
SPSS : Logistic回归分类变量 (哑变量) 的处理及解读

作者：王江源 来源：爱科学

本文原地址：<https://www.iikx.com/news/statistics/6063.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！

SPSS : Logistic回归分类变量 (哑变量) 的处理及解读

一、哑变量的设置方法

Logistic回归中分类变量需要使用哑变量(也叫虚拟变量)来操作。

一般的，n个分类需要设置n-1个哑变量(为什么不是n个?请继续看)。

举个例子，有一个“年龄”变量，分为：青年，中年，老年三类，那么我们可以用两个哑变量来代替：

年龄	变量1	变量2
青年	1	0
中年	0	1
老年	0	0

变量1 = 1代表青年，0代表非青年

变量2 = 1代表中年，0代表非中年

变量1和变量2都等于0代表老年

所以用2个变量就可以表示3个类别。

二、分类变量在SPSS中的操作及结果解读

SPSS中能自动设置哑变量，只需要把变量标记为分类变量即可。

假设我们要分析年龄和病程对某种疾病预后的影响，采用Logistic回归分析。

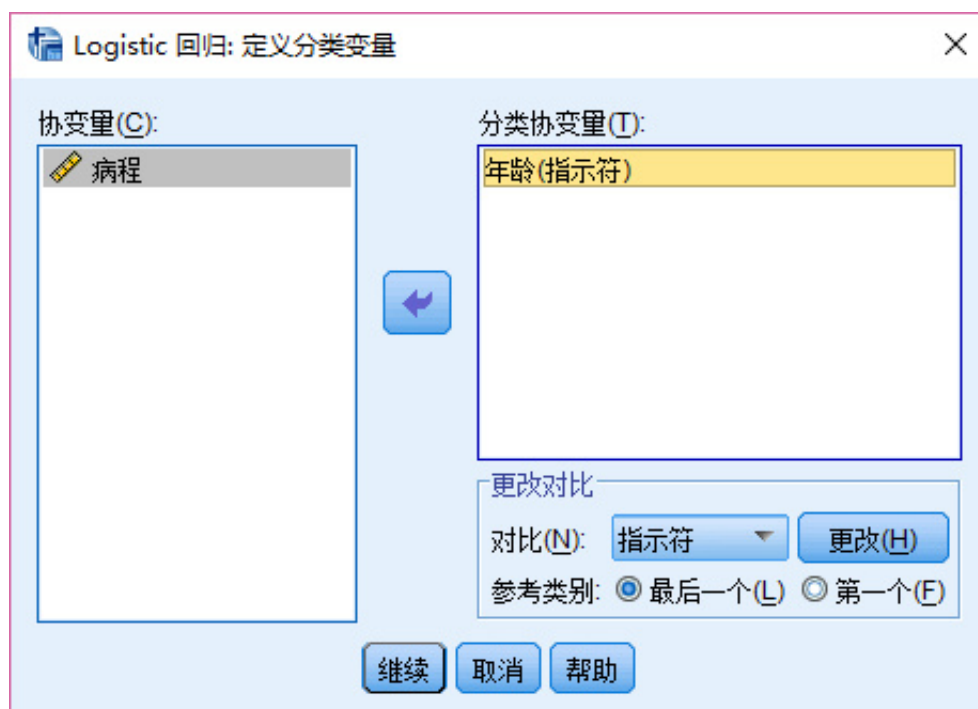
变量赋值如下(数据均为人造，非真实数据)：

预后：因变量，为二分类变量，0=预后差，1=预后好

年龄：自变量，为多分类变量，1=青年，2=中年，3=老年

病程：自变量，为连续变量

(1)首先将年龄设置为分类变量，对比方式默认为“指示符”，参考类别默认为“最后一个”(后面解释为什么)。见下图。





(2)结果输出，有两个主要的表格。

分类变量编码

		频率	参数编码	
			(1)	(2)
年龄	青年	52	1.000	.000
	中年	32	.000	1.000
	老年	16	.000	.000

这是分类变量的编码表格，可以看出，年龄被替换为两个新的变量：年龄(1)和年龄(2)。年龄(1)代表青年人，年龄(2)代表中年人，他们的取值都为0表示老年人，作为青年和中年的参考对象。

方程中的变量

		B	S.E.	Wals	df	Sig.	Exp (B)
步骤 1 ^a	年龄			36.048	2	.000	
	年龄(1)	5.131	1.232	17.333	1	.000	169.150
	年龄(2)	1.821	1.218	2.236	1	.135	6.181
	病程	.058	.080	.530	1	.467	1.060
	常量	-3.070	1.246	6.072	1	.014	.046

a. 在步骤 1 中输入的变量: 年龄, 病程.

这是回归表格，出现了年龄(1)和年龄(2)两个新的变量。可以看出年龄(1)的P为0.000，有统计学意义，年龄(2)的P为0.135，没有统计学意义。

两者不一致，怎么解释？

因为年龄(1)和(2)都是以老年人来作为参照的，所以可以解释为：

(1)青年人相对于老年人，预后更好

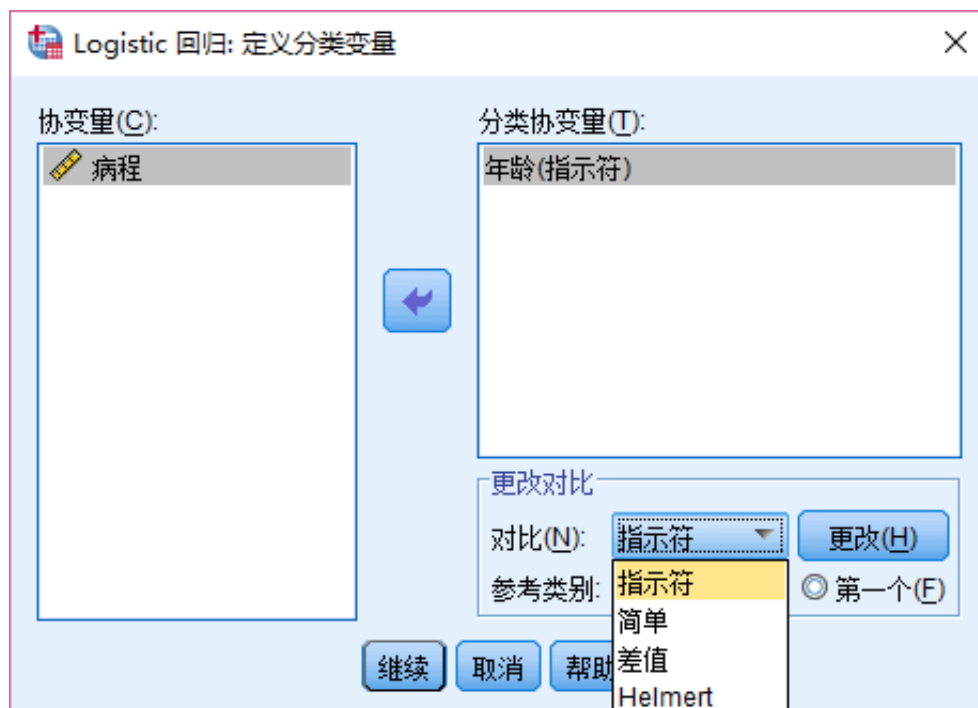
(2)中年人相对于老年人，预后没有统计学差异

(3)青年人比中年人看起来预后好，但需要进一步假设检验。

三、参照方式的选择

分类变量都需要一个参考对象，也就是说跟谁比。

SPSS中提供了多种对比方式，如指示符，简单，差值等等，如下图：



其中默认的“指示符”使用最多，这里仅介绍这一个。

“指示符”表示将每一个类别与参考类别对比。那么哪一个是参考类别呢？SPSS有两个选项：“最后一个”与“第一个”。这里的“最后一个”和“第一个”顺序与上文“分类变量编码表”中的顺序是一样的。如果设置为最后一个，就是以老年为参考类别，如果设置为第一个，就是以青年为参考类别。具体使用哪一个，需要根据分析目的来确定。

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