

# 如何正确运用 Z 检验——定性资料一致性 Z 检验及 SAS 实现

胡纯严<sup>1</sup>, 胡良平<sup>1,2\*</sup>

(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】** 本文目的是介绍定性资料一致性 Z 检验的原理、方法和 SAS 实现。文中涉及两个方面的内容, 包括介绍定性资料一致性 Z 检验的原理和方法以及基于 SAS 软件实现定性资料一致性 Z 检验。文中给出了 SAS 输出结果并对其进行解释, 再分别得出统计学结论和专业结论。

**【关键词】** 定性资料; 列联表; kappa 系数; 一致性检验; Z 检验

中图分类号: R195.1

文献标识码: A

doi: 10.11886/scjsws20201119004

## How to use Z test correctly——consistent Z test of the qualitative data and the SAS implementation

Hu Chunyan<sup>1</sup>, Hu Liangping<sup>1,2\*</sup>

(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 purpose of this paper was to introduce the principles, methods and SAS implementation of the consistent Z test for the qualitative data. Two special titles were involved as follows: one was to introduce the principles and methods of consistent Z test for the qualitative data, the other was to implement consistent Z test of the qualitative data 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】** Qualitative data; Contingency tables; Kappa coefficient; Consistent test; Z test

在临床资料中, 定性资料占有相当大的比重。根据不同的试验设计类型和统计分析目的, 研究者常将定性资料整理成各式各样的列联表<sup>[1-2]</sup>, 其中, 双向有序且属性相同的方形列联表常被用于呈现两位临床医生(或两种测定方法)对同一批某病患者的样本检测的结果。这种列联表也被称为配对设计扩大形式的方表, 对其进行统计分析的目的是判定两位临床医生诊断结果之间是否具有一致性, 需要采用的统计分析方法被称为一致性检验。本文将基于 SAS 软件对此种列联表资料进行统计分析, 对输出结果给出合理的解释并作出统计学结论和专业结论。

### 1 一致性 Z 检验的基本原理和方法

#### 1.1 基本概念

##### 1.1.1 双向有序且属性相同的方形列联表

##### 1.1.1.1 配对设计 2×2 方形列联表

表 1 中列出了采用两种方法同时检测 68 位某

病患者的标本所得的结果。见表 1。

表 1 两种检测方法对同一组受试者检测的结果

试验法 检测结果	例数		合计
	*, +	-	
+	31	4	35
-	3	30	33
合计	34	34	68

注: \*代表金标准检测结果; +代表阳性; -代表阴性

##### 1.1.1.2 配对设计扩大形式 3×3 方形列联表

表 2 呈现了在两个不同时间段上对 446 例流行性出血热病患者的病情转化情况进行记录的结果; 而表 3 呈现了两名医生对 200 例棉屑沉着病可疑患者的诊断结果。见表 2、表 3。

##### 1.1.1.3 配对设计扩大形式 4×4 方形列联表

表 4 呈现了两种测定方法同时检测 73 个样品中的抗体所得的检测结果。见表 4。

表2 446例流行性出血热病情转化情况

早期分型 判断结果	例数			合计
	*:轻型	中型	重型	
轻度	111	21	1	133
中度	5	163	20	188
重度	0	1	124	125
合计	116	185	145	446

注:\*代表最后定型判断结果

表3 甲、乙两名医生对200例棉屑沉着病可疑患者的诊断结果

甲医生的 诊断结果	例数			合计
	*:正常	I期	II期	
正常	78	5	0	83
I期	6	56	13	75
II期	0	10	32	42
合计	84	71	45	200

注:\*代表乙医生的诊断结果

表4 两种测定方法同时检测73个样品中的抗体所得的检测结果

快速法检测结果	样品数				合计
	*:-	+	++	+++	
-	15	0	2	3	20
+	2	19	1	2	24
++	1	3	17	0	21
+++	0	2	0	6	8
合计	18	24	20	11	73

注:\*代表ELISA法检测结果;检测结果分为4档,抗体的浓度由低到高分别用-、+、++和+++表示

### 1.1.2 统计分析方法的合理选择

除了行数和列数不相同之外,以上四个表格在本质上是完全相同的,即都是两种检测方法或两位评价者的测定结果按“配对的方法”整理出来的。对于按这种形式呈现的定性资料,虽然有多种不同的分析目的(例如考察两者之间的“一致性”“对称性”和“相关性”),但在实际工作中,研究者最关注的是两者的“一致性”是否具有统计学意义和实际意义。因此,最常用的统计分析方法是一致性检验。

### 1.1.3 检验方法名称的辨析

在对以上四个表格中的定性资料进行一致性检验时,统计学教科书上常有以下几种说法:即“一致性检验”“kappa检验”和“Z检验(或U检验)”。事实上,它们指的是同一种检验,即计算公式和原理是完全相同的。仔细分析可知,“一致性检验”是目的,“kappa检验”是突出计算公式分子中有一个名叫kappa的一般统计量(它是用来度量两种检测方法检测结果之间呈现一致性程度高低的一个系数),而“Z检验”则是指实现该检验的检验统计量的

名称为Z。在SAS软件中,一般规定检验统计量Z为服从标准正态分布的随机变量,通常记为 $Z \sim N(0, 1)$ 。然而,从检验统计量的公式中是否包含“临床实际意义的界值”来划分,所有的假设检验均可被划分成两大类,一类为不包含临床实际意义的界值的一般差异性检验,例如文献[3-6];另一类为包含临床实际意义的界值的特殊差异性检验,例如非劣效性检验、等效性检验和优效性检验<sup>[7-9]</sup>。

## 1.2 一致性Z检验的基本原理和方法

### 1.2.1 检验假设

给出检验假设,并规定检验水准 $\alpha$ 的值。

$H_0$ : 总体kappa系数=0(两种方法或者两位医生的检查结果不一致); $H_1$ : 总体kappa系数 $\neq 0$ (两种方法或者两位医生的检查结果一致); $\alpha = 0.05$ 。

【说明】kappa系数是度量一致性程度高低的一个数量,包括“简单kappa系数”与“加权kappa系数”两种。

### 1.2.2 与简单kappa系数有关的内容

#### 1.2.2.1 简单kappa系数的计算公式

Kappa(简写成k)系数的计算公式为<sup>[10-12]</sup>:

$$k = \frac{P_a - P_e}{1 - P_e} \quad (1)$$

在式(1)中, $P_a$ 和 $P_e$ 分别代表“实际一致率”与“期望一致率”,其计算公式分别见式(2)和式(3):

$$P_a = \frac{1}{n} \sum_{i=1}^m f_{ii} \quad (2)$$

$$P_e = \frac{1}{n^2} \sum_{i=1}^m R_i C_i \quad (3)$$

在式(2)和式(3)中,“m”代表方形列联表的“行数”或“列数”;“n”代表方形列联表中的总频数; $f_{ii}$ 代表方形列联表中第i行第i列上的频数(即从左上角到右下角对角线上的频数); $R_i$ 与 $C_i$ 分别代表第i行与第i列上的频数之合计值。

#### 1.2.2.2 简单kappa系数的标准误 $S_k$

简单kappa系数的标准误 $S_k$ 可由下面的公式计算:

$$S_k = \frac{\sqrt{P_e + P_e^2 - \frac{1}{n^3} \sum R_i C_i (R_i + C_i)}}{(1 - P_e) \sqrt{n}} \quad (4)$$

### 1.2.2.3 基于简单 kappa 系数的检验统计量 $Z$

由式(1)和式(4)可以构造出检验统计量  $Z$ , 见下式:

$$Z = \frac{k - 0}{S_k} \sim N(0, 1) \quad (5)$$

拒绝域为  $Z > Z_{1-\alpha/2}^\circ$

### 1.2.2.4 基于简单 kappa 系数的置信区间

还可以构造出简单 kappa (简写成  $k$ ) 系数的  $100(1-\alpha)\%$  置信区间, 见下式:

$$\left(k - Z_{1-\alpha/2} S_k, k + Z_{1-\alpha/2} S_k\right) \quad (6)$$

### 1.2.3 与加权 kappa 系数有关的内容

#### 1.2.3.1 概述

简单 kappa 系数在计算时仅与方形列联表的主对角线上的频数有关, 但有时“非主对角线上的频数”对“一致性的估计”也能提供一些“辅助信息”, 将其纳入 kappa 系数的计算之中, 所得到的结果被

称为“加权 kappa 系数”<sup>[10-12]</sup>。在 SAS 的 FREQ 过程中, 引入了两种计算 kappa 系数“权重”的定义, 分别被称为“Cicchetti-Allison 权重 (简称 CA 权重)”和“Fleiss-Cohen 权重 (简称 FC 权重)”<sup>[12]</sup>。

#### 1.2.3.2 加权 kappa 系数

$$\hat{k}_w = \frac{P_{o(w)} - P_{e(w)}}{1 - P_{e(w)}} \quad (7)$$

在式(7)中,  $P_{o(w)}$ 、 $P_{e(w)}$  分别为观察的与期望的加权一致率, 其计算公式见式(8)与式(9):

$$P_{o(w)} = \sum_i \sum_j w_{ij} p_{ij} \quad (8)$$

$$P_{e(w)} = \sum_i \sum_j w_{ij} p_i \cdot p_j \quad (9)$$

在式(8)和(9)中,  $w_{ij}$ 、 $p_{ij}$  分别为方形表中第  $i$  行第  $j$  列上的“权重”与“比例”;  $p_i$  与  $p_j$  分别为方形表中“第  $i$  行上的比例”与“第  $j$  列上的比例”。

#### 1.2.3.3 加权 kappa 系数的方差

加权 kappa 系数的方差见下式:

$$\text{Var}(\hat{k}_w) = \frac{\sum_i \sum_j p_{ij} \left[ w_{ij} - (\bar{w}_i + \bar{w}_j) (1 - \hat{k}_w) \right]^2 - \left[ \hat{k}_w - P_{e(w)} (1 - \hat{k}_w) \right]^2}{[1 - P_{e(w)}]^2 n} \quad (10)$$

在式(10)中,  $\bar{w}_i$  与  $\bar{w}_j$  分别见式(11)与式(12):

$$\bar{w}_i = \sum_j p_j w_{ij} \quad (11)$$

$$\bar{w}_j = \sum_i p_i w_{ij} \quad (12)$$

在式(15)中,  $\text{Var}_0(\hat{k}_w)$  为总体 kappa 系数为 0 时样本 kappa (即  $\hat{k}_w$ ) 的方差, 其计算公式见下式:

$$\text{Var}_0(\hat{k}_w) = \frac{\sum_i \sum_j p_i \cdot p_j \left[ w_{ij} - (\bar{w}_i + \bar{w}_j) \right]^2 - P_{e(w)}^2}{[1 - P_{e(w)}]^2 n} \quad (16)$$

#### 1.2.3.4 加权 kappa 系数的置信区间

总体加权 kappa 系数的  $100(1-\alpha)\%$  置信区间见下式:

$$\left[ \hat{k}_w - Z_{(1-\alpha/2)} \times \sqrt{\text{Var}(\hat{k}_w)}, \hat{k}_w + Z_{(1-\alpha/2)} \times \sqrt{\text{Var}(\hat{k}_w)} \right] \quad (13)$$

【说明】在文献[12]中, 误将  $Z_{(1-\alpha/2)}$  写成了  $Z_{\alpha/2}^\circ$

#### 1.2.3.5 基于加权 kappa 系数的 $Z$ 检验统计量

假定总体 kappa 系数为“ $k_{w(o)} \neq 0$ ”, 则:

$$Z = \frac{\hat{k}_w - k_{w(o)}}{\sqrt{\text{Var}(\hat{k}_w)}} \sim N(0, 1) \quad (14)$$

假定总体 kappa 系数为“ $k_{w(o)} = 0$ ”, 则:

$$Z = \frac{\hat{k}_w}{\sqrt{\text{Var}_0(\hat{k}_w)}} \sim N(0, 1) \quad (15)$$

#### 1.2.3.6 权重系数的计算公式

CA 权重系数的计算公式见下式:

$$w_{ij} = 1 - \frac{|C_i - C_j|}{C_m - C_1} \quad (17)$$

FC 权重系数的计算公式见下式:

$$w_{ij} = 1 - \frac{(C_i - C_j)^2}{(C_m - C_1)^2} \quad (18)$$

在式(17)和式(18)中, 当  $i = j$  时,  $w_{ij} = 1$ ; 当  $i \neq j$ , 且  $i < j$  时,  $C_i$ 、 $C_j$  分别代表第  $i$  列、第  $j$  列上的“得分值”(通常方形表中行、列方向上是相同的有序变量, 分别代表甲、乙两个评价者给出的评价结果, 例如当分 4 档时, 可分为 0、2、4、6 分, 也分为 0、3、7、12 分, 相邻档次之间不要求间隔相等, 可依据专业

知识进行赋值);“ $m$ ”代表方形表的“行数”或“列数”(例如当分 4 档时,  $C_m = C_4$ , 即第 4 档上的“得分值”)。

#### 1.2.4 计算检验统计量 $Z$ 的数值及其对应的尾端概率

计算检验统计量  $Z$  的数值,按照检验水准查  $Z$  值表获得概率分布尾端的概率(对手工计算而言)或由统计软件直接算出标准正态分布曲线下尾端的概率;或者可以求出总体 Kappa 系数的 95% 置信区间。

#### 1.2.5 得出统计学结论和专业结论

根据拒绝还是接受  $H_0$  的结果,得出统计学结论,再结合专业知识给出专业结论。

## 2 一致性 $Z$ 检验的 SAS 实现

现以分析表 4 资料为例,呈现 SAS 程序及结果。

SAS 程序如下:

```
DATA a;
DO A=1 TO 4;
DO B=1 TO 4;
INPUT F @@;
OUTPUT;
END;
END;
CARDS;
15 0 2 3
2 19 1 2
1 3 17 0
0 2 0 6
;
RUN;
PROC FREQ data=a;
WEIGHT F;
TABLES A*B;
TEST KAPPA WTKAP;
RUN;
```

【程序说明】“TEST 语句”中有两个选项,第 1 个选项“KAPPA”要求计算简单 kappa 系数、95% 置信区间并对其进行假设检验;第 2 个选项“WTKAP”要求计算加权 kappa 系数、95% 置信区间并对其进行假设检验。

## 【SAS 主要输出结果及解释】

第 1 部分输出结果:

### 简单 Kappa 系数

Kappa	0.6994
ASE	0.0658
95% 置信下限	0.5704
95% 置信上限	0.8283

由输出结果可知,简单 kappa=0.6994,其渐近标准误  $S_k=0.0658$ ,其 95% 置信区间为 [0.5704, 0.8283]。由于此置信区间未包含“0”,说明总体 kappa 值不等于 0,即两种检测方法检测结果之间的一致性具有统计学意义。

### H0 的检验: Kappa=0

H0 下的 ASE	0.0699
Z	10.0044
单侧 Pr>Z	<0.0001
双侧 Pr> Z	<0.0001

以上是对简单 kappa 进行假设检验的结果,  $Z=10.0044$ ,无论是单侧检验还是双侧检验都得出  $P<0.0001$ ,说明总体 kappa 值不等于 0。

第 2 部分输出结果:

### 加权 Kappa 系数

加权 Kappa	0.6379
ASE	0.0844
95% 置信下限	0.4724
95% 置信上限	0.8034

### H0 的检验: 加权 Kappa=0

H0 下的 ASE	0.0811
Z	7.8683
单侧 Pr>Z	<0.0001
双侧 Pr> Z	<0.0001

第 2 部分输出结果是关于加权 kappa 的计算结果,除加权 kappa 系数、加权检验统计量  $Z$  值等较前面未加权的结果略有变化外,置信区间和假设检验的结果没有发生根本性的改变,结论与基于简单 kappa 计算所得的结论相同,不再赘述。

【结论】快速法与 ELISA 法检测结果之间的一致性具有统计学意义,至于这种一致性是否具有实用价值,取决于研究者的专业需求。若以 ELISA 法检测结果为“金标准”来判断可否采用快速法取代 ELISA 法,关键取决于现在的样本一致率 78.08% 是否满足专业



要求,若满足,则可以取代;否则,就不可以取代。

【说明】样本一致率 78.08% 是按如下方法计算出来的,即采用表 4 中主对角线上的频数之和“57”除以总频数 73 而得。

### 3 讨论与小结

#### 3.1 讨论

如何合理选用“简单 kappa 系数及其有关分析方法”与“加权 kappa 系数及其有关分析方法”,文献[11]认为:加权 kappa 系数法适用于“多级分组的两次检查结果的定性资料”。基于常识可知,若被评价对象的“得分”是客观存在的(例如某病患者的样品的检测结果在本质上应属于某一档是客观真实的),仅由于评价者或测定方法无法准确认定其“得分”高低时,宜选用“简单 kappa 法”(即在方形列联表中,只需要利用主对角线上频数所提供的信息);当被评价对象的“得分”不是客观存在的(例如某声乐参赛者的得分在本质上并不存在一个准确的数量),而是受各评价者主观判断或偏好的影响,此时,宜选用“加权 kappa 法”,即有必要参考“邻近位置上的得分”(即在方形列联表中,不单看主对角线上的频数,也要适当利用非主对角线上频数提供的信息,进行综合考虑)。除采用 SAS 实现一致性检验、非劣效性检验、等效性检验和优效性检验外,还有其他一些统计软件也可以实现前述提及的各种检验,如 R 软件<sup>[13]</sup>。

#### 3.2 小结

本文介绍了一致性 Z 检验的原理、方法和 SAS 实现。在计算原理和方法方面,介绍了“基于简单 kappa 系数法”和“基于加权 kappa 系数法”两方面的内容;在 SAS 实现方面,仅介绍了基于 SAS 中 FREQ 过程直接实现一致性 Z 检验,同时,还给出了总体 kappa 系数的置信区间。

### 参考文献

- [1] 胡良平. 统计学三型理论在统计表达与描述中的应用[M]. 北京: 人民军医出版社, 2008: 153-164.
- [2] 胡良平, 王琪. 定性资料统计分析及应用[M]. 北京: 电子工业出版社, 2016: 1-68.
- [3] 张洪璐, 刘媛媛, 李长平, 等. 正确运用  $t$  检验——两算术均值比较一般差异性  $t$  检验及 SAS 实现[J]. 四川精神卫生, 2020, 33(3): 217-221.
- [4] 于泽洋, 刘媛媛, 李长平, 等. 正确运用  $t$  检验——两几何均值比较一般差异性  $t$  检验及 SAS 实现[J]. 四川精神卫生, 2020, 33(3): 222-225.
- [5] 黄慧杰, 刘媛媛, 李长平, 等. 正确运用  $t$  检验——线性回归模型中参数与 0 比较  $t$  检验及 SAS 实现[J]. 四川精神卫生, 2020, 33(4): 307-311.
- [6] 宋德胜, 刘媛媛, 李长平, 等. 正确运用  $t$  检验——相关系数与 0 比较  $t$  检验及 SAS 实现[J]. 四川精神卫生, 2020, 33(4): 312-316.
- [7] 陈阳, 刘媛媛, 李长平, 等. 正确运用  $t$  检验——两算术均值比较非劣效性  $t$  检验及 SAS 实现[J]. 四川精神卫生, 2020, 33(3): 226-230.
- [8] 姚婷婷, 李长平, 胡良平, 等. 正确运用  $t$  检验——两算术均值比较优效性  $t$  检验及 SAS 实现[J]. 四川精神卫生, 2020, 33(4): 297-301.
- [9] 张甜甜, 刘媛媛, 李长平, 等. 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): 297-301.
- [10] 倪宗瓚. 医学统计学[M]. 北京: 人民卫生出版社, 1990: 255-260.
- [11] 唐军. 中国医学统计百科全书单变量推断统计分册[M]. 北京: 人民卫生出版社, 2004: 44-45.
- [12] SAS Institute Inc. SAS/STAT®15.1 user's guide[M]. Cary, NC: SAS Institute Inc, 2018: 2997-3216.
- [13] 胡良平. 现代医学统计学[M]. 北京: 科学出版社, 2020: 271-272.

(收稿日期:2020-11-19)

(本文编辑:戴浩然)