
Logistic回归中变量赋值的几大常见问题

作者：王晓晓，赵一鸣 来源：临床流行病学和循证医学

本文原地址：<https://www.iikx.com/news/statistics/2061.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

Logistic回归中变量赋值

是很重要的，涉及到结果的准确性。今天，我们说说经常被问到又困惑大家的几个问题。Logistic回归中二分类的变量可以是所关心的结局和相应的对照，鉴于最常见的为疾病和正常对照，文中为方便起见，简化为疾病和正常。特此说明。

首先是二分类的因变量

，分别用两个数字代表正常和疾病组，不管您采用哪两个数字，其实关系不大，因为系统会对因变量进行重新编码，一般默认取值水平高的为阳性结果。

比如，我们赋较大值为疾病组，不管是哪种情况，系统都会默认疾病组为阳性，这样的话，结果解读时，OR大于1提示为危险因素，OR小于1提示为保护因素。倘若，我们将正常和疾病组两个水平弄反了，即赋较大值为正常组，结果解读时就要小心了，这时候OR小于1提示为危险因素，OR大于1提示为保护因素。

因变量编码

原值	内部值
正常 0	0
疾病 1	1
正常 1	0
疾病 2	1
正常 6	0
疾病 7	1
正常 1	0
疾病 100	1

临床流行病学和循证医学

那么问题来了，这两个数字有什么要求吗？必须是0和1吗？还是1和2？6和7可以吗？1和100又怎么样？还是上面那句话，系统会默认取值水平高的为阳性结果，所以说，0和1、1和2、6和7、1和100等等都是可以的，只要疾病组取值水平是高的就可以。但，最简单最推崇的做法是这样的：正常为0，疾病为1。

然后是自变量赋值，即SPSS中的协变量。

对于连续变量，遵循原始数据的样子就可以。

对于二分类变量

，一定要赋值为临近的两个数字，比如0和1、1和2、6和7，注意，当二分类变量作为自变量时，一定要是相差1个单位的两个数字。建议大家采用0和1。

对于有序多分类变量

，可以是连续的几个数字，比如1、2、3、4，或者是5、6、7、8，但一定要是连续的几个数字。建议大家采用1、2、3、4。



对于无序多分类变量

，需进行哑变量处理。点击右上角的“分类”，在弹出的“定义分类变量”的对话框中，在左边的协变量列表框选中需要进行哑变量化的变量，点击移至右边的分类协变量框即可。注意，其他无需进行哑变量化的变量请勿进行任何操作，尤其是连续变量，至于为什么，您看看输出结果就会心领神会的。还要说一句，哑变量化时，默认以取值水平最大的作为参照，您若是想以取值水平最小的作为参照，需要在“参考类别”选择“第一个”，一定记得点击上方的“变化量”按钮。



更多 统计方法 请访问 <https://www.iikx.com/news/statistics/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](http://www.iikx.com)转发