

基于 Excel 的统计分析系统在期刊编辑部审稿中的应用*

高永 张中文 石德文 徐天和

滨州医学院中国医院统计杂志社, 264003, 山东烟台

摘要 期刊审稿过程中非统计专业审稿人很难对统计结果进行审核,导致论文中的统计学错误很多。基于 Excel 的统计分析系统,不需要专业的统计学知识,只要了解一般统计学常识就能方便使用,特别适合期刊编辑部编辑和审稿人员审核稿件中常见的统计学问题。实际应用结果表明,该系统非常实用,对减少期刊论文中的统计错误大有裨益。

关键词 期刊编辑部;审稿;统计分析;Excel

Application of the statistical analysis system based on Excel in journal review // GAO Yong, ZHANG Zhongwen, SHI Dewen, XU Tianhe

Abstract It is difficult for reviewers who are not professional in statistics to examine statistical results in manuscripts; therefore, lots of statistical mistakes might remain in the papers even after peer review. We develop a statistical analysis system based on Microsoft Excel. People who understand general knowledge of statistics can use it well. Journal editors and reviewers can use this system to examine and verify common statistical problems in manuscripts and do not need to master advanced professional knowledge of statistics. The system is very practical and can be applied by a wide range of journal editorial departments to reduce statistical errors in the papers.

Keywords journal editorial department; manuscript review; statistical analysis; Excell

Authors' address Editorial Department of Chinese Journal of Hospital Statistics, Binzhou Medical University, 264003, Yantai, Shandong, China

科技论文,尤其是医学论文,经常涉及到统计学分析,在审稿过程中审稿人很难对统计结果进行审核,导致许多期刊发表的论文存在大量统计学错误^[1]。有学者对我国多种学术期刊的调查发现,论文中统计学应用错误率都超过 50%^[2-3]。错误的统计结果往往得出错误的结论,严重影响了科技期刊的质量。为解决这一问题,我们基于 Excel 软件研究出一套非专业统计人员能方便使用的统计工具——基于 Excel 的统计分析系统,在多家期刊编辑部使用,深受欢迎。本编辑部的编辑和审稿人都是统计专业人员,在处理一般统计资料时仍偏好该统计系统。其最大的特点是使用方便,不像 SAS、SPSS 等专业统计工具那样复杂,只要了解一般统计学知识就能掌握,特别适合编辑部审稿人员使用,

解决稿件中常见的统计问题。该统计工具的使用能有效减少论文中的统计学错误,从而提高期刊质量。

1 基于 Excel 的统计分析系统的基本原理

根据统计学分析的原理和公式^[4],利用 Excel 提供的强大函数功能^[5-6],在 Excel 工作表上编辑相应的公式或函数,对指定的“数据输入区”数据进行运算,通过反复的逻辑判断和运算,将最终结果输出到“结果显示区”,隐藏中间过程,锁定“数据输入区”以外的区域。只要“数据输入区”存在数据,“结果显示区”立刻显示结果,包括错误提示^[7-8]。

2 基于 Excel 的统计分析系统包含的内容

该统计系统包含常见资料类型的统计分析,有 4 大类共计 14 种主要统计方法,每种主要统计方法又由多个相关统计方法构成。

1) 数值变量资料的统计分析:数值变量资料的统计描述、样本均数与总体均数比较的 t 检验、配对资料比较的 t 检验、2 个样本均数的比较、多组资料比较的方差分析(自动运行多个样本均数间两两比较的 q 检验)。

2) 分类变量资料的统计分析:率的检验(样本率与总体率的比较、2 个样本率的比较)、四格表资料的卡方检验、四格表配对资料的卡方检验、行 \times 列表资料的卡方检验。

3) 秩和检验:配对资料的符号秩和检验、2 样本比较的秩和检验、多个样本比较的秩和检验、等级资料比较的秩和检验(自动运行多个样本间两两比较的秩和检验)。

4) 直线相关与回归分析。

3 应用举例

3.1 数值变量资料的统计描述

例 1 现有 130 名 14 岁女孩身高数据,欲了解数据分布情况。

操作 在统计分析系统目录窗口选择“数值变量资料的统计描述”,根据提示输入或粘贴数据,立刻显示出:均数、中位数、几何均数;标准差、方差、全距、变异系数;四分位数及四分位数间距;偏度系数及其标准误、峰度系数及其标准误,正态性检验。如图 1,结果

* 2011 年度全国统计科学研究计划项目(2011LY080)

显示例数 130,均数 143.216 9,按 $\alpha = 0.1$ 水准,资料 为 正 态 分 布。



图 1 数值变量资料的统计描述

3.2 2 个样本均数的比较 (用均数和标准差)

例 2 利用论文中提供的均数和标准差审核检验结果。

操作 在系统目录窗口选择“2 个样本均数的比较(用均数和标准差)”,根据提示输入 2 组样本的例

数、均数、标准差,即可显示组间比较的检验结果。系统根据样本量的大小、正态性检验和方差齐性检验情况,自动选择 t 检验、 t' 检验。如图 2,本例方差不齐,显示 t' 检验结果,单侧、双侧 P 值均 > 0.05 ,按 $\alpha = 0.05$ 水准,差别无统计学意义。



图 2 2 个样本均数的比较 (用均数和标准差)

3.3 2个样本均数的比较(用原始数据)

例3 2种饲料喂养大鼠的体质量增长情况,见表1,问2种饲料喂养大鼠后体质量的增加有无差别?

表1 用2种不同蛋白质质量分数的饲料喂养大鼠后体质量增加情况 g

高蛋白组	134	146	104	119	124	161	107	83	113	129	97	123
低蛋白组	70	118	101	85	107	132	94					

操作 在目录窗口选择“2个样本均数的比较(用原始数据)”,根据提示输入或粘帖2组原始数据,即可显示出组间比较的结果:2组的样本例数、均数、标准差;方差齐性检验;正态性检验;自动选择 t 检验或 t' 检验。如图3,本例为正态分布,方差齐, $t = 1.891$, $P > 0.05$ (双侧),按 $\alpha = 0.05$ 水准,可以认为2种饲料喂养大鼠后体质量的增加无差别。

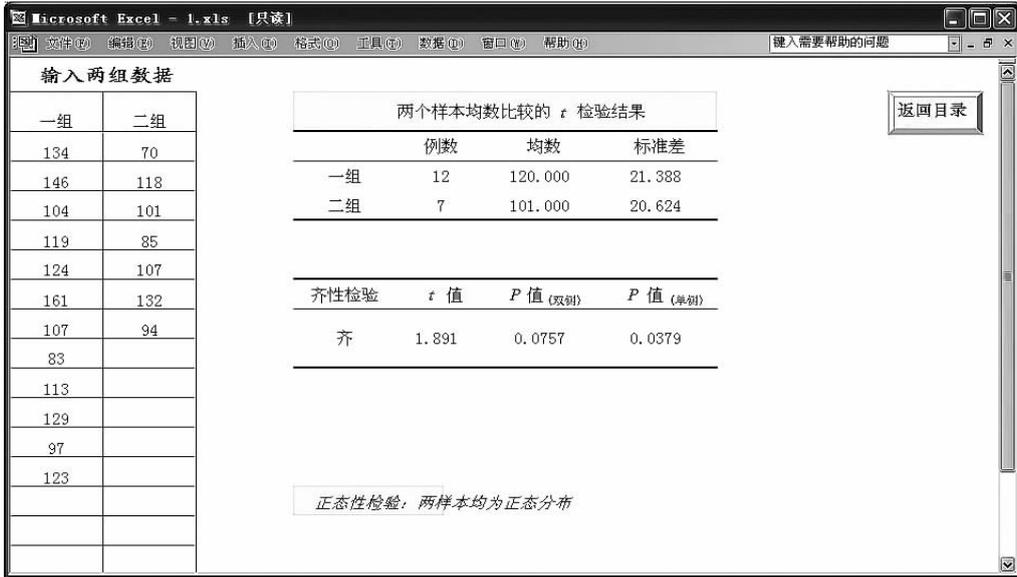


图3 2个样本均数的比较(用原始数据)

3.4 2样本比较的秩和检验

例4 某医师为研究血铁蛋白与肺炎的关系,随机抽查肺炎患者和正常人若干名,测得血铁蛋白($\mu\text{g/L}$)。问肺炎患者与正常人平均血铁蛋白水平有无差别?

表2 测得的肺炎患者和正常人血铁蛋白 $\mu\text{g/L}$

肺炎患者	31	68	237	174	457	492	199	515	599	238						
正常人	177	172	34	47	132	54	47	52	47	294	68	43	277	44	43	95

操作 对于本资料应首先选用“2个样本均数的比较”,输入数据后,提示资料为非正态分布建议秩和检验;返回目录窗口选择“2样本比较的秩和检验”,根据提示将表2中2组数据输入相应位置,即可显示统计分析结果:Wilcoxon 2样本比较的秩和检验统计量 $T = 183.5$ 、 $P = 0.011$ 。按检验水准 $\alpha = 0.05$,差别有统计学意义,即肺炎患者(一组)血铁蛋白平均水平高于正常人(二组),见图4。

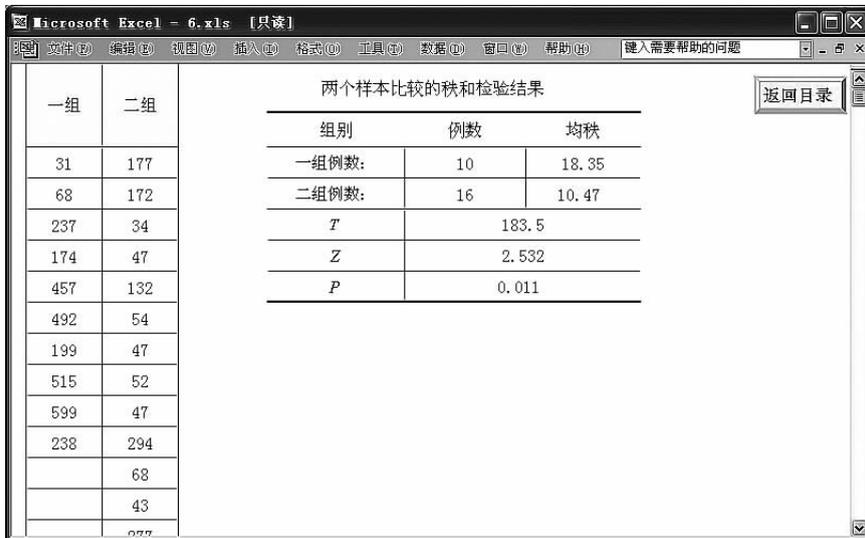


图4 2个样本比较的秩和检验