

实验动物科学研究中生物统计方法应用的常见错误

高 尔 生*

(上海实验动物研究中心)

任何一门学科的研究工作都需要解决研究方法问题,生物统计方法是各种生命科学研究共用的重要研究方法。有生命就有变异,要研究变异事物的规律就必须正确地应用生物统计方法。应用生物统计方法可以帮助研究者:**a.** 正确地进行研究设计,使研究少走弯路,减少不必要的人力、物力和财力的浪费;**b.** 正确地解释研究结果和做出研究结论;**c.** 正确、清楚而简洁地表达其研究结果。

最近,作者认真地阅读了1991年以来发表在国内主要的实验动物科学专业杂志上的论文,感到大多数论文的生物统计方法应用正确,但也发现不少问题。为了提高实验动物科学研究工作的质量,现从这些论文中发现的生物统计方法应用方面常见的错误情况归纳出来,供大家讨论和思考。

1. 统计指标的正确应用

实验动物科学研究的很多结果是以数据的形式出现,要正确、概括地表达研究结果就要应用统计指标予以描述。数据分计数和计量两大类,统计指标也随之有两大类。

表达计数数据平均水平的统计指标是相对数,有:率、构成比和相对比三种。它们的定义及计算公式不同,切忌混用。率是事物发生的强度指标,由事物发生的阳性数除以所有可能发生的事物总数;也就是说,分子都是由分母转化而来,而分母都有可能转化为分子。阳性率和阴性率之和为100%,阳性率不可能超过100%。有人研究各种因素对小鼠IVF胚冷冻效果的影响^[1],把受精胚冷冻后的回收数为总数(N),解冻后对形态异常胚(d)不予培养,而只对形态正常胚培养,培养结果分为发育(a)、未发育(b)及死亡(c)。该文作者分别计算了 a/N , b/N , c/N 和 d/N ,称 a/N 为发育率。这是错误的, a/N 不是率,而是一种构成比,因为分母中有一部分(d)未予培养,显然不可能发育而转化为分子(a)。若计算发育率,公式应为: $a/(a+b+c)$ 或 $a/(N-d)$ 。又如在一次不同浓度TNF对筛窦癌细胞抑制作用的研究中,作者定义:“设对照组集落存活率为100%,治疗组/对照组 $\times 100\%$ 为治疗后集落存活率”^[2]。据此定义计算,有一治疗组“存活率”为110%。实际上作者定义的存活率不是“率”,不表示“筛窦癌细胞”在不同TNF浓度下存活的百分数。此“存活率”实际上是相对比,表示治疗组的集落数相对于对照组集落数的比例。

应用相对数时,另一个十分重要的问题是需要列出绝对数。在实验动物科学研究中,常常应用贵重大动物或濒危动物,因而数量可能很少,此时,若只报道相对数,可能误导读者。如一项关于大鼠肺霉形体(原文为肺支原体)血清抗体阳性率的研究中,作者报道检出阳性率为100%^[3],读者会感到阳性率非常高。该文作者附上了绝对数,显示仅检查3只鼠。读者由此可知,这很有可能来自偶然性,需进一步研究后才能说明问题。在实验动物研究中“敏感性”是一个常用指标,对已知阳性的标本,某一新方法的阳性检出率称新检出方法的敏感性,这是指某一检出方法相对于一个标准检出方法而言。若两个方法中没有一个是标准时,也就没有敏感性而言,更没有敏感性

* 通讯地址:上海实验动物研究中心,上海市斜土路2140号 200032

哪一个高的问题^[4]。特异性与敏感性含义相对,特异性是对已知阴性的标本用新方法检测时的阴性率,同样特异性是相对于标准检测方法而言。敏感性和特异性均表示某新检出方法的正确性。

表示计量数据平均水平的统计指标为平均数,常用的有均数、几何均数和中位数。均数用于正态分布,至少用于分布对称的数据;几何均数是用于取对数后呈正态或对称分布的数据;而中位数可用于各种分布数据,特别是偏态分布、取对数后也不能转化为正态的数据。正态分布的数据通常用标准差(s)表示离散程度,而偏态数据以四分位间距或百分位间距表示其离散程度为好。有一猕猴血清肌酸激酶(CK)的研究,CK标准差(106)和平均数(115)两者相近^[5],说明该指标呈正偏态,因而最好用中位数表示肌酸激酶的平均水平。如果意在表示某一指标的正常生理参考值时,需注意:①动物数一般应在100以上,②范围用 $\bar{x} \pm 1.96s$ (正态分布时)或 $P_{2.5} \sim P_{97.5}$ (第2.5百分位到第97.5百分位间距,正态与偏态均可)表示。动物太少时,数据可靠性就差。用全距表示变异,可因个别特殊的个体使范围波动。

2. 统计分析的正确应用

实验动物科学通常的研究对象是样本,而希望得出关于总体的结论。由于个体有变异,因而样本统计量间有差异,不一定说明总体的参数间也有差异。所以在两样本的统计指标有差别而希望推论总体时,需作统计学检验。如用ELISA和HAI二种方法测定Reo 3病毒抗体时,均测定7⁴个标本,两者阳性率分别为18.91%和9.45%^[6]。如据此认为ELISA的检出率是HAI的2倍,并以此确认“ELISA检出Reo 3病毒抗体比HAI好”,则得出了一个错误结论。对此数据作 χ^2 检验, $\chi^2 = 2.72$, $P > 0.05$,两个阳性率差别在统计学上无意义,因此可以认为ELISA和HAI两种方法对Reo 3病毒抗体检出率是没有差别的。又如为确定清洁垫料是否比普通木屑垫料有助于鼠的生长的一项研究中,测2组鼠50天的平均体重,均数和标准差分别为 21.7 ± 0.5 和 19.4 ± 0.4 。作者结论是“两组动物各期的体重变化无显著性差异”^[7]。从论文提供资料可知,两组动物样本量在400头以上。因而作 t 检验, $P < 0.05$,则结论应为“两组动物50天平均体重差异有统计学意义,用清洁垫料鼠的50天体重高于用普通木屑垫料鼠”。

资料不同,选择的显著性检验方法也不同。在2个样本之间比较时,应注意是否属配对资料。如研究低温灌注对离体心脏湿重是否有影响的显著性检验中,应该用配对 t 检验而不是用成组 t 检验。同理,多组比较时,应注意到研究是否是随机区组设计。例如,比较不同时点上缺乏锌等微量元素对蛋清中脂质过氧化的影响的研究^[8],研究设计为随机区组设计,因而显著性检验应作随机区组的方差分析。值得注意的是,多组比较或两组的多区组重复比较,不能做多个 t 检验,而应作多组完全随机的方差分析(或完全随机区组方差分析),然后再作两两比较。同样,多组的率和构成比比较时,应先作 $R \times C$ 的 χ^2 检验,然后再作各个率间的两两比较。

回归分析也是生物统计中常用的统计方法。回归通常分直线和曲线两大类。一般先将数据在坐标纸上画成点图,观察趋势后再行拟合回归方程。如小鼠与日龄的体重关系^[9],一般呈S形曲线,也称生长曲线,应拟合曲线回归。呈曲线回归的资料则不宜分段拟合直线回归,否则就不能正确显示资料的特征和发展趋势。

3. 统计图表的正确应用

统计图表是资料表达的重要方式,用统计表以精确、简明地表达研究结果;而统计图给人们以形象、直观的感觉。但需正确地应用,否则反而给人以错误印象。

前面已经说过,因为有变异才需应用统计,同样只有表达事物间的差异才需要应用统计表。统计表可避免冗长的文字叙述,又易清楚地表达情况。相反,如各组的研究结果及分布完全相

同^[10,11]，只需用简单文字描述即可，此时不必要应用统计表。

关于统计表正确应用的一个重要方面是正确地分组，各组的组值不能重叠，也不能遗漏。如凝集试验结果^[12]，分为-、+~卅、卅3组就不妥，第2组和第3组重叠，结果“卅”既可在第2组，也可在第3组。又如分娩后时间分为0~24小时、24~48小时二组^[13]，亦是分组重复，若正巧在分娩后24小时，把它分在哪一组就有困难了。

图表中线条的正确应用十分重要。线条一般越少越好。多数统计表只用顶线、底线和纵标目与数字之间的分隔线。表中不宜用斜线，应用斜线一方面表达不清，且印刷上亦不方便。纵线不用为好，特别不要用纵线将表框住^[14]。

表中的数字应正确表达，表中小数点位数应相同，并对齐^[15]。表中不应出现文字描述，若需说明某问题时，可应用注释。标目中需注明单位，但单位不应写在表中间。有用“ $\bar{x} \pm s$ ”或“ $\bar{x} \pm s_x$ ”时应写清，不标明时可认为是“ $\bar{x} \pm s_x$ ”，用“ $\bar{x} \pm s$ ”时需写出样本含量(n)。无数字时不应出现

表1 三种不同血清对半胚培养的影响

培养液	培养半胚数	发育半胚数
PBS+20%犍牛血清	353	252(64.95%, 252/353)
PBS+20%绵羊血清	412	303(73.54%, 303/412)
PBS+20%马血清	388	295(76.03%, 295/388)

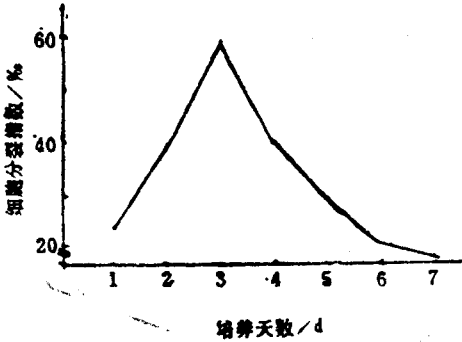


图1 MA-901 C 细胞分裂指数

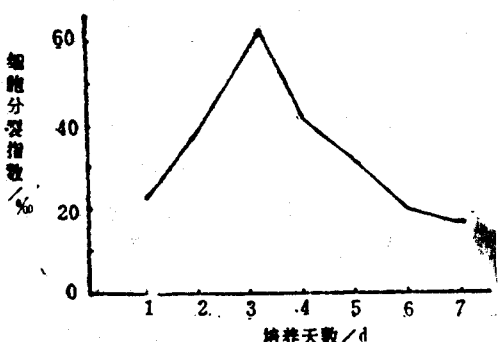


图3 MA-901 C 细胞分裂指数

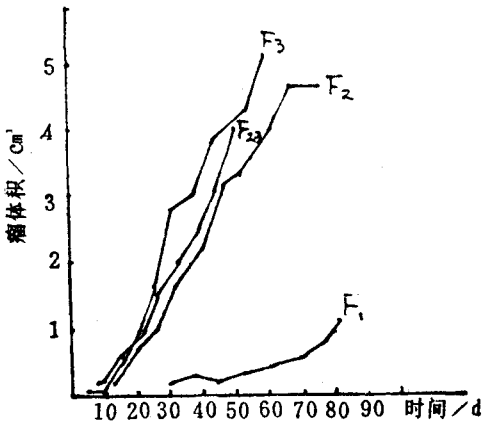


图2 HEC₂ 的生长曲线

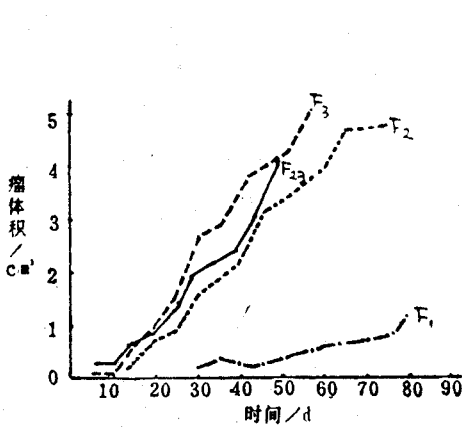


图4 HEC₂ 的生长曲线

表 2 24 只犬华支睾吸虫感染分布情况

项 目	胆囊	肝左叶		肝中央叶	肝右叶			整体(肝)
		外侧叶	中侧叶		乳头叶	尾状叶	外侧叶	
感染例数	13	18	16	17	12	15	15	21
		19			16			
感染率(%)	54.17	75.00	66.67	70.83	50.00	62.50	62.50	87.50
		79.17			66.67			

表 3 3 种不同血清对半胚培养的影响

培养液	培养数	发育率/%
犊牛血清	353	64.95
绵羊血清	412	73.54
马血清	388	76.03

表 5 常用统计图的适用资料及表示意义

种类	适用资料及表示意义
直条图	相互独立的各组资料间的对比关系
直方图	连续性资料,说明频数分布
线图	连续性资料,说明变动趋势或速度
普通线图	示变化趋势
半对数线图	示变化速度
百分构成图	百分比资料,说明事物内部各部份比重
相关(点)图	说明两种事物间的相互关系
统计地图	说明事物的地区分布情况

表 4 24 只犬肝胆华支睾吸虫感染率

部 位	感染率/%
胆囊	54.14
肝脏	87.50
左叶	79.17
(外侧叶)	75.00
(中侧叶)	66.67
中叶	70.83
右叶	66.67
(乳头叶)	50.00
(尾状叶)	62.50
(外侧叶)	62.50

空格,用“-”表示,缺损资料用“……”表示。

统计表的好坏以是否表达清楚、简洁扼要为指标。读者可比较表 1^[16]、表 2^[17]和表 3、表 4,显然后者既清楚地表达了数据特征又十分简洁。

在论文中常用的统计图有直方图、直条图、线图(普通线图和半对数线图)、构成图、相关图和统计地图,它们的适用资料及表示意义如表 5。

一般统计图长宽比例为 7:5,纵横标尺应等距或对数等距,纵轴一般不应折断,否则会造成错觉^[18,19]。直条图间的空隙为直条的 1/2~1 为宜,否则不太好看^[20]。线图中应用粗细虚实的各种线表示,应有图例说明,以便表达清楚^[21]。图 3、图 4 为图 1、图 2 的改进,显然后者表达更清楚、更美观。

作者建议:实验动物科学的研究者在实验设计和撰写论文时,要复习生物统计学并请教生物统计学家,以期保证自己工作及其结论的科学性,尽量少出错误。

参 考 文 献

- 1 刘东军等.昆明小鼠体外受精胚冷冻保存的研究.中国实验动物学杂志,1991,1(3,4):199
- 2 李五一等.关于肿瘤坏死因子及 γ 干扰素体外协同抑制鼻窦癌裸鼠移植瘤的初步研究.中国实验动物学杂志,1992,2(1,2):35

- 3 贺争鸣等. 实验大鼠肺支原体 ELISA 检测方法的建立. 北京实验动物科学, 1992, 9 (3): 10
- 4 贺争鸣等. 检测大鼠细小病毒抗体的 ELISA, IEA 和 HI 法比较. 上海实验动物科学, 1993, 13(3): 132
- 5 宋怀燕等. 成年恒河猴血清生化变量的分布特点及其相关关系的分析. 中国实验动物学杂志, 1992, 2 (1, 2): 26
- 6 吴惠英. 实验大鼠 Reo 3 病毒抗体试剂盒的研制与 Reo 3 病毒血凝条件的探讨. 北京实验动物科学, 1993, 10(1): 10
- 7 郭 健等. 应用玉米芯为原料的清洁实验动物垫料饲养动物实验报告. 北京实验动物科学, 1992, 9(3): 28
- 8 李建瑞等. 缺乏锌等微量营养素对鸡血液和鸡蛋谷胱甘肽过氧化物酶和脂质过氧化的影响. 中国实验动物学杂志, 1993, 3(1, 2): 35
- 9 赵玉杰. 昆明种小鼠体重对日龄的回归分析. 上海实验动物科学, 1992, 12 (4): 223
- 10 张业彬. 清洁级 NIH 系小鼠种群的建立与应用. 上海实验动物科学, 1992, 12(1): 28
- 11 孙 光. 实验动物中单核细胞增生性李氏菌分离方法的研究. 中国实验动物学杂志, 1991, 1(2): 76
- 12 钱心文. 猪丹毒杆菌毒力和敏感性检查试验用家鸽素质的影响. 上海实验动物科学, 1991, 11(1): 32
- 13 王富生. 提高小鼠剖腹取胎育仔成活率的初步探讨. 上海实验动物科学, 1991, 11(2): 81
- 14 赵钧铭. BN 大鼠正常血像和骨髓像的观察. 北京实验动物科学, 1993, 10(4): 44
- 15 石贵山. 用 LA-BRG 系列颗粒饲料饲养豚鼠效果的研究. 中国实验动物学杂志, 1993, 3(1, 2): 43
- 16 陶 涛. 分割小鼠胚胎的简易方法和分割后二分胚的发育能力. 中国实验动物学杂志, 1991, 1(3, 4): 235
- 17 张恒仁. 家犬华支睾吸虫感染肝胆分布情况. 上海实验动物科学, 1992, 12(1): 51
- 18 李 革. BALB/c 小鼠乳腺癌细胞系 (MA-901 c) 的建立及部分生物学特性. 中国实验动物学杂志, 1992, 2(3, 4): 109
- 19 祝小枫. SPF 大鼠失血性休克后肝肾和肠系膜淋巴结中溶菌酶含量的动态观察及其意义. 上海实验动物科学, 1993, 13(2): 72
- 20 陈惠玲. “乳珍”对小鼠生殖系影响的观察. 上海实验动物科学, 1991, 11(4): 212
- 21 蔡海英. 人食管癌裸鼠移植瘤株建立及其应用的研究. 北京实验动物科学, 1992, 9(4): 17