

# 如何正确运用 $\chi^2$ 检验——评分检验与 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】** 本文目的是介绍评分检验与 SAS 实现。具体地说, 评分检验统计量有 5 种常见的形式, 即一般评分检验统计量、稳健评分检验统计量、残差评分检验统计量、广义评分检验统计量和拉格朗日乘数检验统计量; 评分检验主要用于以下 3 种场合, 即检验部分或全部回归系数是否为 0、检验单个效应是否不在回归模型中以及检验平行线假定是否成立。本文通过两个实例, 借助 SAS 软件实现前述提及的各种评分检验, 并着重解释与评分检验有关的输出结果。

**【关键词】** 评分检验; 稳健评分检验; 残差评分检验; 拉格朗日乘数检验;  $\chi^2$  分布

中图分类号: R195.1

文献标识码: A

doi: 10.11886/scjsws20211110007

## How to use $\chi^2$ test correctly——the score test and the implementation of SAS software

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 article was to introduce the score test and the SAS implementation. Concretely, there were five common forms of the score test statistics, namely the general score test statistic, the robust score test statistic, the residual score test statistic, the generalized score test statistic, and the Lagrange multiplier test statistic. The score test was mainly used in the following three situations, namely, testing whether part or all of the regression coefficients were 0, testing whether a single effect was not in the regression model, and testing whether the parallel line hypothesis was established. The paper used two examples and the SAS software to realize the various score tests mentioned above, and focused on explaining the output results related to the score tests.

**【Keywords】** Score test; Robust score test; Residual score test; Lagrange multiplier test;  $\chi^2$  distribution

在构建广义线性回归模型的过程中, 常涉及参数的检验问题, 例如: 检验部分或全部回归系数是否为 0, 检验单个效应是否不在回归模型中以及检验平行线假定是否成立。在统计学上, 实现前述检验的具体方法有评分检验、似然比检验和 Wald 检验。本文仅介绍评分检验及其 SAS 实现。

## 1 基本概念

### 1.1 评分函数

设资料由观测值  $x_1, x_2, \dots, x_N$  组成, 并假定可用依赖于参数  $\theta$  的概率模型  $f(x; \theta)$  来描述该资料。当变量  $x$  只有有限个可能的取值代入函数  $f(x; \theta)$ , 从而指定每一个输出的概率; 当存在无穷多个输出  $f(x; \theta)$  来指定概率密度函数时, 对数似然函数可表示如下:

$$l(\theta) = \sum_{i=1}^N \log f(x_i; \theta) \quad (1)$$

此时, 评分函数<sup>[1]</sup>表示如下:

$$u(\theta) = l'(\theta) = \sum_{i=1}^N \frac{f'(x_i; \theta)}{f(x_i; \theta)} \quad (2)$$

在式(2)中, “'”代表求函数的导数; 若  $\hat{\theta}$  是  $\theta$  的最可能值, 则满足  $u(\hat{\theta})=0$ 。

### 1.2 评分检验统计量的定义

设  $\theta^0$  是  $\theta$  的真值, 则评分就具有均值为 0、方差为  $i(\theta^0)$ , 则下式定义的检验统计量  $Z$  服从标准正态分布:

$$Z = \frac{u(\theta^0) - 0}{[i(\theta^0)]^{1/2}} \sim N(0, 1) \quad (3)$$

基于式(3)定义的检验被称为评分检验<sup>[1]</sup>。文献[1]还给出了 Wald 检验统计量, 见下式:

$$\chi^2 = Z^2 = \frac{(\hat{\theta} - \theta^0)^2}{j^{-1}} \sim \chi^2_1 \quad (4)$$

在式(4)中,  $\hat{\theta}$  是  $\theta$  的最大似然估计值,  $\theta^0$  是  $\theta$  的真值,  $j^{-1}$  是  $\theta$  的方差。

### 1.3 评分检验统计量的用途

在评分检验统计量中,参数 $\theta$ 可以是一维的,也可以是多维的(被称为参数向量)。在统计学上,经常将3种检验统计量(即评分检验统计量、似然比检验统计量和Wald检验统计量)<sup>[1-3]</sup>同时呈现出来,其主要用途如下:检验部分或全部回归系数是否为0、检验单个效应是否不在回归模型中、检验平行线假定是否成立以及检验线性假设是否成立等。

## 2 评分检验统计量的种类

### 2.1 一般评分检验统计量

设回归模型中的回归系数向量 $\beta$ 具有 $K$ 个分量,在SAS/STAT的PHREG过程中,给出的5个检验统计量(即似然比检验统计量、一般评分检验统计量、Wald's检验统计量、稳健评分检验统计量和稳健Wald's检验统计量)都服从自由度 $df=K$ 的 $\chi^2$ 分布。这5种检验都可以用于检验回归模型中全部回归系数是否都等于0,即 $H_0:\beta=0$ 。其中,一般评分检验统计量的定义有以下两种表达形式,见式(5)、式(6)<sup>[2-3]</sup>:

$$\chi^2_s = \left[ \frac{\partial l(\theta)}{\partial \beta} \right]' \left[ -\frac{\partial^2 l(\theta)}{\partial \beta^2} \right]^{-1} \left[ \frac{\partial l(\theta)}{\partial \beta} \right] \quad (5)$$

$$R_n = [S_n(\tilde{\theta})]' [I_n(\tilde{\theta})]^{-1} S_n(\tilde{\theta}) \quad (6)$$

式(6)中, $S_n(\theta) = \partial \log l(\theta) / \partial \theta$ 是评分函数, $\tilde{\theta}$ 是 $H_0$ 成立条件下 $\theta$ 的最大似然估计值或似然方程的根; $H_0:R(\theta)=0, \theta$ 为参数。

### 2.2 稳健评分检验统计量

在SAS/STAT的PHREG过程中,稳健评分检验统计量的定义<sup>[2]</sup>如下:

$$\chi^2_{rs} = \left[ \sum_i L_i^0 \right]' \left[ \sum_i L_i^0 L_i^{0'} \right]^{-1} \left[ \sum_i L_i^0 \right] \quad (7)$$

式(7)中, $L_i^0$ 是第 $i$ 个受试者在 $\beta=0$ 处的评分残差;也就是说, $L_i^0 = L_i(0, \infty)$ ,得分过程 $L_i(\beta, t)$ 的定义如下:

$$L_i(\beta, t) = \int_0^t [Z_i(s) - \bar{Z}(\beta, s)] dM_i(\beta, s) \quad (8)$$

向量 $\hat{L}_i \equiv L_i(\hat{\beta}, \infty)$ 是第 $i$ 个受试者的得分残差。

【说明】在SAS/STAT的PHREG过程中,介绍了4种残差,分别是边际残差、偏差残差、Schoenfeld残差和得分残差。上式中涉及其他一些变量或符号,参见文献[2],此处从略。

### 2.3 残差评分检验统计量

#### 2.3.1 检验部分或全部回归系数是否为0

在累积和广义(或扩展)的logistic回归模型中,假定因变量有 $K(i=1, 2, \dots, K)$ 个类别,共同的自变量有 $S$ 个。也就是说,在每一类中拟合logistic回归模型时,都涉及一个截距项和 $S$ 个自变量(注:指尚未筛选掉某些自变量)。为了检验部分回归系数是否为0,可以采用残差评分 $\chi^2$ 检验。

让 $g(\beta)$ 代表对数似然函数关于参数向量 $\beta$ 的一阶偏导数向量;让 $H(\beta)$ 代表对数似然函数关于参数向量 $\beta$ 的二阶偏导数矩阵。也就是说, $g(\beta)$ 是梯度向量,而 $H(\beta)$ 是哈森矩阵。让 $I(\beta)$ 代表“ $-H(\beta)$ ”的期望。考虑一个无效假设 $H_0$ ,让 $\hat{\beta}_{H_0}$ 是无效假设条件下 $\beta$ 的最大似然估计值(MLE)。在SAS/STAT的LOGISTIC过程中,残差评分检验统计量的定义如下:

$$\chi^2_{rs} = g'(\hat{\beta}_{H_0}) I^{-1}(\hat{\beta}_{H_0}) g(\hat{\beta}_{H_0}) \quad (9)$$

式(9)中的 $\chi^2_{rs}$ 渐近服从自由度为 $df$ 的 $\chi^2$ 分布。其中,参数向量 $\beta$ 、无效假设 $H_0$ 与 $H_0$ 对应的向量 $\beta$ 和自由度 $df$ 都会随着因变量的具体情形不同而有所变化。具体地说,有以下两种情形。

情形一:累积反应变量模型或称多值有序因变量模型。

参数向量 $\beta = (\alpha_1, \alpha_2, \dots, \alpha_K, \beta_1, \beta_2, \dots, \beta_S)'$ ;无效假设 $H_0: \beta_{t+1} = \beta_{t+2} = \dots = \beta_S, t < S$ ;与 $H_0$ 对应的向量 $\beta: \hat{\beta}_{H_0} = (\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_K, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_t, 0, 0, \dots, 0)'$ ;自由度: $df = S - t$ 。

情形二:广义反应变量模型或称多值名义因变量模型(当 $K=2$ 时为二值反应变量)。

参数向量 $\beta = (\alpha_1, \alpha_2, \dots, \alpha_K, \beta'_1, \beta'_2, \dots, \beta'_S)'$ ;其中, $\beta'_i = (\beta_{i1}, \beta_{i2}, \dots, \beta_{iS})$ ;  $i=1, 2, \dots, K$ ;无效假设 $H_0: \beta_{i,t+1} = \beta_{i,t+2} = \dots = \beta_{i,S}, t < S; i=1, 2, \dots, K$ ;与 $H_0$ 对应的向量 $\beta: \hat{\beta}_{H_0} = (\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_K, \hat{\beta}'_{1(t)}, 0'_{(S-t)}, \dots, \hat{\beta}'_{K(t)}, 0'_{(S-t)})'$ ,其中, $0'_{(S-t)}$ 代表由 $S-t$ 个0组成的向量;自由度: $df = K(S-t)$ 。

#### 2.3.2 检验单个效应是否不在回归模型中

解决问题的基本思路:假定有两组自变量,A组中有 $S$ 个自变量(可构建一个全模型);B组中有 $S-1$ 个自变量(可构建一个简化模型);B组中的全部自变量都包含在A组中。将A组比B组多出来的一个自变量记为 $x_t [t \in (1, 2, \dots, S)]$ ,于是,将A组与B组

自变量及其资料分别代入式(4)计算,便可得到两个 $\chi^2$ 值,它们具有不同的自由度。求出这两个 $\chi^2$ 值之差,再基于它们的自由度之差,就可依据 $\chi^2$ 分布求出尾端概率。

自由度的确定方法:若是多值有序因变量,全模型与简化模型的自由度之差等于1;若是多值名义因变量,全模型与简化模型的自由度之差等于 $K$ (当 $K=2$ 时为二值因变量的情形)。当检验结果为 $P<0.05$ ,表明不在B组中的自变量 $x_t[t \in (1, 2, \dots, S)]$ 具有统计学意义;反之亦然。

### 2.3.3 检验平行线假定是否成立

设因变量 $Y$ 为多值有序的且水平数为 $K+1$ (即水平数大于2),让因变量 $Y$ 的取值分别为 $1, 2, \dots, K, K+1$ 。假定有 $S$ 个自变量,在没有平行线假设的条件下,一般的累积logistic回归模型可用下式表达:

$$g[P_r(Y \leq ilx)] = (1, x')\beta_i, 1 \leq i \leq K \tag{10}$$

在式(10)中, $g(\cdot)$ 是连接函数; $\beta_i = (\alpha_i, \beta_{i1}, \beta_{i2}, \dots, \beta_{is})'$ 是 $Y=i$ 条件下的一个未知参数向量,它的分量分别是一个截距、 $S$ 个斜率。这个一般累积模型的参数向量可表示成: $\beta = (\beta'_1, \beta'_2, \dots, \beta'_k)'$ ,它是由 $K$ 个分量组成的一个向量,而每个分量就是前面所呈现的向量[即 $\beta_i = (\alpha_i, \beta_{i1}, \beta_{i2}, \dots, \beta_{is})'$ ]。

给定平行线的无效假设: $H_0: \beta_{1m} = \beta_{2m} = \dots = \beta_{km}, 1 \leq m \leq S$ 。回归系数 $\beta$ 有两个下标,第1个下标代表因变量 $Y$ 的取值或水平代码,第2个下标代表自变量的顺序编号。若用语言描述上述无效假设 $H_0$ ,即在因变量 $Y$ 的 $K$ 个水平条件下的 $K$ 个logistic回归模型中编号相同的自变量具有相等的回归系数。也就是说, $K$ 个logistic回归模型仅截距不同,自变量线性组合中只有一组共同的回归系数。

用统计学语言表述如下:让 $\beta_1, \beta_2, \dots, \beta_s$ 代表共同斜率系数,让 $\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_k$ 和 $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_s$ 分别代表截距参数与斜率参数的最大似然估计值。那么,在 $H_0$ 成立的条件下, $\beta$ 的最大似然估计见式(11):

$$\hat{\beta}_{H_0} = (\beta'_1, \beta'_2, \dots, \beta'_k)', \text{其中 } \hat{\beta}_i = (\hat{\alpha}_i, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_s)', 1 \leq i \leq K \tag{11}$$

此时,与无效假设 $H_0$ 对应的残差评分 $\chi^2$ 检验统计量见式(12):

$$\chi^2_{RS} = g'(\hat{\beta}_{H_0})I^{-1}(\hat{\beta}_{H_0})g(\hat{\beta}_{H_0}) \sim \chi^2_{df}, df = S(K-1) \tag{12}$$

【说明】对平行线假定的检验可视为对参数有线性约束的检验(简称约束检验<sup>[6]</sup>)的特例,后者常被冠以拉格朗日乘数检验(参见下文)。

### 2.4 广义评分检验统计量

让 $\tilde{\beta}$ 为在限制条件 $L'\beta=0$ 下从广义估计方程中求解得到的回归参数;让 $S(\tilde{\beta})$ 为在 $\tilde{\beta}$ 时广义估计方程的值。Boos和Rotnitzky与Jewell描述了在广义估计方程中将评分检验应用于检验 $L'\beta=0$ ,其中 $L'$ 是用户指定的 $r \times p$ 阶比较矩阵。于是,广义评分检验统计量<sup>[2]</sup>见下式:

$$T = S(\tilde{\beta})' \sum_m L(L' \sum_c L)^{-1} L' \sum_m S(\tilde{\beta}) \tag{13}$$

在式(13)中, $\sum_m$ 是基于模型的协方差估计值, $\sum_c$ 是经验协方差估计值, $T$ 服从自由度 $df=r$ 的 $\chi^2$ 分布。

### 2.5 拉格朗日乘数(或约束评分)检验统计量

在SAS/STAT的GENMOD过程中的定义如下:当用户给回归模型中的参数施加某些限制(例如:NOINT,即强制性要求截距为0;又例如:NOSCALE,即强制性要求某些定量随机变量的概率分布中的尺度参数取固定值)时,需要有相应的检验方法来评价这些限制的有效性。这个检验方法被称为拉格朗日乘数检验或约束评分检验。检验统计量<sup>[2]</sup>见式(14):

$$\chi^2_{LM} = \frac{S^2}{V} \sim \chi^2_1 \tag{14}$$

式(14)中, $S$ 是在与被限制参数对应的限制的最大值点上估计的评分向量的成分;而 $V = I_{11} - I_{12}I_{22}^{-1}I_{21}$ ,其中矩阵 $I$ 是信息矩阵,下标中的1指被限制的参数,下标中的2指未被限制的其他参数。

【说明】因篇幅所限,在SAS/ETS的ENTROPY过程中的定义、在SAS/ETS的MDC、MODEL和QLIM过程中的定义以及在SAS/ETS的PANEL过程中的定义从略。

## 3 实例与SAS实现

### 3.1 问题与数据

【例1】一项关于成年人心理健康问题的研究,调查了某城市40例成年居民的心理健康状况(1=无心理问题,2=存在轻微症状,3=存在中度症状,4=存在精神损害),考察的两个原因变量分别是社会经济地位(0=低,1=高)和生活事件评分,资料见表1<sup>[4]</sup>。试分析心理健康状况与两个原因变量的关系。

表1 成年人心理健康状况调查资料

编 号	心理健康状况	社会经济地位	生活事件评分
1	1	1	1
2	1	1	9
...	...	...	...
39	4	0	8
40	4	0	9

注:详细数据见下文 SAS 程序

【例2】随机纳入 30 例在某医院就诊的仅有胸闷症状的非器质性心脏病患者,收集患者在 24 小时中的早搏数(y),试问早搏是否与吸烟(x<sub>1</sub>)、性别(x<sub>2</sub>)和喝咖啡(x<sub>3</sub>)有关。资料见表 2<sup>[5]</sup>。

表2 某医院伴有胸闷症状的非器质性心脏病患者情况

编 号	x <sub>1</sub>	x <sub>2</sub>	x <sub>3</sub>	y
1	0	1	1	11
2	0	0	0	7
...	...	...	...	...
29	1	1	0	9
30	0	0	1	5

注:x<sub>1</sub>=1 表示吸烟,x<sub>1</sub>=0 表示不吸烟;x<sub>2</sub>=1 表示喜欢喝咖啡,x<sub>2</sub>=0 表示不喜欢喝咖啡;x<sub>3</sub>=1 表示男性,x<sub>3</sub>=0 表示女性;详细数据见下文 SAS 程序

### 3.2 分析与解答

#### 3.2.1 对例 1 的分析与解答

【分析与解答】设所需要的 SAS 程序如下:

```
data abc;
input id psych society life @@;
cards;
1 1 1 1 21 2 1 9
2 1 1 9 22 2 0 3
3 1 1 4 23 2 1 3
4 1 1 3 24 2 1 1
5 1 0 2 25 3 0 0
6 1 1 0 26 3 1 4
7 1 0 1 27 3 0 3
8 1 1 3 28 3 0 9
9 1 1 3 29 3 1 6
10 1 1 7 30 3 0 4
11 1 0 1 31 3 0 3
12 1 0 2 32 4 1 8
13 2 1 5 33 4 1 2
14 2 0 6 34 4 1 7
15 2 1 3 35 4 0 5
16 2 0 1 36 4 0 4
17 2 1 8 37 4 0 4
```

```
18 2 1 2 38 4 1 8
19 2 0 5 39 4 0 8
20 2 1 5 40 4 0 9
;
run;
proc logistic data=abc descending;
model psych=society life/selection=stepwise
sle=0.5 sls=0.05;
run;
proc logistic data=abc descending;
model psych=society life;
test1:test 0.5*society+1*life=0;
test2:test society=life;
run;
```

【SAS 输出结果及解释】

第 1 个过程步的主要输出结果:

比例优比假设的评分检验

卡方	自由度	Pr>卡方
0.8865	2	0.6419

以上是模型中含有截距项和自变量 life 时的比例优比假设的评分检验结果,也被称为平行线假设检验结果。因  $\chi^2_{RS}=0.8865, df=2, P=0.6419$ ,说明资料满足构建 logistic 回归模型的前提条件。

检验全局原假设: BETA=0

检验	卡方	自由度	Pr>卡方
似然比	6.5150	1	0.0107
评分	5.9514	1	0.0147
Wald	6.2871	1	0.0122

以上是采用 3 种检验方法检验此时的 logistic 回归模型(含截距项和一个自变量 life)是否成立,由于 3 种检验结果均有  $P<0.05$ ,说明此时的 logistic 回归模型成立。

逐步选择汇总

步	效应		自由度	个数	评分卡方	Wald 卡方	Pr>卡方
	已进入	已删除					
1	life		1	1	5.9514		0.0147
2	society		1	2	3.3672		0.0665
3		society	1	1		3.2719	0.0705

以上是逐步选择汇总,从第 2 列与倒数第 3 列可知,选择自变量进入回归模型时采用的是评分  $\chi^2$  检验;而从第 3 列与倒数第 2 列可知,删除自变量时采用的是 Wald  $\chi^2$  检验。

第 2 个过程步的主要输出结果:

比例优比假设的评分检验

卡方	自由度	Pr>卡方
2.3255	4	0.6761

以上输出的是平行线假定的检验结果,因  $P=0.6761>0.05$ ,说明资料满足构建 logistic 回归模型的前提条件。

检验全局原假设:  $BETA=0$

检验	卡方	自由度	Pr>卡方
似然比	9.9442	2	0.0069
评分	9.1431	2	0.0103
Wald	8.5018	2	0.0143

以上是采用 3 种检验方法检验此时的 logistic 回归模型[含截距项和两个自变量(life 和 society)]是否成立,由于 3 种检验结果均有  $P<0.05$ ,说明此时的 logistic 回归模型成立。

线性假设检验结果

标签	Wald 卡方	自由度	Pr>卡方
test1	0.6221	1	0.4303
test2	4.7700	1	0.0290

以上输出的是线性假设检验结果。第 1 行的结果说明  $0.5 \times \beta_1 + 1 \times \beta_2 = 0$  的假设成立;第 2 行的结果说明  $\beta_1 = \beta_2$  的假设不成立。这里采用的是 Wald  $\chi^2$  检验,而不是似然比  $\chi^2$  检验和拉格朗日  $\chi^2$  检验。

【说明】为节省篇幅,以上仅呈现了与评分检验有关的输出结果,故不适合给出统计结论与专业结论。

### 3.2.2 对例 2 的分析与解答

【分析与解答】设所需要的 SAS 程序如下:

```
DATA abc;
INPUT x1-x3 y @@;
CARDS;
0 1 1 11 0 0 0 7 0 0 0 3 1 0 1 5
0 0 0 2 1 1 1 13 0 1 0 6 1 0 1 10
0 0 0 4 1 0 1 7 0 0 0 1 0 0 1 9
0 0 1 6 1 1 1 17 0 0 0 5 1 0 0 11
0 1 1 8 1 0 1 9 0 0 0 8 1 0 0 5
0 1 1 5 1 1 0 8 1 1 0 13 0 0 1 8
1 0 0 6 0 0 1 4 0 0 0 6 1 1 1 13
1 1 0 9 0 0 1 5
;
```

```
RUN;
PROC GENMOD data=abc;
MODEL y=x1-x3/LINK=LOG DIST=NB noscale
aggregate;
RUN;
PROC GENMOD data=abc;
MODEL y=x1-x2/LINK=LOG DIST=NB noscale
aggregate;
RUN;
PROC GENMOD data=abc;
MODEL y=x1-x2/LINK=LOG DIST=POISSON;
RUN;
【SAS输出结果及解释】
```

第 1 个过程步的主要输出结果:

Lagrange 乘数统计量		
参数	卡方	Pr>卡方
离散度	8.6315	0.0017*
*单侧 P 值		

以上输出的是拉格朗日乘数检验结果,  $\chi^2_{LM}=8.6315, P=0.0017$ ,说明资料存在过度离散现象,不适合直接拟合 Poisson 分布回归模型。

第 2 个过程步的主要输出结果:

Lagrange 乘数统计量		
参数	卡方	Pr>卡方
离散度	0.9727	0.1620*
*单侧 P 值		

以上输出的是拉格朗日乘数检验结果,  $\chi^2_{LM}=0.9727, P=0.1620$ ,说明此时的资料(不包括自变量 x3)不存在过度离散现象,可直接拟合 Poisson 分布回归模型。

第 3 个过程步的输出结果与评分检验无关,为节省篇幅,此处从略。

【说明】为节省篇幅,以上仅呈现了与评分检验有关的输出结果,故不适合给出统计结论与专业结论。

## 4 讨论与小结

### 4.1 讨论

在进行广义线性模型<sup>[6]</sup>分析,特别是进行多重 logistic 回归模型<sup>[7-10]</sup>分析时,在多个环节上需要对参数进行假设检验,有时还需要在参数处于特定约束条件下对其进行检验。通常涉及 3 种检验,即评

分检验、似然比检验和 Wald 检验;当有约束条件时,其检验被称为约束检验<sup>[6]</sup>或拉格朗日乘数检验<sup>[2]</sup>,但它们也被称为评分检验<sup>[2]</sup>。当在某种场合下同时应用以上提及的多种检验方法时,检验结果可能不尽相同,但通常情况下不会出现截然相反的结果。事实上,在上述提及的几种检验统计量的推导过程中,一般都基于大样本条件下给出渐近的结果,而不是精确计算的结果。

#### 4.2 小结

本文介绍了与评分检验有关的基本概念,介绍了 5 种评分检验统计量,即一般评分检验统计量、稳健评分检验统计量、残差评分检验统计量、广义评分检验统计量和拉格朗日乘数(或约束评分)检验统计量。在介绍残差评分检验统计量这部分内容时,详细呈现了在构建广义线性模型过程中的 3 种适用场合;在介绍拉格朗日乘数(或约束评分)检验统计量这部分内容时,详细呈现了此种检验统计量的多种不同定义形式,该检验主要用于对参数有线性约束的场合。最后,基于两个实例并借助 SAS 软件,说明了评分检验的常用场合和呈现形式。

#### 参考文献

- [1] Armitage P, Colton T. Encyclopedia of biostatistics [M]. 2<sup>nd</sup> edition. John Wiley & Sons, Inc, 2005: 3027-3039, 4605-4607, 4658-4659.
- [2] SAS Institute Inc. SAS/STAT<sup>®</sup> 15.1 user's guide [M]. Cary, NC: SAS Institute Inc, 2018: 3405-3608, 5749-6006, 7223-7490, 7991-8092.
- [3] 邵军. Mathematical statistics [M]. 2 版. 北京: 世界图书出版公司, 2009: 393-470.
- [4] 胡良平. 面向问题的统计学: (2) 多因素设计与线性模型分析 [M]. 北京: 人民卫生出版社, 2012: 508-517.
- [5] 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: 532-535.
- [6] 陈希孺. 广义线性模型的拟似然法 [M]. 合肥: 中国科学技术大学出版社, 2011: 132-142.
- [7] 王济川, 郭志刚. Logistic 回归模型: 方法与应用 [M]. 北京: 高等教育出版社, 2001: 123-144, 237-262.
- [8] 凤思苑, 李长平, 胡良平. 非配对设计多值有序资料多水平多重 Logistic 回归分析 [J]. 四川精神卫生, 2019, 32(6): 481-485.
- [9] 巩晓文, 李长平, 胡良平. 非配对设计多值名义资料—水平多重 Logistic 回归分析 [J]. 四川精神卫生, 2019, 32(6): 486-489.
- [10] 刘媛媛, 李长平, 胡良平. 复杂抽样调查设计多值名义资料—水平多重 Logistic 回归分析 [J]. 四川精神卫生, 2019, 32(6): 490-494.

(收稿日期: 2021-11-10)

(本文编辑: 戴浩然)