

• 思路与方法学 •

临床医师实用

统计学讲座

第四讲 计量资料的重要参数

皖南医学院 孙瑞元 黄志力

计量资料即量反应资料。要对计量资料进行统计分析,其方法可归纳为两大类,参数统计和非参数统计。参数统计常需要一个总体分布的前提,在临床医学统计中最主要的是要求数据资料的分布情况符合“常态分布”。即一组同质的计量资料虽然有大有小,但中等居多,集中分布在均数附近,特别大或特别小的数据很少。数据的分布可形成一个高峰位于中央(均数所在处),两侧均匀、对称的钟形曲线,即常态曲线(见附图a)。在很多情况下,一组计量数据的分布情况是符合常态规律的,因而参数统计方法是常规的、重要的分析方法。医学实践中也会遇到一些数据资料不符合常态分布,有时甚至分布情况不能确定的数据。在这种情况下,除有时可以通过数据转换的方法转为常态分布,再采用相应的分析方法外,通常的统计方法就不适用了,这时可采用非参数统计方法。

在对计量资料进行统计分析前,应首先认识资料中包含的信息,做到心中有数,如资料是否符合常态、是否有异常数据、是否需要数据转换等等,这样才能选择适宜的方法进行统计分析,以免盲目套用公式。

一、计量资料的基本参数

1. 均数(算术均数 $\bar{X}$ 或 M)

$$\text{公式(1): } \bar{X} = \frac{\sum X}{n} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

算术均数反映了一组测量数据的集中趋势,它具有下列性质:(1)均数与例数的乘积即为各值总和 $n \cdot \bar{X} = \sum X$;(2)各值与均数差值之和恒为零 $\sum (X - \bar{X}) = 0$;(3)各值与均数差值的平方和即 $\sum (X - \bar{X})^2$ 恒为最小值。

式(1)是计算均数的基本公式,对于大样本资料均数的计算比较麻烦,这时可采用加权法、简捷法,由于近年来计算器的普遍使用,可以准确、方便地计算均数,故近似的方法不再介绍。

均数除算术均数外,还有几何均数、调和均数……等,前者可以通过对数转换,后者可以通过倒数转换,仍可按算术均数的方法进行统计分析。

均数是反应数据集中趋势的统计指标。当计量资料呈非常态分布时,或数据偶而出现特别大或特别小

的数值时,算术均数就不能正确代表(反映)数据的集中情况,此时可用中位数表示之。中位数是一个特定的百分位数,在全部观察值中,有一半比它大,有一半比它小,它是全部观察值按大小顺序排列,位数居中的那个数值。

2. 标准差(S, SD, δ_{n-1})

$$\begin{aligned} \text{公式(2): } S &= \sqrt{\frac{\sum (X - \bar{X})^2}{n-1}} = \sqrt{\frac{\sum X^2 - \bar{X} \sum X}{n-1}} \\ &= \sqrt{\frac{n \sum X^2 - (\sum X)^2}{n(n-1)}} \end{aligned}$$

应注意: $\sum X^2$ 和 $(\sum X)^2$ 的含义是截然不同的。 $\sum X^2$ 为各值平方后总和,即 $\sum X^2 = X_1^2 + X_2^2 + \dots + X_n^2$; $(\sum X)^2$ 是各值总和的平方,即 $(\sum X)^2 = (X_1 + X_2 + \dots + X_n)^2$ 。一般计算器上都有统计功能键,可直接算出标准差,已取代用式(2)硬算的方法。

标准差是反映测量值离散性的参数,数据差距越大,标准差就越大;反之,数据愈集中,标准差愈小。

前已述及,均数是反映测量值集中性的参数,和标准差结合起来,应用常态分布规律能够对测量数据的分布情况作出概括的估计: $\bar{X} \pm \frac{2}{3}S$ 范围内约有50%的数据; $\bar{X} \pm S$ 范围内约有68%的数据; $\bar{X} \pm 2S$ 范围内约有95%的数据。利用这一点可以心算大致验证S值计算是否存在差错。例如:若在 $\bar{X} \pm 2S$ 范围内包括的数据还不到一半,则肯定算出的S值偏小,应找出原因,重新计算。

$\bar{X}$ 值一般大于S值,若出现 $\bar{X}$ 值小于S值,可能提示:(1)数值中有正有负,有无正负可以从观察值中直接看出;(2)排除(1)的情况后,可认为数据属偏态。

3. 例数(n)

不论计量资料还是计数资料,样本例数都是重要的基本参数,离开了它,统计分析无法进行。

均数、标准差、样本例数是计量资料的基本参数,报告论文,三者不可缺少,其它参数可由它们派生而来。

二、计量资料的派生参数

1. 标准误($S_{\bar{X}}$)

$$\text{公式(3): } S_{\bar{X}} = \frac{S}{\sqrt{n}}$$

标准误就是均数的标准差,它是用来判断标本均数在多大程度上可以用来估计总体均数(即样本均数)可靠性的统计指标。标准误小,说明抽样误差小,样本均数和总体均数较接近,用样本均数代替总体均数的可靠性较大;反之,标准误越大则表示样本均数越不可靠。样本均数的可靠程度可用均数加减标准误的范围来估计。可见 $S_{\bar{X}}$ 越小,由样本均数估计总体均数的误差范围也越小,均数的可信限越好。 $S_{\bar{X}}$ 恒小于 S ,有人感到用 $\bar{X} \pm S_{\bar{X}}$ 来表达数据外观更为好看,但两者含义并不相同,作为一组数据的代表参数,近年来许多杂志规定用 $\bar{X} \pm S$ 来表达,而不用 $\bar{X} \pm S_{\bar{X}}$ 。

2. 可信限

公式(4): 95% 可信限 = $(\bar{X} - t_{0.05} S_{\bar{X}}) \sim (\bar{X} +$

$t_{0.05} S_{\bar{X}})$

公式(5): 99% 可信限 = $(\bar{X} - t_{0.01} S_{\bar{X}}) \sim (\bar{X} +$

$t_{0.01} S_{\bar{X}})$

自由度 $f = n - 1$

可信限是由样本均数估计总体均数所在可信范围的参数。以95%可信限最为多用,通常不指明的可信限均提95%可信限。式(4)、式(5)中 $t_{0.05}$ 、 $t_{0.01}$ 需根据自由度 f 查 t 值表而得。当 $f \geq 10$ 时也可用下面的孙一宋氏 t 值公式计算。

公式(6): $t_{0.05} = 1.960 + \frac{2.376}{f-1.143} \approx 1.96 + \frac{2.4}{f-1}$

公式(7): $t_{0.01} = 2.576 + \frac{4.96}{f-1.66} \approx 2.58 + \frac{4.8}{f-2}$

常用 t 值简表

自由度(f)	5	8	10	15	20	30	40	80	120	∞
$t_{0.05}$	2.57	2.31	2.23	2.13	2.09	2.04	2.02	1.99	1.98	1.96
$t_{0.01}$	4.03	3.36	3.17	2.95	2.84	2.75	2.70	2.69	2.62	2.58

$t_{0.05} = 1.958 + \frac{2.45}{f-1}$ (取二位小数,适用于 $f=3 \rightarrow \infty$)

3. 变异系数(CV)

当标准差的单位不同或单位相同而均数悬殊时,为了便于比较,必须有一个共同的标准做依据,这个标准即是变异系数。即标准差占均数的百分比,计算公式为 $CV = \frac{S}{\bar{X}} \times 100\%$

均数、中位数等描述数据的集中趋势,标准差、标准误、变异系数、可信限等表示数据的离散程度。

三、数据有无偏态的判断

判断资料是否为偏态,也即是与常态分别的差别有无显著意义,其方法有正态概率纸目测法、矩法及 D 检验法,但计算均较繁复,请参阅有关专著。下面提供一种简单判断法,结合例子说明计算步骤。

1. 步骤

(1) 求出均数及标准差: 如数据为6, 8, 8, 8, 8, 9, 9, 9, 20, $n=9$; 算出 $\bar{X}=9.44$, $S=4.06$ 。

(2) 求内外例数差(绝对值): 所谓“内”、“外”是指数据分布在 $\bar{X} \pm 0.65S$ 范围内与外,本例 $\bar{X} \pm 0.65S$ 的范围为6.8~12.1,其中6, 20两例位于该范围以外,余者7例均在该范围内。所以 $n_0=2$, $n_1=7$,则内外例数差的绝对值为 $|n_1 - n_0| = |7 - 2| = 5$ 。

(3) 求高低例数差(绝对值): 这组数据中只有20大于均数9.44,故高于均数侧的例数为1例, $n_h=1$,其余8例数据均小于9.44,故低侧例数为8例, $n_l=8$,高低例数差的绝对值为 $|n_h - n_l| = |1 - 8| = 7$ 。

(4) 判断: $|n_h - n_l| \geq 2\sqrt{N}$ 表示资料在对称性上有偏态(见附图b)。 $|n_1 - n_0| \geq 2\sqrt{N}$ 表示资料在峰坡性上有偏态(见附图c)。

在对称性上或峰坡性上任一种有偏态时,即可认为资料分布与常态分布的差别有显著意义($P < 0.05$)。

本例 $2\sqrt{N} = 2\sqrt{9} = 6$, $|n_h - n_l| = 7$, $|n_1 - n_0| = 5$, $|n_h - n_l| > 2\sqrt{N}$, 对称性有偏态。该资料分布不符合常态,不应采用均数,标准差等参数进行统计分析。应进行数据转换或采用非参数统计法。

2. 讨论

(1) 本方法可由 X^2 基本公式导出,如判断资料对称性与常态分布是否有显著差别时,由常态分布的对称性可知: 大于均数和小于均数的例数应相等。所以无效假设为 $n_h = n_l$ 。据此假设,高低两侧的理论频数均应为 $\frac{N}{2} = \frac{n_h + n_l}{2}$, 由 X^2 基本公式:

$$\begin{aligned}
 X^2 &= \sum \frac{(A - T)^2}{T} = \frac{(n_h - \frac{n_h + n_l}{2})^2}{\frac{n_h + n_l}{2}} \\
 &\quad + \frac{(n_l - \frac{n_h + n_l}{2})^2}{\frac{n_h + n_l}{2}} = \frac{4(\frac{n_h - n_l}{2})^2}{n_h + n_l} \\
 &= \frac{(n_h - n_l)^2}{N}
 \end{aligned}$$

以 $X^2_{0.05}=3.84$ 为判断显著性水平,若差异有显著意义($P\leq 0.05$)则必须 $X^2=3.84$ 。

$$\text{即 } \frac{(n_h - n_l)^2}{N} \geq 3.84 \quad \text{则有 } (n_h - n_l)^2 \geq 3.84N$$

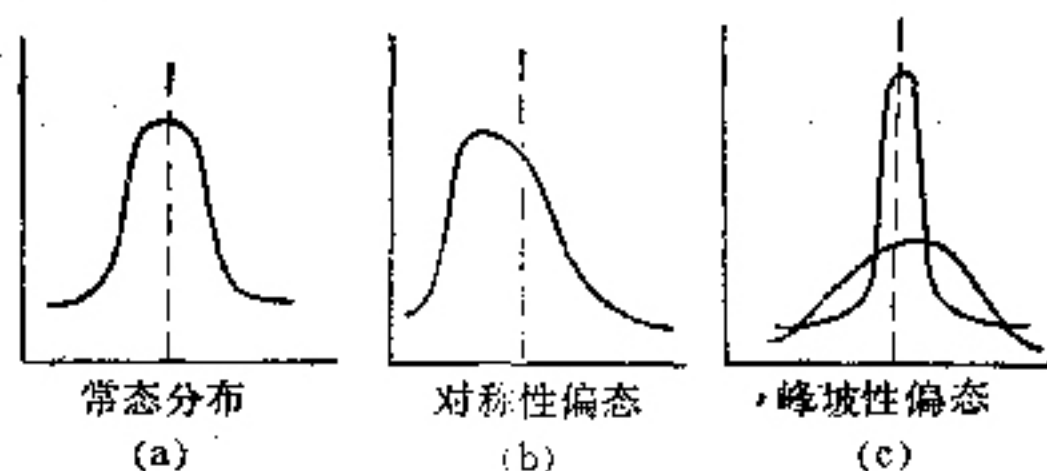
$$\text{两边开方取绝对值: } |n_h - n_l| \geq \sqrt{3.84N}$$

为方便起见,我们没有采用 X^2 公式中的校正数0.5,使 X^2 值有所增大,这里将 X^2 显著性的临界值也适当增大,上式写成: $|n_h - n_l| \geq 2\sqrt{N}$

同样,常态分布在 $\bar{X} \pm 0.65 S$ 范围内包括了50%的数据,即该范围内外的例数应基本相同,因此同理可以得到判定资料在峰坡性上是否有偏态的式子。

(2)应该指出:本法只判断“差值大于 $2\sqrt{N}$,则资料有偏态”,并不表示“差值小于 $2\sqrt{N}$,则资料肯定

没有偏态”,只是这时数据分布与常态分布的差异不显著罢了。该方法虽较粗简,但在统计分析前可迅速、简便地判断资料有无偏态情况,对于选择统计方法颇有价值。



附图 常态和偏态示意图 (待 续)

乌梅四物汤治疗胃下垂2例报告

甘肃省天水市中滩中心卫生院 王显周

乌梅四物汤出自《雷氏家传验方集》,近年来用此方治疗胃下垂患者2例,获得了比较满意的效果,现报告如下。

例1 王××,女,52岁,已婚。1972年3月初诊,全身乏力,气短,语言低微,肌肉消瘦,面色萎黄;食后胃脘坠胀,暖气,时有恶心呕吐;大便溏薄,纳差,微动则汗出;脉搏濡弱,舌淡苔少。X线胃肠钡餐透视,诊断为胃下垂。患者曾服补中益气汤、六君子汤等治疗,均未见效,服药后恶心、呕吐等症反有加重。脉证合参,治宜补血行滞,扶胃敛收,遂投乌梅四物汤治之。方药:白芍(醋炒)、红糖各15g,当归(醋炒)、熟地各10g、乌梅7枚,每剂水煎分2次服,每日1剂。从第2剂开始,每服1剂,加乌梅1枚;从第8剂开始,每剂减乌梅1枚,减至7枚时,共服13剂,为1疗程。患者服完第1疗程后,自觉腹胀减轻,再服1疗程,腹胀明显减轻,食纳有

增,脉能应指,舌质淡红、润。随后每服两个疗程,停药10~15天。先后共服12个疗程后,诸症相继消失,肌肤红润有力;面色光泽有神,食纳如常人。随访观察9年,病情无反复。

例2 吴××,女,36岁。1979年12月患腹泻,久治不愈,后时有复发,1年后感觉腹胀下坠,时有呕吐,食纳极少,且完谷不化;经期迁延,经血色淡量少;其形骨瘦如柴,面色晄白,乏力少语,脉搏微弱,舌体瘦小,色淡少津。X线胃肠钡剂造影检查,诊断为胃下垂。血虚之象明显,加之每有腹泻,治宜补血以温中土,固脱以扶脾胃。投方乌梅四物汤加砂仁。服1疗程后,食欲有增,再无呕吐,自觉周身有力,脉舌如前。唯患者自觉有寒凉之感,遂于原方加肉桂5g,焦白术10g,再服1疗程,诸症遂减,肌体渐复。后以原方配散剂冲服两个月即停药。随访8年,无复发。

“全国第三届中西医结合研究急腹症学术交流会”征文通知

全国第三届中西医结合研究急腹症学术交流会将于1988年9月在沈阳市召开,现将征文事宜通知如下。

一、征稿内容:(1)急腹症中西医结合临床研究的新经验、新技术。(2)近几年来急腹症实验研究的新进展。(3)特殊类型急腹症中西医结合治疗经验(老年、妊娠、小儿急腹症)。

二、要求:(1)稿件要求有全文及1000字以内摘

要各一份。(2)要用稿纸书写,字迹清楚。(3)稿件上要写清楚省、市及工作单位、作者姓名(请注明第一作者的年龄)。

三、截稿日期:1988年4月30日以前,以邮戳为准。

四、稿件加盖公章报送本省分会,由分会统一寄沈阳市、辽宁省中医研究院贺瑞麟收。