SPSS：Logistic回归(Logistic regression)概述

作者：writer 来源：临床流行病学和循证医学

本文原地址：https://www.iikx.com/news/statistics/6284.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(Logistic regression)概述。在医学研究中，经常需要分析疾病与各危险因素之间的定量关系，如食管癌的发生与吸烟、饮酒、不良饮食习惯等危险因素的关系，为了正确说明这种关系，需要排除一些混杂因素的影响。谈到如何很好的解决这个问题，就不得不提我们今天的主题了——Logistic回归。

Logistic回归(Logistic regression)属于概率型非线性回归，是研究二分类(可扩展到多分类)观察结果和一些影响因素之间关系的一种多变量分析方法
。Logistic回归常用于疾病的影响因素分析，用Logistic回归可以很好的解决我们刚刚提出的问题。

首先我们来认识下Logistic回归模型，这会有助于理解即将推出的倾向评分(PS)原理和实施。设应变量Y是一个二分类变量，Y=1(出现阳性结果)，Y=0(出现阴性结果)，另有影响Y取值的m个自变量x1，x2，…xm，P表示在m个自变量作用下阳性结果发生的概率。Logistic回归模型可以表示如下，其中g(x)表示阳性与阴性结果发生概率之比的自然对数，称为logitp。

$$g(x)=\beta_0+\beta_1x_1+\beta_2x_2+....+\beta_ix_i+....+\beta_mx_m$$

$$g(x)=\ln\left[\frac{P}{1-P}\right]$$

模型参数对结果解读相当重要，其中β0为常数项，β1，β2……。βm分别为m个自变量的偏回归系数。

由上述公式可知，偏回归系数βm表示扣除了其他自变量的影响，自变量xm改变一个单位时logitp的改变量。

知道了Logistic回归的原理，SPSS的操作马上呈现给大家。例：比较新疗法与旧疗法治疗某种疾病的疗效。共40例患者，20例接受新疗法，20例接受旧疗法治疗。根据专业知识，患者的病情严重程度、年龄对疗效也有影响。如何评价新旧疗法的疗效(注：作为举例，本例样本量仅为40例，由于样本量太小，Logistic回归的结论仅作为参考)

logistic – SPSS Data Editor

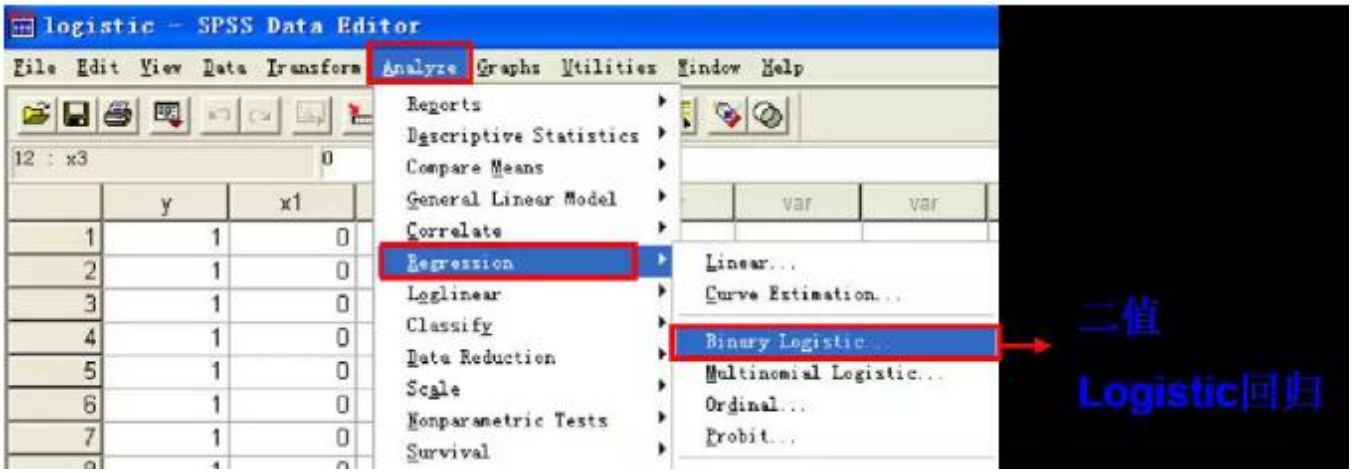
File Edit View Data Transform Analyze Graphs Utili

25 :

	y	x1	x2	x3
1	1	0	20	1
2	1	0	23	1
3	1	0	32	1
4	1	0	38	1
5	1	0	25	0
6	1	0	20	0
7	1	0	24	0
8	1	0	28	0
9	1	0	30	0
10	1	0	32	0
11	1	0	38	0
12	1	0	26	0
13	1	0	29	0

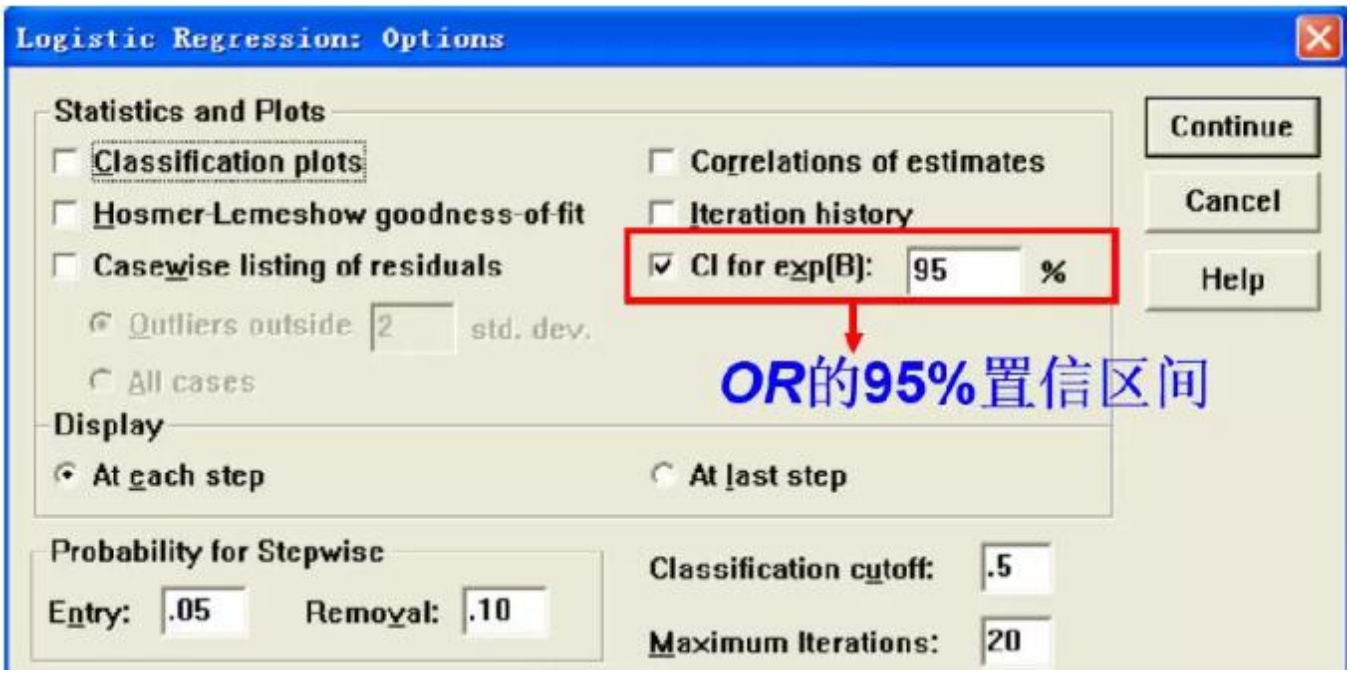
变量说明：Y：治愈情况，1=治愈;0=未治愈;X1：病情严重程度，0=不严重，1=严重;X2：年龄。X3：治疗方法，0=新疗法，1=旧疗法。

执行Analyze-Regression-Binary Logistic





将y选入因变量，x1，x2，x3选入自变量。这里注意一下，当自变量的个数较多时，为了使建立的Logistic回归模型比较稳定和便于解释，应尽可能将回归效果显著的自变量选入模型中，将作用不显著的自变量排除在外。具体方法有前进法、后退法和逐步法(后退法最好)，一般默认为Enter(全部进入)。



在一系列点击菜单的操作后，我们终于看到了输出结果。首先是对Logistic回归模型的检验。

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	13.951	3	.003
	Block	13.951	3	.003
	Model	13.951	3	.003

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	40.597	.294	.396

北医三院临床流行病和循证医学中心

经统计学检验，模型c2=13.951，P=0.003，Logistic回归模型有显著性。

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	X1	-1.595	.858	3.452	1	.063	.203	.038	1.092
	X2	-.077	.071	1.176	1	.278	.926	.806	1.064
	X3	-2.273	.857	7.028	1	.008	.103	.019	.553
	Constant	3.734	2.335	2.556	1	.110	41.826		

a. Variable(s) entered on step 1: X1, X2, X3.

北医三院临床流行病和循证医学中心

回归系数 标准误 Wald值 P值 OR OR置信区间

根据模型，病情严重程度与治疗方法对患者的治愈情况有影响;其中病情严重组相对于不严重组，OR=0.203，95%置信区间为(0.038，1.092)(区间包括1，缺乏实际意义,不多作解释);旧疗法组相对于新疗法组，OR=0.103，95%置信区间为(0.019，0.553)。

鉴于
Logistic回归的自变量既可以是连续变量，也可以是分类变量。对于连续变量、二分类变量和等级变量无需特殊处理，但如果自变量是名义变量，我们需要将名义变量哑变量化，可通过Categorical按钮来实现。

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转发