

如何正确运用 Z 检验——两总体率比较 等效性 Z 检验及 SAS 实现

胡纯严¹, 胡良平^{1,2*}

(1. 军事科学院研究生院, 北京 100850;

2. 世界中医药学会联合会临床科研统计学专业委员会, 北京 100029

*通信作者: 胡良平, E-mail: lphu927@163.com)

【摘要】 本文介绍了两总体率比较等效性 Z 检验的原理、方法和 SAS 实现。文中涉及两个方面内容, 包括两总体率比较等效性 Z 检验的原理和方法以及基于 SAS 软件实现两总体率比较等效性 Z 检验。文中给出了 SAS 输出结果并对其进行解释, 并分别得出统计学结论和专业结论。

【关键词】 总体率; 正态分布; 等效性检验; 等效性界值

中图分类号: R195.1

文献标识码: A

doi: 10.11886/scjsws20201119002

How to use Z test correctly——comparison of two population rates for the equivalence Z test and the SAS implementation

Hu Chunyan¹, Hu Liangping^{1,2*}

(1. Graduate School, Academy of Military Sciences PLA China, Beijing 100850, China;

2. Specialty Committee of Clinical Scientific Research Statistics of World Federation of Chinese Medicine Societies, Beijing 100029, China

*Corresponding author: Hu Liangping, E-mail: lphu927@163.com)

【Abstract】 The paper introduced the principles, methods and SAS implementation of the comparison of two population rates by using the equivalence Z test. Two special titles were involved as follows: one was to introduce the principles and methods of the equivalence Z test for the comparison of two population rates, the other was to implement the equivalence Z test for the comparison of two population rates based on the SAS software. The SAS output results were presented and explained, and also the statistical conclusions as well as the professional conclusions were given in this paper.

【Keywords】 Population rate; Normal distribution; Equivalence test; Equivalence dividing value

在对临床资料进行统计分析时, 研究者常选择的统计分析方法是差异性检验。其中, 一般差异性检验^[1-3]仅考虑差异是否具有统计学意义, 而 3 种特殊差异性检验, 即非劣效性检验、等效性检验和优效性检验^[4]除了考虑差异应具有统计学意义外, 还需要考虑应具有“临床意义”。本文将结合临床实际, 介绍与“两总体率比较的等效性检验”有关的内容。

1 两总体率比较等效性 Z 检验的基本原理和方法

1.1 基本概念

等效性检验是指主要研究目的为显示两种治疗效果之间的差别大小在临床上并无重要意义的试验。在试验设计阶段需要设定等效性界值(δ_L, δ_U)来界定两种药物或两种治疗方案的等效性^[4-5]。

1.2 问题与数据结构

【例 1】 一个新的抗肿瘤药物 A 与临床有效药物 B 对照进行临床试验, 选取 300 例符合要求的患者并随机均分成两组, 分别接受 A 药和 B 药治疗。试验结果为 A 药的有效率是 58.00% (有效例数 $n=87$), B 药的有效率是 46.00% (有效例数 $n=69$)。现根据临床实际, 设置等效性界限为 10%, 试问两种药物是否等效^[6]? 临床资料见表 1。

表 1 两组患者的治疗效果

药物种类	例数(n)	有效率(%)
A	150	58.00
B	150	46.00

【对数据结构的分析】 该资料的试验因素为“药物种类”, 其两个水平分别为“药物 A”与“药物 B”, 观测指标为“有效率”, 所以该资料为成组设计一元定性资料。

【统计分析方法的选择】 研究者的研究目的是

评价两种药物抗肿瘤效果是否等效,并且设定了等效性界值 $\delta_L = -10\%$ 、 $\delta_U = 10\%$,这时应选用等效性检验。

【例2】在伊曲康唑治疗手足癣的平行对照临床试验中,335例患者入选,166例患者服用伊曲康唑400 mg/d,服药1周(治愈92例),169例患者服用伊曲康唑100 mg/d,连续服药4周(治愈98例)。设定等效性界值为12%,试问两种治疗方案是否等效?临床资料见表2。

表2 两组患者的真菌感染治愈率

治疗方案	例数(n)	治愈率(%)
A	166	55.42
B	169	57.99

注:本例作为练习,请读者完成

1.3 基本原理

1.3.1 检验假设

规定整个等效性检验的检验水准 α (每个单侧检验的检验水准为 $\alpha' = \alpha/2$)的值。

$H_{0(1)}: \pi_T - \pi_R \leq \delta_L, H_{1(1)}: \pi_T - \pi_R > \delta_L, \alpha' = 0.025$ (单侧)

$H_{0(2)}: \pi_T - \pi_R \geq \delta_U, H_{1(2)}: \pi_T - \pi_R < \delta_U, \alpha' = 0.025$ (单侧)

1.3.2 检验统计量

$$Z_1 = \frac{P_T - P_R - \delta_L}{\sqrt{\frac{P_T(1-P_T)}{n_T} + \frac{P_R(1-P_R)}{n_R}}} \sim N(0, 1) \quad (1)$$

拒绝域为 $Z_1 > Z_{1-\alpha/2}$

$$Z_2 = \frac{P_T - P_R - \delta_U}{\sqrt{\frac{P_T(1-P_T)}{n_T} + \frac{P_R(1-P_R)}{n_R}}} \sim N(0, 1) \quad (2)$$

拒绝域为 $Z_2 < Z_{\alpha/2}$

在式(1)和式(2)中,检验统计量 Z 服从标准正态分布; (P_T, n_T) 与 (P_R, n_R) 分别为试验组(T)与对照组(R)的“样本率、样本含量”; δ_L 与 δ_U 分别为等效性界值的下限(下限取负值)与上限。

1.3.3 基于检验统计量 Z 的取值计算单侧尾端概率

根据检验水准查表或借助相应的SAS函数进行计算,获得检验统计量 Z 的临界值(针对手工计算而言)和(或)标准正态分布尾端的概率(针对统计软件而言)。

1.3.4 得出统计学结论和专业结论

根据拒绝还是不能拒绝 H_0 的结果,得出统计学

结论,最后结合专业知识得出专业结论。需注意,须同时拒绝 $H_{0(1)}$ 和 $H_{0(2)}$,方能得出等效性结论。

2 两总体率比较等效性检验的SAS实现

2.1 基于SAS语言按公式编程间接实现计算

基于SAS语言和标准正态分布函数等知识^[7]并按式(1)和式(2)进行编程,就可间接实现两总体率比较等效性 Z 检验。

SAS程序如下:

```
%let n_t=150;
%let n_r=150;
%let p_t=0.580;
%let p_r=0.460;
%let delta_L=-0.10;
%let delta_U=0.10;
data a;
A1=(%p_t-%p_r)-&delta_L;
B1=(%p_t-%p_r)-&delta_U;
A2=%p_t*(1-%p_t)/%n_t;
A3=%p_r*(1-%p_r)/%n_r;
Z1=A1/SQRT(A2+A3);
Z2=B1/SQRT(A2+A3);
p1=1-probnorm(Z1);
p2=probnorm(Z2);
run;
proc print noobs;
var Z1 p1 Z2 p2;
run;
```

【程序说明】变量名后面的“_t”与“_r”分别代表“试验组”与“对照组”相应的量(包括样本含量与有效率)。等效性下界值应采用负值来表示。

【SAS输出结果】

Z1	P1	Z2	P2
3.84137	0.000061176	0.34922	0.63654

【统计学结论】因 $Z_1=3.841, P=0.000061$ (单侧概率) <0.025 ,所以,应拒绝 $H_{0(1)}$,接受 $H_{1(1)}$,又因 $Z_2=0.349, P=0.63654$ (单侧概率) >0.025 ,所以,应接受 $H_{0(2)}$ 。

【专业结论】在等效界值为 $(-0.10, 0.10)$ 的前提下,药物A与药物B不等效。

2.2 基于SAS中FREQ过程直接实现计算

由于SAS中FREQ过程包含了几乎绝大多数定

性资料假设检验方法^[8],所以,此处将采用此过程来直接实现两总体率比较等效性 Z 检验。

SAS 程序如下:

```
data a;
do a=1 to 2;
do b=1 to 2;
input f @@;
output;
end;
end;
cards;
87 63
69 81
;
run;
proc freq data=a;
tables a*b/riskdiff(equiv MARGIN=0.10);
weight f;
run;
```

【程序说明】各组中的数据应为“有效例数”与“无效例数”,而不适合直接采用“有效率”。程序中 equiv 表示进行等效性检验, MARGIN=0.10 表示等效界值(即下界值为-0.10、上界值为0.10),若下界值与上界值的绝对值不相等,需要明确写出,例如: MARGIN=(-0.12, 0.15)(注意:下界值必须采用负数)。

【SAS 主要输出结果及解释】

第 1 部分输出结果:

两个单侧检验 (TOST)

检验	Z	P-Value
边际下限	3.8414	Pr>Z <0.0001
边际上限	0.3492	Pr<Z 0.6365
全部		0.6365

第 2 部分输出结果:

等效性限制	90% 置信区间
-0.1000 0.1000	0.0258 0.2142

【统计学结论】由第 1 部分输出结果可知:因 $Z_1=3.8414, P<0.0001$ (单侧概率) <0.025 ,所以,应拒绝 $H_{0(1)}$,接受 $H_{1(1)}$ 。又因 $Z_2=0.3492, P=0.6365$ (单侧概率) >0.025 ,所以,应接受 $H_{0(2)}$ 。由第 2 部分输出结果可知:两总体率之差的 90% 置信区间为 $[0.0258, 0.2142]$,该区间包含了等效性上界值 $\delta_U=0.10$,说明与上界值对应的等效性检验未通过,即不等效。

【专业结论】在等效界值为(-0.10, 0.10)的前提下,药物 A 与药物 B 不等效。

3 讨论与小结

3.1 影响等效性检验结果的因素

等效性检验结果不是绝对的,随着条件的改变,最终的检验结果也会发生相应的改变。具体地说,影响其结果的因素有“显著性水平 α ”“等效性界值(δ_L, δ_U)”和“样本含量”。

3.1.1 α 的取值对等效性检验结果的影响

等效性检验结果与显著性水平 α 有密切关系。现以例 2 的数据为例,若取每个单侧检验的显著性水平 $\alpha=0.025$,等效界值取为(-0.12, 0.12),计算结果为 $Z_1=1.7428, P=0.0407>\alpha=0.025$; $Z_2=-2.6911, P=0.0037<\alpha=0.025$,结论为两药不等效。若取每个单侧检验的显著性水平 $\alpha=0.05$,由前面的计算结果(两个 P 值都小于 0.05)可知,结论为两药等效。

3.1.2 等效性界值(δ_L, δ_U)的取值对等效性检验结果的影响

等效性检验结果与等效性界值(δ_L, δ_U)的取值也有密切关系。现仍以例 2 的数据为例,若取每个单侧检验的显著性水平 $\alpha=0.025$,等效界值取为(-0.12, 0.12),计算结果为 $Z_1=1.7428, P=0.0407>\alpha=0.025$; $Z_2=-2.6911, P=0.0037<\alpha=0.025$,结论为两药不等效。若取每个单侧检验的显著性水平 $\alpha=0.025$,等效界值取为(-0.14, 0.14),计算结果为 $Z_1=2.1123, P=0.0173<\alpha=0.025$; $Z_2=-3.0606, P=0.0011<\alpha=0.025$,结论为两药等效。

3.1.3 样本含量对等效性检验结果的影响

若保持“3.1.3”中“两药不等效”情况的大部分条件不变,仅改变样本含量,例如,两组样本含量都扩大 10 倍,即 A 药组有效例数为 920 例,无效例数为 740 例; B 药组有效例数为 980 例,无效例数为 710 例。计算结果为 $Z_1=5.5113, P<0.0001$; $Z_2=-8.5100, P<0.0001$,结论就变为“两药等效”。即使样本含量扩大 2 倍,也能得出“两药等效”的结论。需注意的是,在拟开展等效性试验研究之初期,就应当按照已知的条件[包括两组估计的有效率、显著性水平 α 的值、检验效能 $(1-\beta)$ 的值和等效性界值(δ_L, δ_U)的值],找到相应的计算公式,估计出合适的样本含量^[4]。

3.2 小结

本文介绍了两总体率比较等效性 Z 检验的原理、方法和 SAS 实现。在 SAS 实现方面,介绍了基于 SAS 语言编程间接实现两总体率比较等效性 Z 检验以及基于 SAS 中 FREQ 过程直接实现与前面提及的相同计算。在后者中,还给出了两总体率之差的置信区间。基于置信区间也可以判断等效性结论是否成立。

参考文献

- [1] 葛祥丽,李慧敬,崔明湖,等.艾司西酞普兰联合奥氮平治疗抑郁症的效果及安全性的 Meta 分析[J].四川精神卫生,2019,32(6):523-530.
- [2] 孙桂芝,崔明湖,都萌萌,等.阿戈美拉汀治疗首发抑郁障碍效果及安全性的 Meta 分析[J].四川精神卫生,2020,33(3):257-262.
- [3] 许国生,车鹏,胡俊,等.氨磺必利和利培酮治疗老年精神分裂症的效果和安全性 Meta 分析[J].四川精神卫生,2020,33(4):342-348.
- [4] 胡良平,陶丽新.临床试验设计与统计分析[M].北京:军事医学科学出版社,2013:71-74,88-100,111-118.
- [5] 张甜甜,刘媛媛,李长平,等.如何正确运用 t 检验——两算术均值比较等效性 t 检验及 SAS 实现[J].四川精神卫生,2020,33(4):302-306.
- [6] 胡良平.科研设计与统计分析[M].北京:军事医学科学出版社,2012:337-338.
- [7] 胡良平,胡纯严.SAS 语言基础与高级编程技术[M].北京:电子工业出版社,2014:37-97.
- [8] SAS Institute Inc. SAS/STAT®15.1 user's guide[M]. Cary, NC: SAS Institute Inc, 2018: 2997-3216.
- [9] 胡良平.现代医学统计学[M].北京:科学出版社,2020:244-257.

(收稿日期:2020-11-19)

(本文编辑:戴浩然)



科研方法专题策划人——胡良平教授简介

胡良平,男,1955年8月出生,教授,博士生导师,曾任军事医学科学院研究生部医学统计学教研室主任和生物医学统计学咨询中心主任、国际一般系统论研究会中国分会概率统计系统专业理事会常务理事、中国生物医学统计学会副会长、北京大学口腔医学院客座教授和《中华医学杂志》等10余种杂志编委;现任世界中医药学会联合会临床科研统计学专业委员会会长、国家食品药品监督管理局评审专家和3种医学杂志编委;主编统计学专著48部、参编统计学专著10部;发表第一作者和通信作者学术论文300余篇、发表合作论文130余篇;获军

队科技成果和省部级科技成果多项;参加并完成三项国家标准的撰写工作、参加三项国家科技重大专项课题研究工作。在从事统计学工作的30年中,为几千名研究生、医学科研人员、临床医生和杂志编辑讲授生物医学统计学,在全国各地作统计学学术报告100余场,举办数十期全国统计学培训班,培养20多名统计学专业硕士和博士研究生。近几年来,参加国家级新药和医疗器械项目评审数十项、参加100多项全军重大重点课题的统计学检查工作。归纳并提炼出有利于透过现象看本质的“八性”和“八思维”的统计学思想,独创了逆向统计学教学法和三型理论。擅长于科研课题的研究设计、复杂科研资料的统计分析和 SAS 与 R 软件实现、各种层次的统计学教学培训和咨询工作。