

如何正确运用 χ^2 检验——拟合优度检验与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】 本文目的是介绍拟合优度检验及其SAS实现, 主要内容包括以下四个方面: ①Pearson's 拟合优度检验; ②偏差或称似然比拟合优度检验; ③Hosmer-Lemeshow 拟合优度检验; ④稀疏资料拟合优度检验。前述提及的第四个方面的内容包含六种具体的检验方法, 即“信息矩阵检验”“信息矩阵对角线检验”“Osious-Rojek 检验”“未加权的残差平方和检验”“Spiegelhalter 检验”和“Stukel 检验”。本文结合一个实例并借助SAS软件实现前述提及的四类拟合优度检验, 对输出结果做出解释, 并给出统计结论和专业结论。

【关键词】 拟合优度检验; 失拟检验; 似然比检验; 稀疏资料; χ^2 分布

中图分类号: R195.1

文献标识码: A

doi: 10.11886/scjsws20210915004

How to use χ^2 test correctly—— the goodness of fit tests and the implementation of the SAS software

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 purpose of this article was to introduce the goodness of fit test and its SAS implementation. The main contents included the following four aspects: ① Pearson's goodness of fit test; ② deviance or likelihood ratio goodness of fit test; ③ Hosmer-Lemeshow goodness of fit test; ④ goodness of fit test for the sparse data. In the aforementioned “fourth aspect”, there were six specific test approaches, namely “information matrix test” “information matrix diagonal test” “Osious-Rojek test” “unweighted residual sum of squares test” “Spiegelhalter test” and “Stukel test”. The paper implemented the four types of the goodness of fit tests mentioned above with the help of the SAS software through an example, explained the output results, and made statistical and professional conclusions.

【Keywords】 Goodness of fit test; Lack of fit test; Likelihood ratio test; Sparse data; χ^2 distribution

在进行列联表资料的独立性分析和/或广义线性回归模型^[1]分析中, 常会涉及“拟合优度检验”和“失拟检验”等内容^[1-4]。本文将介绍有关概念、各种检验方法的计算公式以及基于SAS软件^[5]实现计算的方法。

1 有关基本概念

1.1 拟合优度检验的定义和种类

所谓拟合优度检验, 就是评价基于某种假设或回归模型推算出按顺序排列的一组期望频数与相应观察频数之间一致性的一种检验方法^[1,5]。拟合优度检验常涉及以下两个统计学问题: 其一, 检验某种假设是否成立。例如: 在列联表资料分析中可能会提出“某些属性变量之间存在独立性”的假设,

此时的检验被称为“独立性拟合优度检验”; 又例如: 在展示一组单变量资料的频数分布时, 可能会提出“该组资料服从正态分布、Poisson 分布或二项分布”的假设, 此时的检验被称为“分布拟合优度检验”。其二, 检验所拟合的某个回归模型对所描述资料的变化趋势是否吻合, 此时的检验被称为“回归模型拟合优度检验”。

从功能上来划分, 拟合优度检验可分为四种: ①独立性拟合优度检验(常用于列联表资料的独立性检验); ②线性拟合优度检验(常用于定量因变量的线性回归分析中, 以判断是否需要采取更复杂的方法或非线性回归模型); ③分布拟合优度检验(以判断给定的资料是否服从某种特定的概率分布, 如正态分布、Poisson 分布或二项分布); ④回归模型拟合优度检验(以判断当前所构建的回归模型与给定

资料的吻合程度)”。从检验统计量所服从的分布来划分,拟合优度检验可分为三种:① F 分布(常用于线性回归模型的分析之中);②经验分布(即柯尔莫哥洛夫检验);③ χ^2 分布(常用于除线性回归模型分析之外的其他场合)。从构造检验统计量的方法来划分,拟合优度检验可分为三种:①方差分析法(常用于线性回归模型的分析);②似然比法(常用于除线性回归模型分析之外的其他场合);③观测频数与理论频数之偏差的平方和法(常用于除线性回归模型分析之外的其他场合)。这三种划分方法在内容上存在交叉和重叠。因篇幅所限,本文仅介绍基于“ χ^2 检验统计量”的拟合优度检验方法^[5-9]。

1.2 失拟检验及其与拟合优度检验的关系

所谓失拟检验,就是对回归函数是否为自变量的线性函数进行的检验^[2]。失拟检验最早出现在简单线性回归模型的分析过程中,用以检验采用简单线性回归模型是否可以取代曲线回归模型。若失拟检验的结果为 $P>0.05$,表明用简单线性回归模型可以取代曲线回归模型;反之亦然。在 SAS/STAT 的 REG 过程中,在进行多重线性回归分析时,也可以进行失拟检验^[5]。值得注意的是:在线性回归分析中所做的失拟检验是 F 检验(也叫做方差分析),而不是 χ^2 检验。在 SAS/STAT 的 LOGISTIC 过程中,将“Hosmer-Lemeshow 检验”列入“拟合优度检验(Goodness of fit tests)”标题之下,但在指定“选项(lackfit=)”时,却又将其归类于“失拟检验”之中;然而,在 SAS/STAT 的 PROBIT 过程中,将“失拟检验”与“拟合优度检验”视为同一种检验^[5]。正因如此,在下文中,统一称为“拟合优度检验”,而不特别区分何时为“拟合优度检验”,何时为“失拟检验”。

2 Pearson's 拟合优度检验

2.1 可以应用的三种场合

Pearson's 拟合优度检验可以应用于下列三种场合^[1-5]:①列联表资料独立性假定的拟合优度检验;②单变量频数分布的分布拟合优度检验;③除线性回归模型分析之外的其他回归模型的拟合优度检验。

2.2 Pearson's 拟合优度检验统计量

2.2.1 独立性拟合优度检验统计量

在进行 $R \times C$ 列联表(包括 2×2 列联表)资料的独

立性检验时,Pearson's 拟合优度检验统计量的形式见式(1):

$$\chi^2 = \sum_{i=1}^R \sum_{j=1}^C \frac{(O_{ij} - T_{ij})^2}{T_{ij}} \sim \chi_{df}^2, \quad df = (R-1)(C-1) \quad (1)$$

在式(1)中, χ^2 服从自由度为 $df=(R-1)(C-1)$ 的 χ^2 分布; R 和 C 分别代表二维列联表的“行数”与“列数”。

2.2.2 分布拟合优度检验统计量

在推断一组单变量定量资料服从何种频数分布时,Pearson's 拟合优度检验统计量的形式见式(2):

$$\chi^2 = \sum_{i=1}^k \frac{(f_i - nP_i)^2}{nP_i} \sim \chi_{df}^2, \quad df = k - r - 1 \quad (2)$$

在式(2)中, χ^2 服从自由度为 $df=k-r-1$ 的 χ^2 分布; k 为资料中不同数据个数或分组数; r 为假定概率分布函数中待估参数的个数; f_i 为定量变量的第 i 观测值(需要由小到大排序,在实际使用中,应是第 i 个组的组中值)所对应的观测频数; P_i 为与 f_i 对应的频率; n 为总样本含量; nP_i 为基于设定概率分布推算出来的理论频数。

值得注意的是:各 nP_i 应大于等于5,否则,需要将相邻的数据或组进行合并,以满足前述的前提条件。

2.2.3 回归模型拟合优度检验统计量

2.2.3.1 在 SAS/STAT 的 GENMOD 过程中的定义

在 SAS/STAT 的 GENMOD 过程中,Pearson's 拟合优度检验统计量的定义见式(3):

$$\chi^2 = \sum_{i=1}^k \frac{f_i w_i (y_i - \mu_i)^2}{V(\mu_i)} \sim \chi_{df}^2, \quad df = k - r \quad (3)$$

在式(3)中, χ^2 服从自由度为 $df=k-r$ 的 χ^2 分布; k 为资料中不同数据个数或分组数; f_i 、 w_i 、 y_i 、 μ_i 和 $V(\mu_i)$ 分别代表第 i 个组的“频数”“权重”“因变量的取值”“ y_i 的期望值”和“ y_i 的方差”; r 为所构建的回归模型中待估参数的个数(包括截距项)。值得注意的是:广义线性模型涵盖面很宽泛,其因变量通常涉及以下常见概率分布,例如:正态分布、Poisson 分布、伽玛分布、逆高斯分布、二项分布、多项分布、负二项分布、零膨胀 Poisson 分布和零膨胀负二项分布

等;依据不同的概率分布, y_i 、 μ_i 和 $V(\mu_i)$ 将取不同的值或计算公式,限于篇幅,不再赘述。

2.2.3.2 在SAS/STAT的LOGISTIC过程中的定义

在SAS/STAT的LOGISTIC过程中,Pearson's拟合优度检验统计量的定义见式(4):

$$\chi^2_P = \sum_{j=1}^m \sum_{i=1}^{k+1} \frac{(r_{ij} - n_j \hat{\pi}_{ij})^2}{n_j \hat{\pi}_{ij}} \sim \chi^2_{df}, df = mk - P \quad (4)$$

在式(4)中, χ^2_P 服从自由度为 $df=mk-P$ 的 χ^2 分布; r_{ij} 代表第 j 组因变量的第 i 个水平下观测的事件数(或称为阳性数); n_j 为第 j 组的总频数; $\hat{\pi}_{ij}$ 代表第 j 组因变量的第 i 个水平下事件(或称为阳性)出现的概率; m 为资料中不同数据个数或分组数; $k+1$ 代表定性因变量的水平数; P 为所构建的回归模型中待估参数的个数(包括截距项)。

2.2.3.3 在SAS/STAT的PROBIT过程中的定义

在SAS/STAT的PROBIT过程中,Pearson's拟合优度检验统计量的定义见式(5):

$$\chi^2_P = \sum_{i=1}^m \sum_{j=1}^k \frac{(r_{ij} - n_i \hat{P}_{ij})^2}{n_i \hat{P}_{ij}} \sim \chi^2_{df},$$

$$df = (k-1) \times m - q \quad (5)$$

在式(5)中, χ^2_P 服从自由度为 $df=(k-1) \times m - q$ 的 χ^2 分布; r_{ij} 代表第 i 组因变量的第 j 个水平下观测的事件数(或称为阳性数); n_i 为第 i 组的总频数; $\hat{P}_{ij}$ 代表第 i 组因变量的第 j 个水平下事件(或称为阳性)出现的拟合概率; m 为资料中不同数据个数或分组数; k 代表定性因变量的水平数; q 为所构建的回归模型中待估参数的个数(包括截距项)。

3 偏差或称似然比拟合优度检验

3.1 可以应用的场合

偏差或称似然比拟合优度检验的应用场合较少,目前仅应用于回归模型拟合优度检验和检验整个回归模型中全部参数是否都为0。

3.2 偏差拟合优度检验统计量

3.2.1 在SAS/STAT的GENMOD过程中的定义

在SAS/STAT的GENMOD过程中^[5],偏差拟合优度检验统计量的定义见式(6):

$$D(Y, \mu) = 2[l(Y, Y) - l(Y, \mu)] \quad (6)$$

在式(6)中, Y 与 μ 分别代表因变量的向量与因

变量预测均值向量; $l(Y, Y)$ 和 $l(Y, \mu)$ 分别代表含有全部自变量的回归模型的似然函数的自然对数与含有部分自变量的回归模型的似然函数的自然对数; $D(Y, \mu)$ 服从自由度为 $df=n-q$ 的 χ^2 分布,这里, n 和 q 分别代表总样本含量与所构建的回归模型中待估参数的个数(包括截距项)。

3.2.2 在SAS/STAT的LOGISTIC过程中的定义

在SAS/STAT的LOGISTIC过程中^[5],偏差拟合优度检验统计量的定义见式(7):

$$\chi^2_D = 2 \sum_{j=1}^m \sum_{i=1}^{k+1} r_{ij} \ln \left(\frac{r_{ij}}{n_j \hat{\pi}_{ij}} \right) \sim \chi^2_{df},$$

$$df = m \times k - P \quad (7)$$

在式(7)中, χ^2_D 服从自由度为 $df=m \times k - P$ 的 χ^2 分布;其他符号的含义与式(4)后面的内容相同,此处从略。

3.2.3 在SAS/STAT的PROBIT过程中的定义

在SAS/STAT的PROBIT过程中^[5],偏差拟合优度检验统计量的定义见式(8):

$$\chi^2_D = 2 \sum_{i=1}^m \sum_{j=1}^k r_{ij} \ln \left(\frac{r_{ij}}{n_i \hat{P}_{ij}} \right) \sim \chi^2_{df},$$

$$df = (k-1) \times m - q \quad (8)$$

在式(8)中, χ^2_D 服从自由度为 $df=(k-1) \times m - q$ 的 χ^2 分布;其他符号的含义与式(5)中的相应内容相同,此处从略。

4 Hosmer-Lemeshow 拟合优度检验

4.1 概述

前面介绍的Pearson's拟合优度检验与偏差拟合优度检验都是基于数据分组之上的,当资料中存在连续性变量时,就不可避免地会出现某些组中的样本含量很少,甚至只有1~2个样本,这种现象被称为“稀疏资料”。事实上,前面提及的两种拟合优度检验要求各组样本含量应大于等于5,否则,其检验结果不可靠。针对“稀疏资料(某些组中没有重复数据)”,Hosmer和Lemeshow于2000年提出了一种新的拟合优度检验方法,该拟合优度检验统计量近似服从 χ^2 分布^[5]。

4.2 二值因变量条件下 Hosmer-Lemeshow 拟合优度检验统计量

二值因变量条件下 Hosmer-Lemeshow 拟合优度

检验统计量见式(9):

$$\chi_{il}^2 = \sum_{j=1}^g \frac{(O_j - F_j \bar{\pi}_j)^2}{F_j \bar{\pi}_j (1 - \bar{\pi}_j)} \sim \chi_{df}^2, df = g - r \quad (9)$$

在式(9)中, χ_{il}^2 为服从自由度为 $df=g-r$ 的 χ^2 分布; g 代表组数; r 代表模型中参数的数目(可在“model 语句”的选项“LACKFIT”后通过子选项“DFREDUCE=”来指定, 系统默认值为 $r=2$); F_j 为第 j 组中的总样本含量; O_j 为第 j 组中的总事件数(或阳性数); $\bar{\pi}_j$ 为第 j 组中出现事件的预测概率的平均值。

4.3 多值因变量条件下 Hosmer-Lemeshow 拟合优度检验统计量

多值因变量条件下 Hosmer-Lemeshow 拟合优度检验统计量^[5]见式(10):

$$\chi_{il}^2 = \sum_{j=1}^g \sum_{k=1}^K \frac{(O_{jk} - E_{jk})^2}{E_{jk}} \sim \chi_{df}^2 \quad (10)$$

式(10)中的 χ_{il}^2 为服从 χ^2 分布的随机变量, 自由度有两种不同的取值(假定所有模型中斜率相等), 当因变量为多值有序变量时, $df=(g-r)(K-1)+K-2$; 当因变量为“多值名义变量”时, $df=(g-r)(K-1)$ 。 g 代表组数; K 代表因变量的水平数; O_{jk} 与 E_{jk} 分别代表第 j 组因变量的第 k 个水平条件下观测频数与基于模型计算的预测频数。

5 稀疏资料拟合优度检验

5.1 概述

尽管 Hosmer-Lemeshow 拟合优度检验对稀疏资料具有一定的校正能力, 但它仍然需要将资料分组后才能进行计算。当资料中“稀疏情形”非常严重时, 其检验效能将会明显下降。为此, 统计学家提出了另外 6 种稀疏资料拟合优度检验方法。迄今为止, 这些方法仅适用于因变量为二值变量的 logistic 回归模型之中^[5]。设 $\hat{\pi}_j$ 代表事件出现的预测概率, $\hat{V}$ 代表拟合模型的协方差矩阵。

5.2 不需对资料进行分组的稀疏资料拟合优度检验法

5.2.1 信息矩阵检验

Orme 于 1988 年将 White 于 1982 年提出的“一般错误指定检验”应用于二项反应资料。设 X_j 为第 j 个观测的设计向量(即由全部协变量形成的向量), 于

是, $X_j X_j'$ 就是信息矩阵。通过信息矩阵 $X_j X_j'$ 的右上角矩阵, 将设计向量 X_j 展开成为一个新向量, 记为 $\text{Vech}(X_j)$ 。通过 $\sqrt{\hat{\pi}_j(1-\hat{\pi}_j)}$ 将 X_j 尺度化, 通过 $\sqrt{\hat{\pi}_j(1-\hat{\pi}_j)}(1-2\hat{\pi}_j)$ 将 $\text{Vech}(X_j)$ 尺度化。使用残差 $e_j = Y_j - \hat{\pi}_j$ 的二项形式创建新反应变量的值 Y^* , 其表达式见式(11):

$$Y^* = \frac{e_j}{\sqrt{\hat{\pi}_j(1-\hat{\pi}_j)}} \quad (11)$$

构建 Y^* 关于展开后的协变量的多重线性回归模型, 求出该模型的回归平方和, 这个平方和服从自由度为 df 的 χ^2 分布。这里, df 等于尺度化后的 $\text{Vech}(X_j)$ 中未退化的协变量的数目。基于此方法产生的检验结果在拟合优度检验表中的标签为“信息矩阵”。

5.2.2 信息矩阵对角线检验

Kuss 于 2002 年修改了上述“信息矩阵检验”方法, 即仅通过信息矩阵 $X_j X_j'$ 对角线上的元素, 将设计向量 X_j 展开成为一个新向量, 记为 $\text{Vech}(X_j)$ 。基于此方法产生的检验结果在拟合优度检验表中的标签为“信息矩阵对角”。

5.2.3 Osius-Rojek 检验

Osius 和 Rojek 于 1992 年用固定格渐近方法推导出 Pearson's χ^2 统计量的均值和方差, 均值为 m , 即子总体的个数(即“组数”); 方差的计算方法见式(12):

$$\sigma^2 = 2m + \sum_{j=1}^m \frac{1}{w_j n_j} \left(\frac{1}{\hat{\pi}_j(1-\hat{\pi}_j)} - 6 \right) - c' \hat{V} c \quad (12)$$

在式(12)中, $c = \sum_{j=1}^m (1-2\hat{\pi}_j) X_j$; 标准化的 Pearson's χ^2 统计量见式(13):

$$B = \frac{\chi_p^2 - m}{\sigma} \quad (13)$$

基于式(13)可得到 Osius-Rojek 检验统计量见式(14):

$$\chi_{OR}^2 = B^2 \sim \chi_1^2 \quad (14)$$

在式(14)中, B 的定义见式(13); χ_{OR}^2 为服从自由度为 1 的 χ^2 分布的检验统计量。

5.2.4 未加权的残差平方和检验

Hosmer 等于 1997 年在 Copas 于 1989 年提出的基本公式的基础上, 构造出未加权的残差平方和检

验统计量,见式(15):

$$\chi_{URSS}^2 = \left[\frac{S - \text{trace}(\hat{V})}{\sigma} \right]^2 \sim \chi_1^2 \quad (15)$$

在式(15)中, σ 为回归的残差标准差,其他符号的含义分别如下:

$$S = \sum_{i=1}^2 w_{ij} (r_{ij} - n_j \hat{\pi}_{ij})^2 \quad (16)$$

$$E(S) = \text{trace}(\hat{V}) \quad (17)$$

5.2.5 Spiegelhalter 检验

Spiegelhalter 于 1986 年基于 Brier 评分推导出一个检验统计量,见式(18):

$$\chi_{SPI}^2 = \left[\frac{B - E(B)}{\sqrt{\text{Var}(B)}} \right]^2 \sim \chi_1^2 \quad (18)$$

在式(18)中,各符号的含义如下:

$$B = \frac{1}{W} w_j (Y_j - \hat{\pi}_j)^2 \quad (19)$$

$$W = \sum_{j=1}^N w_j \quad (20)$$

$$E(B) = \frac{1}{W} w_j \hat{\pi}_j (1 - \hat{\pi}_j) \quad (21)$$

$$\text{Var}(B) = \frac{1}{W^2} w_j (1 - 2\hat{\pi}_j)^2 \hat{\pi}_j (1 - \hat{\pi}_j) \quad (22)$$

5.2.6 Stukel 检验

Stukel 于 1988 年提出了如下方法,在现有模型中增加两个新变量,然后检验这两个新变量是否为无统计学意义的变量。对于二值反应变量而言,设 $\eta_j = X_j' \beta$, $I(\cdot)$ 代表指示变量,增加的两个新变量如下:

$$Z_a = \eta_j^2 I(\eta_j \geq 0) \quad (23)$$

$$Z_b = \eta_j^2 I(\eta_j < 0) \quad (24)$$

然后,应用“评分检验(说明:此检验常被用于检验模型中部分回归系数是否为 0)”检验这个大模型(指增加两个新变量后的模型)与已拟合的模型(即未增加新变量的模型)之间差异是否有统计学意义。若检验的结果为“ $P > 0.05$ ”,说明原先已拟合的模型对资料的拟合效果满意。

6 实例与 SAS 实现

6.1 问题与数据

【例 1】为探讨新生儿乙肝病毒宫内感染的危险

因素,某医科大学流行病学教研室于 2004 年-2005 年收集了 145 例 HBsAg 阳性孕妇及其 145 例新生儿作为研究对象,检测新生儿出生 24 小时内血清中 HBV DNA,以判断是否发生了乙肝病毒宫内感染,新生儿血清 HBV DNA 阳性者为发生宫内感染,阴性者为未发生宫内感染。同时检测孕妇血清中 HBV DNA(阴性=0,阳性=1)、HbeAg(阴性=0,阳性=1),所得资料见表 1^[10],试分析新生儿乙肝宫内感染的影响因素。

表 1 145 例孕妇及其新生儿血清 HBV 相关指标检测结果

孕妇血中 HBV DNA(X_1)	孕妇血中 HbeAg(X_2)	例数		
		Y:	1	0
0	0		5	72
0	1		1	15
1	0		1	3
1	1		19	29

注:Y 代表“是否出现乙肝病毒宫内感染”,Y=1 代表“感染”,Y=0 代表“未感染”

【分析与解答】设所需要的 SAS 程序如下:

```
data abc;
input X1 X2 @@;
do y=1 to 0 by -1;
input f @@;
output;
end;
cards;
0 0 5 72
0 1 1 15
1 0 1 3
1 1 19 29
;
run;
proc logistic data=abc descending;
model y=X1 X2/cl stb rsq scale=none
aggregate lackfit gof;
freq f;
run;
```

【程序说明】“model 语句”的选项“scale=none aggregate lackfit gof”要求进行各种拟合优度检验。

【SAS 输出结果及解释】

拟合优度检验

准则	值	自由度	Pr>卡方
偏差	0.2014	1	0.6536
Pearson	0.1919	1	0.6613
信息矩阵	0.1919	1	0.6613
信息矩阵对角线	0.1919	1	0.6613
Osius 和 Rojek	1.8939	1	0.1688
残差平方和	0.1919	1	0.6613
Spiegelhalter	0.0002	1	0.9900
Stukel	0.1919	1	0.6613

以上输出的是“拟合优度检验”结果,前两行是最常用的方法,后六行是稀疏资料拟合优度检验结果, P 均 >0.05 ,说明所构建的回归模型对资料的拟合效果好。

模型拟合统计量

准则	值
AIC	119.6366
AICC	119.8068
SC	128.5668
-2 Log L	113.6366
-2 Log L(仅截距)	136.4004
误分类率	0.1793
平均平方误差	0.1232
R 方	0.1453
最大重新换算 R 方	0.2383
Aldrich-Nelson R 方	0.1357
Efron R 方	0.1625
Estrella R 方	0.1578
Estrella 的修正 R 方	0.1161
McFadden R 方	0.1669
Tjur R 方	0.1618
Veall-Zimmerman R 方	0.2799

以上输出的是“模型拟合统计量”,它们只能相对反映模型对资料的拟合效果,因为它们不是假设检验的结果,只有在有两个或两个以上模型时,才可根据模型拟合统计量的数值,判断哪个模型对资料的拟合效果相对更好(例如,各种“ R^2 ”值越大越好,AIC、AICC 和 SC 的值越小越好)。

Hosmer 和 Lemeshow 拟合优度检验

卡方	自由度	Pr>卡方
0.1919	2	0.9085

以上输出的是“Hosmer 和 Lemeshow 拟合优度检验”结果,因 $P=0.9085>0.05$,说明所构建的回归模型对资料的拟合效果较好,不存在“失拟”问题。

检验全局原假设: $BETA=0$

检验	卡方	自由度	Pr>卡方
似然比	22.7638	2	<0.0001
评分	23.3250	2	<0.0001
Wald	18.7762	2	<0.0001

以上输出的是“检验全局原假设: $BETA=0$ ”的结果,三种检验方法所得结果和结论基本一致,即认为回归模型中的回归系数不全为 0,所构建的回归模型具有一定的实用价值。

优比估计

效应	点估计	95% Wald 置信限
X1	7.293	1.694 31.393
X2	1.344	0.305 5.924

以上输出的是“优比估计”结果, X_2 的 95% Wald 置信区间包括“1”,表明 X_2 的两个水平对结果的影响接近相等; X_1 的 $\widehat{OR}_{X_1} = 7.293$,其 95% Wald 置信区间为 $[1.694, 31.393]$,未包括“1”,表明 X_1 的两个水平之间对结果的影响是不同的,即“孕妇血中 HBV DNA(X_1)”为阳性条件下“新生儿出现乙型肝炎病毒宫内感染”的概率是“孕妇血中 HBV DNA(X_1)”为阴性条件下“新生儿出现乙型肝炎病毒宫内感染”概率的 7.293 倍。

用于 Hosmer 和 Lemeshow 检验的分区

组	合计	y = 1		y = 0	
		观测	期望	观测	期望
1	77	5	4.71	72	72.29
2	16	1	1.29	15	14.71
3	4	1	1.29	3	2.71
4	48	19	18.71	29	29.29

以上输出的是“用于 Hosmer 和 Lemeshow 检验的分区”结果,实际上就是将全部资料划分成 4 个组。

最大似然估计分析

参数	自由度	估计	标准误差	Wald 卡方	Pr>卡方	标准化估计
Intercept	1	-2.7307	0.4515	36.5727	<0.0001	
X1	1	1.9869	0.7447	7.1177	0.0076	0.5272
X2	1	0.2957	0.7568	0.1527	0.6960	0.0812

以上输出的是“最大似然估计分析”结果,此结果表明: X_2 的回归系数与0之间的差别无统计学意义,表明“孕妇血中HbeAg(X_2)”是否为“阳性”对“新生儿是否出现乙肝病毒宫内感染”的影响可以忽略不计。要想获得更简洁的回归模型,可以淘汰掉自变量 X_2 后重新拟合logistic回归模型。

【统计结论与专业结论】由多种拟合优度检验结果可知,所拟合的回归模型对资料的拟合效果满意;由回归系数的检验结果可知,孕妇血中HBV DNA(X_1)是新生儿乙肝宫内感染的主要影响因素。

7 讨论与小结

7.1 讨论

拟合优度检验与失拟检验在本质上是相同的,都是用来评价某种假设(例如独立性假设或关于资料来自某种分布的假设)是否成立或所拟合的回归模型与资料的吻合程度。其微小差别在于:拟合优度检验一般要求将资料整理成一维频数表或高维频数表资料形式,当检验结果为 $P>0.05$ 时,表明所做的假设成立或所拟合的回归模型与资料的变化趋势是吻合的;当检验结果为 $P\leq 0.05$ 时,表明所做的假设可能不成立或所拟合的回归模型与资料的变化趋势不够吻合;然而,失拟检验一般要求在自变量或自变量向量取特定值的条件下应做多次重复试验,采用失拟检验的目的是揭示资料中所包含的“曲线变化的成分”是否足够多。当检验结果为 $P>0.05$ 时,表明所拟合的回归模型与资料的变化趋势是吻合的;当检验结果为 $P\leq 0.05$ 时,表明所拟合的回归模型与资料的变化趋势不够吻合,分析者需要构建更复杂的回归模型或非线性回归模型。

7.2 小结

本文介绍了9种拟合优度检验方法,它们的主

要应用场合是评价广义线性回归模型(包括logistic回归模型)对资料的拟合效果。其中,偏差拟合优度检验和Pearson's拟合优度检验最常用;Hosmer-Lemeshow拟合优度检验对稀疏资料具有一定的校正作用;当资料的稀疏程度十分严重时,需要使用专门用于稀疏资料的6种拟合优度检验。本文结合一个实例并借助SAS软件实现了全部9种拟合优度检验,对SAS输出结果做了解释,给出了统计结论和专业结论。

参考文献

- [1] Armitage P, Colton T. Encyclopedia of biostatistics [M]. 2nd edition. John Wiley & Sons, 2005: 2269-2295, 2389-2412.
- [2] 茆诗松. 统计手册[M]. 北京: 科学出版社, 2003: 169-173, 448-459, 617-618.
- [3] 方积乾, 陆盈. 现代医学统计学[M]. 北京: 人民卫生出版, 2002: 596-597.
- [4] 万崇华, 罗家洪. 高级医学统计学[M]. 北京: 科学出版社, 2014: 275.
- [5] SAS Institute Inc. SAS/STAT®15.1 user's guide[M]. Cary, NC: SAS Institute Inc, 2018: 3405-3608, 5749-6006, 7223-7490, 7991-8092.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 李长平, 张甜甜, 宋德胜, 等. 非配对设计多值名义资料多水平多重 Logistic 回归模型[J]. 四川精神卫生, 2019, 32(6): 495-500.
- [10] 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, 2012: 418-432.

(收稿日期:2021-09-15)

(本文编辑:陈霞)