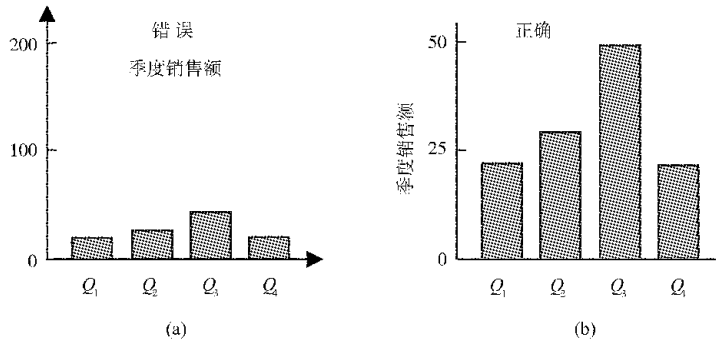

图 2.2　某行业季度销售额数据（单位：百万元）

图 2.2（a）中以 100（百万元）为纵轴单位，看上去四个季度的销售额差不多。实际上，如果以 25（百万元）作为纵轴单位（见图 2.2（b）），同样的数据就会充分反映出第一、二、三季度销售额不断增加而第四季度锐减的状态。

例 2　美国《时代》杂志 1979 年 4 月 9 日用如图 2.3（a）所示的象形图描述石油价格的变化。

图 2.3（a）表示的是一桶原油在运离沙特阿拉伯时的价格。1973 年是 2.41 美元，1974 年是 10.95 美元，……，1979 年是 13.34 美元。1979 年的价格与 1973 年相比，大约涨了 5 倍。从图中可以看到，绘图者为了表示出 6 倍价格的变化，在图 2.3（a）中将 1979 年的油桶画成 1973 年油桶的 6 倍高和 6 倍直径，这样 1979 年大桶的容积就是 1973 年的 $6^3 = 216$ 倍了。如果 1979 年这么大的一桶油只售 13.34 美元，那么原油价格非但没有上涨，反而下降了很多。同时，在反映原油实际价格变化的时候，还应该考虑物价指数的变化，即应将通货膨胀因素扣除掉。从 1973 年到 1979 年，美国的通货膨胀率上涨了近 1 倍，原油实际价格上涨了约 3.5 倍。图 2.3（b）真

实地反映了20世纪50年代末到70年代末美国实际油价的变化。

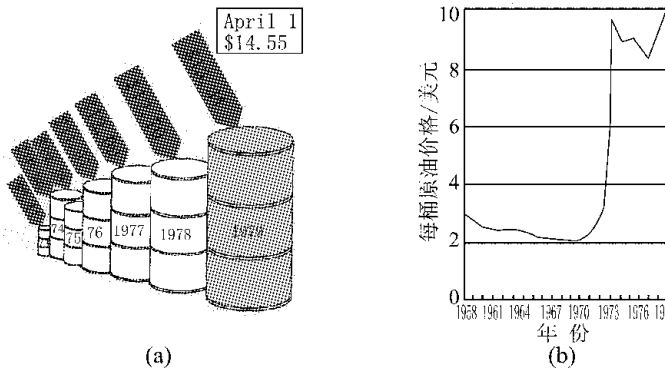

图2.3　1973 – 1979年美国原油价格的变化

(资料来源:*T. H. Wonnacoot, R. J. Wonnacoot. Introductory Statistics. John Willy & Sons*,1990)

例3　美国《纽约邮报》(*New York Post*)1981年4月刊登了如图2.4所示的统计图,并配上《纽约邮报》发行量在惊人地攀升的标题。

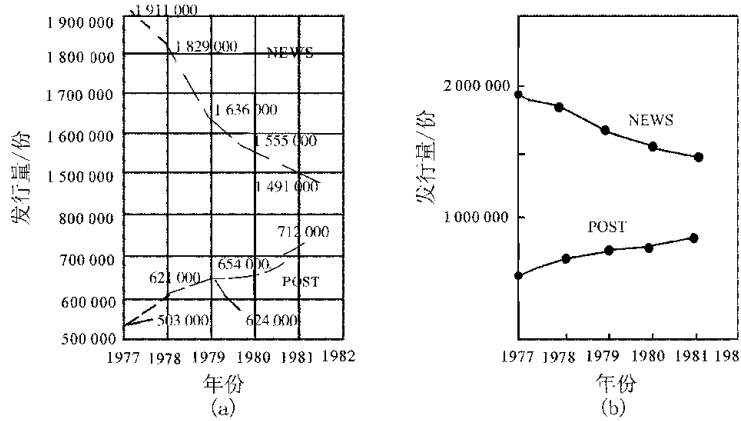

图2.4　美国《纽约邮报》和《新闻报》的发行量

(资料来源:*T. H. Wonnacoot, R. J. Wonnacoot. Introductory Statistics. John Willy & Sons*,1990)

乍一看图 2.4(a),《纽约邮报》(POST) 和《新闻报》(NEWS) 的发行量正在接近,似乎已经没有什么差别,但仔细观察,发现左边图中有两处错误。一是纵轴的发行量是从 500 000 为起点,而不是 0;二是纵轴从 800 000 直接跳到 1 500 000,数据间断又没有注明,就人为地造成两种报刊发行量接近的错觉。正确的画法如图 2.4(b) 所示。

绘制统计图中出现的错误还有很多的表现形式,尤其是在指标的选择、规范性要求以及数据的引用等方面更为普遍,我们一定要把握统计图制作的基本原理和规范化要求,从中发现、修改和消除工作中遇到的类似错误。