如何正确运用Z检验——两率比较一般差异性Z检验及SAS实现

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】 本文介绍了用Z检验进行率比较的原理、方法和SAS实现。包括一个未知总体率与一个已知总体率比较的Z检验和两未知总体率比较的Z检验。在SAS实现方面，采取了两种途径：①基于计算公式，用SAS语言编程；②直接借助SAS中的“FREQ过程”。

【关键词】 总体率；正态分布；二项分布；卡方分布；Z检验；精确检验
中图分类号：R195.1 文献标识码：A doi：10.11886/scjsws20200916006

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

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 the rates by using the *Z* test, including *Z* test for the comparison of an unknown population rate with a known population rate and *Z* test for the comparison of two unknown population rates. In the aspect of SAS implementation, two measures were taken as below: the first was writing SAS program by using SAS language based on the calculation formula, the second was directly applying the procedure FREQ in SAS software.

【**Keywords**】 Population rate; Normal distribution; Binomial distribution; χ^2 distribution; *Z* test; Exact test

涉及一个未知总体率与一个已知总体率比较时，需要借助二项分布原理进行计算，也可以采用正态近似法来实现；而涉及两个总体率比较时，其统计分析方法有χ^2检验（包括未校正的、校正的、似然比、连续校正的χ^2检验）、Fisher's精确检验以及基于标准正态分布的Z检验。本文将着重介绍基于标准正态分布的方法，同时，也给出用其他类似统计分析方法计算的结果。

1 一个未知总体率与一个已知总体率比较的Z检验

1.1 基本概念

在某些统计学教科书上，把“一个未知总体率与一个已知总体率比较”表述为“样本率与总体率比较”，这种简化的表述存在欠妥之处。因为“样本”与“总体”是不对等的两个事物，故它们之间是没有可比性的。类似地，“样本均数与总体均数比较”“两样本率比较”“多个样本率比较”“两样本标准化率比较”等表述也是欠妥的。虽然“A样本率”与“B样本率”是对等的两个事物，但通常两个“样本率”在数量上一般都是不等的，故无需再做比较。统计学的价值就在于基于“样本所提供的信息去推论所讨论的问题在总体中的规律”，就是从两个“样本率”出发来推断它们各自所代表的“总体率”之间的数量关系。其目的是进行“两总体率比较”，而非“两样本率比较”。

1.2 问题与数据结构

【例1】文献[1]的目的是探讨新冠肺炎（COVID-19）疫情期间居家儿童青少年焦虑症状检出率及影响因素，为给予其心理支持提供参考。采用电子问卷调查方式，共收回有效问卷5 392份，其中焦虑组1 045人（19.4%），非焦虑组4 347人（80.6%）。

由此可知，COVID-19 疫情期间居家隔离儿童青少年焦虑症状检出率为 19.4%。同时，文献[1]还援引了其他文献报道的同类检出率大约为 22.0%~36.9%。现以后者的平均值 29.45% 为“已知总体率”，试问：

问题 1：文献[1]的总体率（未知）与已知总体率之间的差别是否具有统计学意义？

问题 2：文献[1]的总体率（未知）是否明显低于已知总体率？

【统计分析方法的选择】此问题属于“一个未知总体率与一个已知总体率比较的问题”，可以运用二项分布原理进行计算，也可以采取正态近似法进行计算。对“问题 1”而言，属于“双侧检验”问题；而对“问题 2”而言，则属于“下单侧检验”问题。

1.3 假设检验方法

1.3.1 基于二项分布原理的假设检验方法

由于对每位受试者来说，其调查结果都是“二值变量（是否出现焦虑）”的一种取值，要么是“出现焦虑”，要么是“未出现焦虑”，这是一个“两点分布”问题。多个两点分布叠加起来就形成了一个“二项分布”。此分布可用作研究“率或比”比较的最直接方法[2~4]，但其计算原理比较深奥，因篇幅所限，此处从略。

1.3.2 基于标准正态分布原理的假设检验方法

1.3.2.1 检验统计量

由于满足一定条件的二项分布的计算问题可以采用正态分布来近似计算，而标准正态分布应用广泛，且其检验统计量形式简单、计算方便。对于“一个未知总体率与一个已知总体率比较”的检验统计量[2~4]分别见式（1）和式（2）：

$$Z=\frac{|X-nP_0|}{\sqrt{nP_0(1-P_0)}}=\frac{|P-P_0|}{\sqrt{P_0(1-P_0)/n}},\quad P_0\text{为已知总体率} \tag{1}$$

$$Z=\frac{|X-nP_0|-0.5}{\sqrt{nP_0(1-P_0)}}=\frac{|P-P_0|-0.5}{\sqrt{P_0(1-P_0)/n}},\quad P_0\text{为已知总体率} \tag{2}$$

式（1）和式（2）中的 Z 服从标准正态分布，式（1）属于未校正的公式，而式（2）属于校正公式。其中，X 为样本阳性数、n 为样本含量、P 为样本率、P_0 为已知总体率，而式（2）分子上的“0.5”为连续性校正数，当 $|X-nP_0|\leqslant 0.5$ 时不适合进行校正。

1.3.2.2 前提条件

当 P_0 很小时，可基于 Poisson 分布原理进行检验；当 P_0 不太靠近 0 或 1 时，可基于二项分布原理进行检验；而当样本含量 n 足够大时，可基于标准正态分布原理进行检验。事实上，当 $nP_0\geqslant 5$ 且 $n(1-P_0)\geqslant 5$ 时，用标准正态分布取代二项分布进行计算，其误差极小。

1.4 SAS 实现

SAS 程序如下：

```
/*Z检验*/
%let P0=0.2945;
%let n=5392;
%let X=1045;
/*Z1、P1是按未校正公式计算的*/
/*Z2、P2是按校正公式计算的*/
data a1;
P=&X/&n;
D=abs(P-&P0);
Z1=D/sqrt(&P0*(1-&P0)/&n);
Z2=(D-0.5)/sqrt(&P0*(1-&P0)/&n);
if D<=0.5 then Z2=Z1;
P1=2*(1-probnorm(Z1));
P2=2*(1-probnorm(Z2));
proc print data=a1;
var Z1 P1 Z2 P2;
run;
/*基于二项分布原理的假设检验*/
data a2;
input group count;
cards;
1 1045
2 4347
;
run;
proc freq data=a2;
weight count;
exact binomial/pformat=exact point;
tables group / binomial(p=0.2945 exact);
run;
```

【程序说明】第 2 个“注释语句”之前为“基于标准正态分布进行二项分布的近似计算”，而该语句之后为“基于二项分布进行精确计算，同时，也包括基于标准正态分布进行二项分布的近似计算”。前者是基于公式手工编程计算的，而后者是直接使用 SAS 中的“FREQ 过程”并添加一些“选项”来完成的。其中，“exact 语句”的作用是实现精确检验；而“tables 语句”中“选项”的作用是估计精确置信区间。

【SAS 主要输出结果及解释】

第 1 部分（基于公式编程计算）SAS 程序的输出结果如下：

Obs	Z1	P1	Z2	P2
1	16.2214	0	16.2214	0

因 $Z_1=Z_2$，说明此资料不符合进行校正计算的条件；$P<0.0001$，说明文献[1]的调查样本所来自的总体的焦虑检出率与已知的总体率 29.45% 之间的差别具有统计学意义。因样本率 19.40% 小于已知的总体率 29.45%，说明文献[1]的调查样本所来自的总体的焦虑检出率明显偏低。

第 2 部分（基于 FREQ 过程计算）SAS 程序的输出结果如下：

H_0 检验：比例=0.2945

H_0 下的 ASE	0.0062
Z	−16.2214
单侧 Pr<Z	<0.0001
双侧 Pr>|Z|	<0.0001

精确检验	
单侧 Pr≤P	3.378E−64
点 Pr=P	1.436E−64
双侧=2*单侧	6.757E−64

样本大小=5392

以上是关于“未知总体率是否等于已知总体率”的两种假设检验结果，“H_0 检验：比例=0.2945”部分是基于标准正态分布近似法得到的结果，$Z=-16.2214$（注：SAS 软件在计算时，未取绝对值），无论是基于单侧检验还是双侧检验，都得到 $P<0.0001$ 的结果，即未知总体率不等于已知总体率，结合具体的样本率数据可知，未知总体率小于已知总体率。“精确检验”部分是基于二项分布进行精确计算得到的结果，单侧检验的概率和双侧检验的概率都极小，说明样本（注：样本率为 19.38%）所对应的总体率（未知）小于已知总体率 29.45%。

【结论】文献[1]中 COVID-19 疫情期间居家隔离儿童青少年总体焦虑症状检出率（未知）低于其他文献报导的相应检出率（29.45%）。

2 两未知总体率比较的 Z 检验

2.1 问题与数据结构

【例 2】文献[1]中 COVID-19 疫情期间居家隔离儿童青少年焦虑症状男性检出率为 16.81%，女性检出率为 22.45%。试问：

问题 1：两性别焦虑症状总体检出率之间的差别是否具有统计学意义？

问题 2：女性焦虑症状总体检出率是否一定高于男性？

【统计分析方法的选择】此问题属于“两未知总体率比较的问题”，可以运用二项分布原理进行计算，也可以采取正态近似法进行计算。对“问题 1”而言，属于“双侧检验”问题；而对“问题 2”而言，则属于“上单侧检验”问题。

2.2 基于标准正态分布原理的假设检验方法

2.2.1 检验统计量

依据两率差的标准误计算方法和是否进行校正，对于“两个未知总体率比较”的检验统计量有 4 个计算公式，分别见式(3)~式(6)：

情形一，按各自率求率差的标准误且不校正，见式(3)：

$$Z_1=\frac{|P_1-P_2|}{\sqrt{\frac{P_1(1-P_1)}{n_1}+\frac{P_2(1-P_2)}{n_2}}} \tag{3}$$

情形二，按各自率求率差的标准误且要校正，见式(4)：

$$Z_2=\frac{|\frac{X_1-0.5}{n_1}-\frac{X_2+0.5}{n_2}|}{\sqrt{\frac{P_1(1-P_1)}{n_1}+\frac{P_2(1-P_2)}{n_2}}} \tag{4}$$

情形三，按平均率求率差的标准误且不校正，见式(5)：

$$Z_3=\frac{|P_1-P_2|}{\sqrt{PC(1-PC)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}},$$

$$PC=\frac{X_1+X_2}{n_1+n_2} \tag{5}$$

情形四，按平均率求率差的标准误且要校正，见式(6)：

$$Z_4 = \frac{\left|\frac{X_1 - 0.5}{n_1} - \frac{X_2 + 0.5}{n_2}\right|}{\sqrt{PC(1 - PC)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}},$$

$$PC = \frac{X_1 + X_2}{n_1 + n_2} \quad (6)$$

以上各式中的检验统计量(Z_1~Z_4)均服从标准正态分布；PC为两样本率的平均率；0.5为连续性校正数。

2.2.2 Z检验统计量与χ^2检验统计量之间的关系

由于“两率之间的一般差异性检验问题”可以转化为“非配对设计四格表资料的独立性检验问题”，后者可以采用χ^2检验。事实上，由χ^2分布的定义可知，当自由度为1时，$Z^2 = \chi^2$。

2.3 SAS实现

SAS程序如下：

由于SAS中FREQ过程的计算结果中包含了“基于标准正态分布近似法计算”和“基于二项分布原理计算”两种结果，又因篇幅所限，故第1部分“基于公式编程实现正态近似法的SAS程序”及其输出结果均从略。

```
/*第1部分：基于公式编程的Z检验SAS程序从略*/
/*第2部分：卡方检验，加上选项RISKDIFF(EQUAL)*/
data data1;
do a=1 TO 2;
do b=1 TO 2;
input f @@;
output;
end;
end;
cards;
494 2444
551 1903
;
run;
proc freq;
weight f;
tables a*b / chisq riskdiff(equal);
run;
proc freq data=a2;
weight f;
tables a*b / chisq riskdiff(equal correct);
run;
```

【程序说明】第2部分SAS程序较简洁，第1个过程步的关键在于在“tables语句”中使用了两个选项，即“chisq”和“riskdiff(equal)”。前者进行卡方检验，后者进行正态近似计算；而第2个过程步的选项中增加了“correct”，即求置信区间时进行“校正”，连续性校正的WALD置信限按下面的公式计算(基于四格表而言)[3]：

$$Est \pm \left[Z_{\alpha/2} \times se(Est) + cc\right] \quad (7)$$

在式(7)中，cc为校正数，对于第1行风险率，$cc=1/2n_1$；对于第2行风险率，$cc=1/2n_2$；对于风险率差，$cc=(1/n_1+1/n_2)/2$。对于第1列和第2列风险率，使用与前面类似的校正数。n_1和n_2分别为第1行与第2行的合计频数。

【SAS主要输出结果及解释】

因篇幅所限，第1部分输出结果从略。

由第1部分输出结果(未显示)可知，无论“是否使用平均率”和“是否进行校正”，Z检验统计量的数值相差无几，双侧概率$P<0.0001$，说明男性与女性的焦虑症状总体阳性率不等，结合样本率数据可知，女性的总体焦虑率高于男性的总体焦虑率。

第2部分未使用校正的输出结果：

首先输出的是卡方检验结果，$\chi^2=27.2127$，$P<0.0001$，说明男、女总体焦虑检出率不等。接着输出Fisher's精确检验结果，双侧概率$P<0.0001$，说明男、女总体焦虑检出率不等。此处省略了“与总体率置信区间估计”有关的输出结果和结果解释。

比例(风险)差值检验

$H_0: P_1-P_2=0$

比例差值	−0.0564
渐近标准误差(样本)	0.0109
Z	−5.1788
单侧 Pr<Z	<0.0001
双侧 Pr>|Z|	<0.0001

最后输出的是基于标准正态分布近似法进行两未知总体率比较的结果，$Z=-5.1788$(注意：未取绝对值)，双侧概率$P<0.0001$。

第2部分使用校正算法的输出结果(因篇幅所限，这部分输出结果从略)。

3 讨论与小结

3.1 讨论

可用于两率比较的统计分析方法很多，其中最精确的方法是基于二项分布原理推导出的计算方法和 Fisher's 精确检验法，其次是卡方检验和 Z 检验两种方法。然而，这些统计分析方法之间还是存在区别的。若仅关注“一般差异性检验”，则“基于二项分布原理推导出的计算方法”是最合理的选择；若考虑其他检验类型（例如非劣效性检验、优效性检验或等效性检验），适合选择“Z 检验”。

本文涉及“正态分布”“二项分布”“卡方分布”和“超几何分布（Fisher's 精确检验法的理论依据）”等概率分布知识，需要了解这方面知识的读者，可参阅文献[5-6]。因篇幅所限，这些内容此处从略。

3.2 小结

本文结合两个实例，介绍了“一个未知总体率与一个已知总体率比较的 Z 检验”和“两未知总体率比较的 Z 检验”的基本原理和 SAS 实现。同时，还给出了“基于二项分布原理”“卡方分布原理”和“超几何分布原理”实现统计计算的 SAS 程序和输出结果。由本文的做法和输出结果可知：相对于“基于计算公式，用 SAS 语言编程”来说，直接使用 SAS 中的 FREQ 过程来完成率的比较，不仅更加简洁和方便，而且还能提供更多的算法和相应的输出结果。

参考文献

[1] 莫大明，闫军伟，李欣，等. 新冠肺炎疫情下儿童青少年焦虑症状检出率及影响因素[J]. 四川精神卫生，2020，33(3)：202-206.

[2] 杨树勤. 中国医学百科全书 医学统计学[M]. 上海：上海科学技术出版社，1985：92-94.

[3] SAS Institute Inc. SAS/STAT®9.3 user's guide[M]. Cary，NC：SAS Institute Inc，2011：2270-2437.

[4] 胡良平. 现代医学统计学[M]. 北京：科学出版社，2020：235-285.

[5] 方开泰，许建伦. 统计分布[M]. 北京：科学出版社，1987：62-211.

[6] 茆诗松. 统计手册[M]. 北京：科学出版社，2006：1-33.

（收稿日期：2020-09-16）

（本文编辑：戴浩然）