

## 诊断 - 二分类 logistic 回归

Characteristics	Total(N)	OR(95% CI) Univariate analysis	P value Univariate analysis	OR(95% CI) Multivariate analysis	P value Multiv
Age	228	1.039 (1.006 - 1.072)	0.020	1.009 (0.969 - 1.051)	0.6
Weight loss	214	1.006 (0.983 - 1.029)	0.609		
Sex	228				
Male	138	Reference		Reference	
Female	90	0.333 (0.183 - 0.605)	< 0.001	0.285 (0.137 - 0.595)	< 0.
Grade	228				
0	40	Reference		Reference	
1	92	0.171 (0.021 - 1.362)	0.095	0.211 (0.026 - 1.740)	0.1
2	96	0.024 (0.003 - 0.179)	< 0.001	0.026 (0.003 - 0.205)	< 0.
Stage	227				
Stage1	63	Reference		Reference	
Stage2	113	1.859 (0.970 - 3.560)	0.062	1.696 (0.754 - 3.818)	0.2
Stage3	51	5.270 (1.961 - 14.163)	< 0.001	4.754 (1.255 - 18.014)	0.0

网址: <https://www.xiantao.love>



更新时间: 2023.03.28

## 目录

基本概念 .....	3
应用场景 .....	3
分析流程 .....	4
结果解读 .....	5
数据格式 .....	6
参数说明 .....	8
阈值控制 .....	8
数据处理 .....	9
置信区间 .....	10
结果说明 .....	11
主要结果 .....	11
补充结果 .....	13
SPSS 验证 .....	16
方法学 .....	22
如何引用 .....	23
常见问题 .....	24

## 基本概念

二分类 logistic 回归分析：当因变量是二分类变量时，常使用 logistic 回归分析自变量与因变量的关联。

### ➤ 结果构成

Characteristics	Total(N)	OR(95% CI) Univariate analysis	P value Univariate analysis	OR(95% CI) Multivariate analysis	P value Multiv
Age	228	1.039 (1.006 - 1.072)	0.020	1.009 (0.969 - 1.051)	0.8
Weight loss	214	1.006 (0.983 - 1.029)	0.609		
Sex	228				
Male	138	Reference		Reference	
Female	90	0.333 (0.183 - 0.605)	< 0.001	0.285 (0.137 - 0.595)	< 0.
Grade	228				
0	40	Reference		Reference	
1	92	0.171 (0.021 - 1.362)	0.095	0.211 (0.026 - 1.740)	0.1
2	96	0.024 (0.003 - 0.179)	< 0.001	0.026 (0.003 - 0.205)	< 0.
Stage	227				
Stage1	63	Reference		Reference	
Stage2	113	1.859 (0.970 - 3.560)	0.062	1.696 (0.754 - 3.818)	0.2
Stage3	51	5.270 (1.961 - 14.163)	< 0.001	4.754 (1.255 - 18.014)	0.0

[Logistic回归.xlsx](#)
[Logistic回归.docx](#)
[预测值-系数.xlsx](#)

## 应用场景

logistic 回归在流行病学、实验研究、临床试验评价及等方面均有广泛应用。

## 分析流程



注：

模型评价包括：似然比检验、C 指数评价模型的预测能力、模型拟合度评价

## 结果解读

Characteristics	Total(N)	OR(95% CI) Univariate analysis	P value Univariate analysis	OR(95% CI) Multivariate analysis	P value Multiv
Age	228	1.039 (1.006 - 1.072)	0.020	1.009 (0.969 - 1.051)	0.8
Weight loss	214	1.006 (0.983 - 1.029)	0.609		
Sex	228				
Male	138	Reference		Reference	
Female	90	0.333 (0.183 - 0.605)	< 0.001	0.285 (0.137 - 0.595)	< 0.
Grade	228				
0	40	Reference		Reference	
1	92	0.171 (0.021 - 1.362)	0.095	0.211 (0.026 - 1.740)	0.1
2	96	0.024 (0.003 - 0.179)	< 0.001	0.026 (0.003 - 0.205)	< 0.
Stage	227				
Stage1	63	Reference		Reference	
Stage2	113	1.859 (0.970 - 3.560)	0.062	1.696 (0.754 - 3.818)	0.2
Stage3	51	5.270 (1.961 - 14.163)	< 0.001	4.754 (1.255 - 18.014)	0.0

[Logistic回归.xlsx](#)
[Logistic回归.docx](#)
[预测值-系数.xlsx](#)

二分类 logistic 回归分析表格

- 表中包含了参与单变量 logistic 分析的变量及样本数，OR 值和 95% 置信区间以及在筛选变量后的多因素 logistic 分析及结果（当满足有变量纳入到因素时才会有）
- 满足单因素 logistics 回归有意义的变量才纳入到多因素中进行分析，默认设置的 p 值阈值为：(p < 0.1)。
- 当单因素分析结果中所有自变量都没有统计学意义的时候，将不进行多因素 logistic 回归分析，此时可以考虑适当增加 P 值的大小（宽松度）。
- 模型预测值表格可以用来做 ROC 分析，预测值就可以代表联合指标

## 数据格式

outcome	Age	Weight loss	Sex	Grade	Stage	Score
1	80		Male	0	Stage2	100
1	82	15	Male	0	Stage1	90
0	42	15	Male	2	Stage1	90
1	57	11	Male	0	Stage2	60
1	60	0	Male	2	Stage1	90
0	74	0	Male	2	Stage2	80
1	68	10	Female	0	Stage3	60
1	71	1	Female	2	Stage3	80
1	53	16	Male	1	Stage2	80
1	61	34	Male	0	Stage3	70
1	57	27	Male	1	Stage2	80
1	68	23	Female	1	Stage3	70
1	68	5	Female	0	Stage2	90
1	60	32	Male	0		70
1	57	60	Male	0	Stage2	70
1	67	15	Male	0	Stage2	90
1	70	-5	Male	1	Stage2	100
1	63	22	Male	2	Stage3	70
1	56	10	Female	0	Stage3	60
1	57		Male	0	Stage2	80
1	67	17	Male	0	Stage2	80

表格 1: 诊断数据

A	B	C
Sex	Stage	Grade
Male	Stage1	0
Female	Stage2	1
	Stage3	2
	Stage4	

表格 2: 分类变量的顺序

- 第一列因变量（必须是二分类），缺失值不能超过第一列长度的 85%。第 1 列中分类的前后出现的顺序会被参考的顺序，先出现的分类会被当做参考组。（影响 OR 值和置信区间计算）
- 至少需要 2 列数据,最多不能超过 20 列,最少需要 20 行,最多不能超过 30000 行，样本量需要至少 4 倍以上变量数量，样本过少拟合模型效果相对较差。
- 第二列及以后为预测的变量，可以是数值类型，也可以是分类类型
  - 如果变量是数值变量，请以数值纳入，只要含有非数值（除空值）或者是无穷值外，则此列有可能没有办法纳入到分析

- 数值变量如果其分类个数 < 10 个（如 Grade 变量只有 0 1 2）则会按照等级变量来处理
- 如果变量是等级变量，建议以具体的名字纳入，比如上图中的 Stage，也可以（类似 Grade）以数字 0 1 2 的形式纳入，但是，如果以数字编码的形式纳入，如果种类超过 5 个，需要在 excel 的表 2 中设置等级参考顺序，否则该变量会以数值纳入（等级超过 8 个将没办法纳入）
- ◆ 等级变量在不同等级之间的 OR 是不同的，比如结果表格中的 Stage 变量，可以看到 Stage2 和 Stage1 与 Stage4 和 Stage3 之间的 OR 是不同的。尤其要注意不要随意对一个等级资料编码为 0 1 2 3，如果在上传数据进行了此类编码，则这个变量会被认为是数值变量而产生上述数值变量的效果而出现错误。如果是进行了数字编码的等级变量，比如图中 Grade 变量，假设我们设置了 Grade 变量的等级是 0 1 2，可以在表 2 中设定该变量的等级顺序
- 如果变量是分类变量，默认是以等级资料纳入。二分类变量以等级或者以分类资料或者数值纳入结果都是一样的。如果是多分类非等级资料，则需要以哑变量（暂不考虑）的形式纳入
- 数值变量

下方表格:（表 2-可以不提供）:

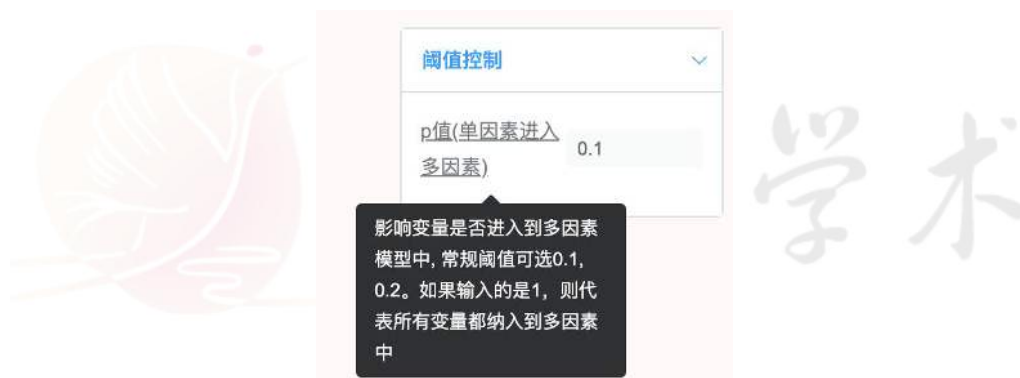
- 对应（表 1）预测变量（分类类型）中各分类的顺序
  - 比如 Stage 想要设置 Stage1, Stage2, Stage3, Stage4 的顺序，就可以如上图设置。注意，设置了等级顺序后，多因素 Logistic 回归的结果都是以第一个作为参考，其他的等级顺序与第一个等级进行对比。另外，如果在表 1 中的分类变量没有设置等级顺序，则默认以在表 1 中各个分组出现的顺序作为等级顺序。此外，如果是以 0 1 2 编码的等级变量，如果没有在这个表中进行设置，则会以数值类型纳入（可见 Grade 列）

- 如果其取值跟表 1 预测变量完全一致，则会按照其顺序对上方对应的变量分类顺序进行分析。比如 Grade 变量在表 2 中各分类的顺序为 0、1、2，与表 1 的 Grade 变量中变量名还有具体值完全一样，则会按照表 2 变量法分类的顺序进行分析，如果不是则按照表 1 中变量分类的顺序进行分析。

## 参数说明

(说明：标注了颜色的为常用参数。)

## 阈值控制



- P 值（单因素进入多因素）：单因素分析 P 值 < 阈值的变量将纳入多因素分析



## 数据处理



- **缺失值处理**：默认是单因素后多因素前处理变量缺失，也可以选择单因素分析前统一删除缺失值

## 置信区间

置信区间

计算方法

Wald方法

- 计算方法：包含有：Wald 方法、profile 方法(MASS 包)、传统计算方法。其中，Wald 方法得到的置信区间是和 SPSS 是一致的，传统计算方法为  $(OR \pm 1.96 * SE)$ ，传统计算方法对应原本生成置信区间的方式。建议选择 Wald 方法。



## 结果说明

## 主要结果

主要结果   补充结果   方法学
保存结果   下载整份报告

### 二分类Logistic

· Logistic回归: 二分类模型, 用于因变量是二分类变量时分析自变量与因变量的关联分析。  
 · P值阈值: 0.1

Characteristics	Total(N)	OR(95% CI) Univariate analysis	P value Univariate analysis	OR(95% CI) Multivariate analysis	P value Multivariate analysis
Age	228	1.039 (1.006 - 1.072)	0.020	1.009 (0.969 - 1.051)	0.609
Weight loss	214	1.006 (0.983 - 1.029)	0.609		
Sex	228				
Male	138	Reference		Reference	
Female	90	0.333 (0.183 - 0.605)	< 0.001	0.285 (0.137 - 0.595)	< 0.001
Grade	228				
0	40	Reference		Reference	
1	92	0.171 (0.021 - 1.362)	0.095	0.211 (0.026 - 1.740)	0.109
2	96	0.024 (0.003 - 0.179)	< 0.001	0.026 (0.003 - 0.205)	< 0.001
Stage	227				
Stage1	63	Reference		Reference	
Stage2	113	1.859 (0.970 - 3.560)	0.062	1.696 (0.754 - 3.818)	0.202
Stage3	51	5.270 (1.961 - 14.163)	< 0.001	4.754 (1.255 - 18.014)	0.020

Logistic回归.xlsx
Logistic回归.docx
预测值-系数.xlsx

· Characteristics: 变量以及分组

· OR(95% CI) Univariate analysis: 单因素分析得到的 OR 以及对应的置信区间。其中Reference代表等级资料的参考组, 变量中其他组在这个组进行比较

· P value Univariate analysis: 单因素分析得到的自变量对应的 p 值 (一般是满足设定的p值阈值 就会纳入到多因素模型中)

· OR(95% CI) Multivariate analysis: (只有变量满足进入多因素 Logistic 模型的 p值阈值才会有值) 多因素分析得到的 OR 以及对应的置信区间。

主要结果格式为图片格式, 提供 xlsx, docs 格式下载。另外还提供了模型预测值表格的下载。

其中预测预测值-系数中提供了：

	A	B	C	D	E	F	G	H
1	outcome	Age	Sex	Grade	Stage	Score	linear_predictors	predicted_probability
2	0	42	Male	2	Stage1	90	-0.306265778	0.424026481
3	1	80	Male	0	Stage2	100	4.197987931	0.985196653
4	1	82	Male	0	Stage1	90	3.705226356	0.975995727
5	1	57	Male	0	Stage2	60	4.057732222	0.983005621
6	1	60	Male	2	Stage1	90	-0.141710681	0.464631499
7	0	74	Male	2	Stage2	80	0.532326659	0.630025605
8	1	68	Female	0	Stage3	60	3.933613065	0.980802914
9	1	74	Female	0	Stage2	80	0.532326659	0.630025605

前面的列内容对应纳入多因素的结果，linear\_predictors 代表多因素模型中每个样本对应线性预测值，对应的就是各个变量\*系数+常量后的值。predicted\_probability 代表多因素模型给每个样本预测得到的概率值。这些内容后续可以用于比如 ROC 模块中代表联合多个指标的情况。

		riskscore		多因素模型情况		+	
	A	B					
1	多因素模型变	系数					
2	(Intercept)	3.113107027					
3	Age	0.00914195					
4	SexFemale	-1.25519315					
5	Grade1	-1.55359087					
6	Grade2	-3.64581414					
7	StageStage2	0.528547759					
8	StageStage3	1.559060308					
9	Score	-0.00175023					
10							

第 2 个表提供了多因素变量对应变量的系数情况（其中分类类型变量对应有除参考组以外其他分类的系数）

## 补充结果

### 1. 变量情况表

<div> <a href="#">主要结果</a> <a href="#">补充结果</a> <a href="#">方法学</a> </div> <div> <a href="#">保存结果</a> <a href="#">下载整份报告</a> </div>					
<b>变量情况</b>					
各个变量识别出来的类型以及是否纳入进行分析					
变量	类型	分类数量	缺失数量	是否纳入分析	补充说明
outcome	分类变量	2	0	纳入	
Age	数值变量	-	0	纳入	
Weight loss	数值变量	-	14	纳入	
Sex	分类变量	2	0	纳入	
Grade	分类变量	3	0	纳入	
Stage	分类变量	3	1	纳入	
Score	数值变量	-	3	纳入	
总样本数: 228 · 如果某个分类变量的分类>10, 将无法识别为分类变量/等级变量 · 如果变量的分组是以 0 1 2此类进行编码, 如果分类数量<5, 会被识别为分类变量; 如果>5, 会被识别为数值变量 · 如果数据中含有无穷值, 无穷值会被当做缺失处理 补充说明: 单因素分析前, 会先去掉 结局列中的缺失的样本(结局缺失的样本是无法纳入进行分析的) 缺失处理策略: 单因素后多因素前处理变量缺失					

### 2. 单因素 logistic 结果

<b>单因素Logistic</b>					
变量	类型	数量	OR	置信区间	p值
Age	数值变量	228	1.039	1.006 - 1.072	0.0205
Weight loss	数值变量	214	1.006	0.983 - 1.029	0.6088
Sex	等级变量	228			
Male		138	Reference		
Female		90	0.333	0.183 - 0.605	0.0003
Grade	等级变量	228			
0		40	Reference		
1		92	0.171	0.021 - 1.362	0.0953
2		96	0.024	0.003 - 0.179	0.0003
Stage	等级变量	227			
Stage1		63	Reference		
Stage2		113	1.859	0.970 - 3.560	0.0615
Stage3		51	5.270	1.961 - 14.163	0.0010
单因素中满足 $p < 0.1$ 就会纳入到多因素logistic回归中					

### 3. 多因素 logistic 分析

#### 多因素Logistic

模型对应二分类结局(因变量): 1 vs. 0 (其中参考组: 0)

变量	系数 $\beta$	OR	置信区间	p值
Age	0.0091419	1.009	0.969 - 1.051	0.6576
Sex				
Male		Reference		
Female	-1.2552	0.285	0.137 - 0.595	0.0008
Grade				
0		Reference		
1	-1.5536	0.211	0.026 - 1.740	0.1486
2	-3.6458	0.026	0.003 - 0.205	0.0005
Stage				
Stage1		Reference		
Stage2	0.52855	1.696	0.754 - 3.818	0.2015
Stage3	1.5591	4.754	1.255 - 18.014	0.0218
Score	-0.0017502	0.998	0.970 - 1.028	0.9056

模型常数/截距(Intercept): 3.1131

原始数据一共有228个, 变量信息缺失的样本有4个, 最终纳入的样本数: 224

备注: 如果出现纳入了多因素但是对应的统计量为空的情况, 说明(1)这个变量在去除变量信息缺失后某个分类数目过少(只有1个或者0个)或者是(2)存在严重共线性导致这个变量导致没办法计算。

备注: 当如果多因素中出现OR异常大或者异常小时(OR一般在0.33-3范围内), 说明这个变量的这个分类数量过少或者是存在共线性问题或者是数值类型变量变异比较小导致

备注: 当如果多因素中出现OR异常大或者异常小时(OR一般在0.33-3范围内), 说明这个变量的这个分类数量过少或者是存在共线性问题或者是数值类型变量变异比较小导致

·(分类/等级)变量只要某个分类的p值满足阈值就会纳入到多因素中

△ 模型全局情况:

·· AIC值: 205.917 (一般来说AIC越小, 模型拟合越好)

### 4. 方差膨胀因子统计表格

#### 方差膨胀因子(VIF)

方差膨胀因子可用于分析模型中的变量是否存在多重共线性问题

变量	类型	VIF
Age	数值变量	1.0729
Sex	等级变量	
Male		Reference
Female		1.0771
Grade	等级变量	
0		Reference
1		7.6686
2		7.6986
Stage	等级变量	
Stage1		Reference
Stage2		1.3075
Stage3		1.6947
Score	数值变量	1.3035

一般认为, 当  $0 < VIF < 10$ , 不存在多重共线性(补充: 也有认为  $VIF > 4$  就存在多重共线性); 当  $10 \leq VIF < 100$ , 存在较强的多重共线性; 当  $VIF \geq 100$  或者是出现NaN, 多重共线性非常严重

## 5. 模型评价

### 模型评价

评价方向	评价内容	统计量	p值
模型检验	似然比检验	卡方值: 76.253	7.98e-14
区分度评价	C指数	C指数: 0.845 (0.791 - 0.899)	1.09e-37
校准度评价	拟合优度检验	卡方值: 7.1254	0.5232

1. 似然比检验: 若P值小于检验水准 ( $P < 0.05$ ), 表示本次拟合的模型中至少有一个变量的OR值有统计学意义, 即模型总体有意义
2. 模型的区分度能力采用C指数来评价, 一般而言, 0.51-0.7认为是一般的准确性, 0.71-0.9为中等的准确性,  $> 0.9$ 为高度的准确性;
3. 校准度评价使用方法是Hosmer-Lemeshow Goodness of Fit 拟合优度检验, 一般而言, 若检验结果显示有统计学显著性( $P < 0.05$ ), 则表明模型预测值和实际观测值之间存在一定的差异, 模型校准度一般; 如果 $P > 0.05$ , 说明预测值与观测值没有显著差异, 因此模型拟合度较好



## SPSS 验证

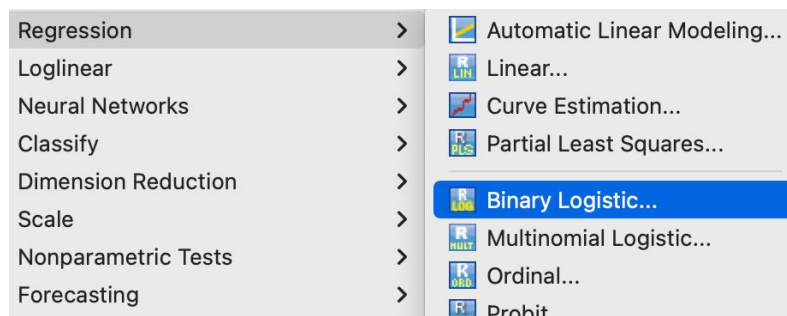
数据输入+编码：

	outcome	Age	Weight_loss	Sex	Grade	Stage	Score
1	.00	42.00	15.00	Male	2.00	1	90.00
2	1.00	80.00	.	Male	.00	2	100.00
3	1.00	82.00	15.00	Male	.00	1	90.00
4	1.00	57.00	11.00	Male	.00	2	60.00
5	1.00	60.00	.00	Male	2.00	1	90.00
6	.00	74.00	.00	Male	2.00	2	80.00
7	1.00	68.00	10.00	Female	.00	3	60.00
8	1.00	71.00	1.00	Female	2.00	3	80.00
9	1.00	53.00	16.00	Male	1.00	2	80.00
10	1.00	61.00	34.00	Male	.00	3	70.00
11	1.00	57.00	27.00	Male	1.00	2	80.00
12	1.00	68.00	23.00	Female	1.00	3	70.00
13	1.00	68.00	5.00	Female	.00	2	90.00
14	1.00	60.00	32.00	Male	.00	.	70.00

编码最好都是数字编码（因为 Sex 和 Grade 是没有缺失，不一定要编码）

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	outcome	Numeric	8	2		None	None	8	Right	Scale	Input
2	Age	Numeric	8	2		None	None	8	Right	Scale	Input
3	Weight_loss	Numeric	8	2		None	None	8	Right	Scale	Input
4	Sex	String	6	0		None	None	6	Left	Nominal	Input
5	Grade	String	8	0		None	None	8	Left	Nominal	Input
6	Stage	Numeric	6	0		None	None	6	Right	Scale	Input
7	Score	Numeric	8	2		None	None	8	Right	Scale	Input

选择二分类 logistic 回归：





选择计算置信区间：

勾选结局变量和待分析的协变量：

下面展示单因素 logistic 回归结果：

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1 <sup>a</sup>	Age	.038	.016	5.370	1	.020	1.039	1.006	1.072
	Constant	-1.378	1.011	1.859	1	.173	.252		

a. Variable(s) entered on step 1: Age.

仙桃结果（完全一致，p 0.205 是因为从原来的 0.2049 进一位约分上来的，当置信区间计算方法选择 Wald 方法置信区间也会一样）

单因素Logistic					
变量	类型	数量	OR	置信区间	p值
Age	数值变量	228	1.039	1.006 - 1.072	0.0205
Weight loss	数值变量	214	1.006	0.983 - 1.029	0.6088
Sex	等级变量	228			
Male		138	Reference		
Female		90	0.333	0.183 - 0.605	0.0003
Grade	等级变量	228			
0		40	Reference		
1		92	0.171	0.021 - 1.362	0.0953
2		96	0.024	0.003 - 0.179	0.0003
Stage	等级变量	227			
Stage1		63	Reference		
Stage2		113	1.859	0.970 - 3.560	0.0615
Stage3		51	5.270	1.961 - 14.163	0.0010

针对多分类协变量单因素结果：

Logistic Regression

Dependent: outcome

Block 1 of 1

Previous Next

Covariates: Grade(Cat)

Method: Enter

Selection Variable:

Rule...

Buttons: Categorical..., Save..., Options..., Style..., Bootstrap..., Reset, Paste, Cancel, OK

设置比较水平

**Logistic Regression: Define Categorical Variables**

**Covariates:**

**Categorical Covariates:**  
Grade(Simple(first))<

**Change Contrast**

Contrast: Simple

Reference Category: ☐ Last ☒ First

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1 <sup>a</sup>	Grade			37.896	2	.000			
	Grade(1)	-1.766	1.059	2.782	1	.095	.171	.021	1.362
	Grade(2)	-3.747	1.033	13.153	1	.000	.024	.003	.179
	Constant	1.826	.360	25.791	1	.000	6.208		

a. Variable(s) entered on step 1: Grade.

仙桃结果（完全一致，当置信区间计算方法选择 Wald 方法，置信区间也会一致）

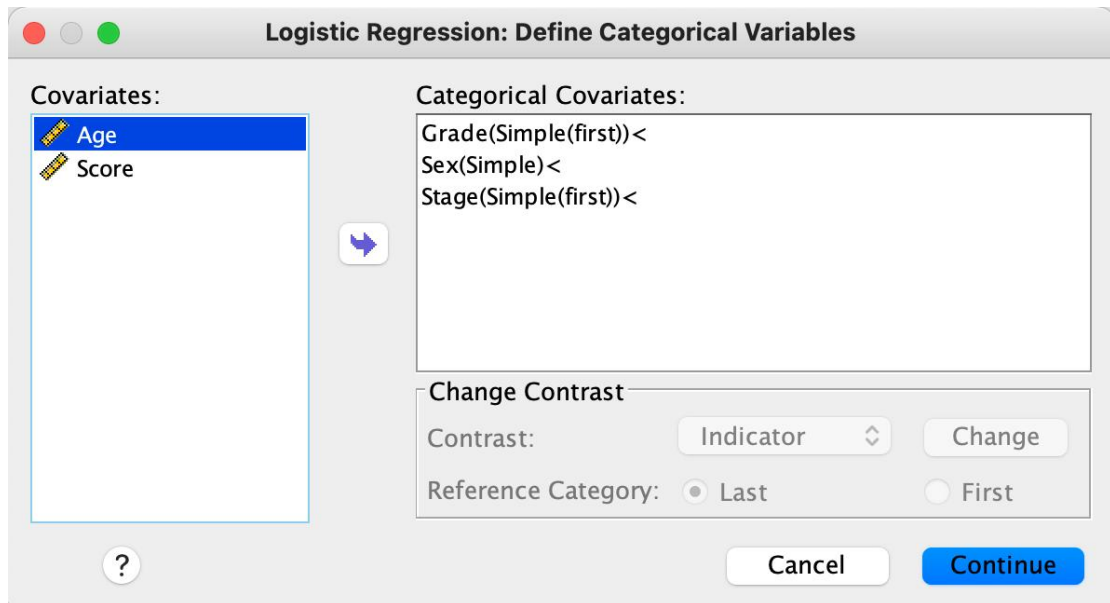
单因素Logistic

变量	类型	数量	OR	置信区间	p值
Age	数值变量	228	1.039	1.006 - 1.072	0.0205
Weight loss	数值变量	214	1.006	0.983 - 1.029	0.6088
Sex	等级变量	228			
Male		138	Reference		
Female		90	0.333	0.183 - 0.605	0.0003
Grade	等级变量	228			
0		40	Reference		
1		92	0.171	0.021 - 1.362	0.0953
2		96	0.024	0.003 - 0.179	0.0003
Stage	等级变量	227			
Stage1		63	Reference		
Stage2		113	1.859	0.970 - 3.560	0.0615
Stage3		51	5.270	1.961 - 14.163	0.0010

单因素中满足  $p < 0.1$  就会纳入到多因素logistic回归中

多因素结果：

(控制多分类变量参考组)



**Logistic Regression: Define Categorical Variables**

**Covariates:**

- Age
- Score

**Categorical Covariates:**

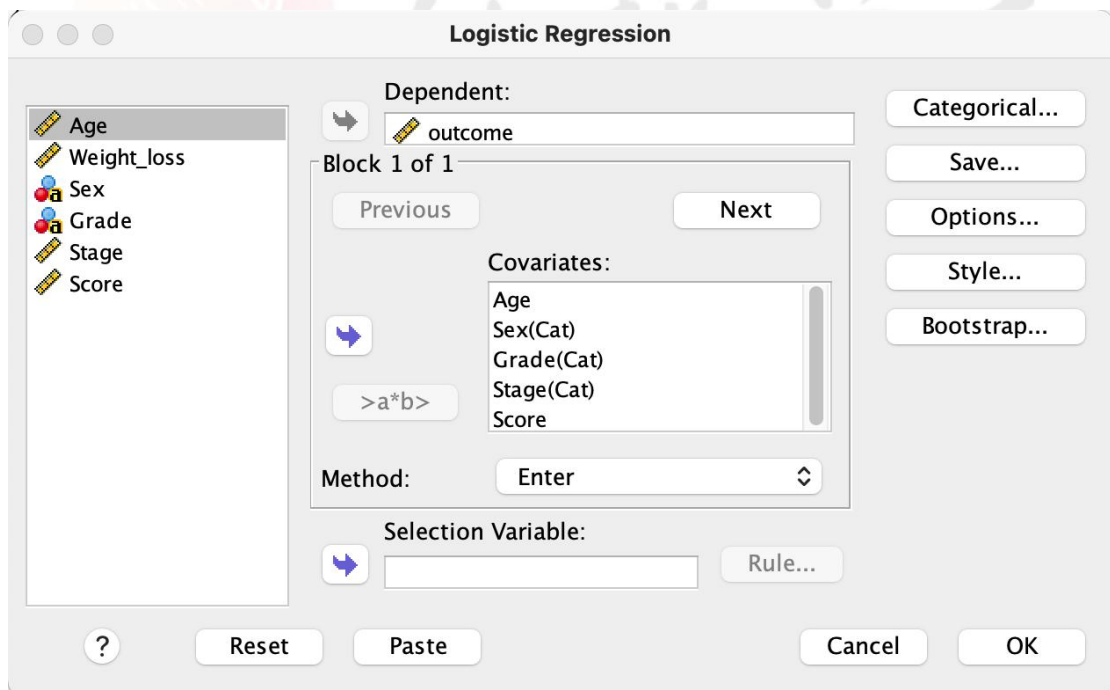
Grade(Simple(first))<  
Sex(Simple)<  
Stage(Simple(first))<

**Change Contrast**

Contrast: Indicator

Reference Category: ☒ Last ☐ First

模型变量纳入情况



**Logistic Regression**

**Dependent:** outcome

**Block 1 of 1**

**Covariates:**

- Age
- Sex(Cat)
- Grade(Cat)
- Stage(Cat)
- Score

**Method:** Enter

**Selection Variable:**

分析结果：

Variables in the Equation									
		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1 <sup>a</sup>	Age	.009	.021	.196	1	.658	1.009	.969	1.051
	Sex(1)	-1.255	.375	11.186	1	.001	.285	.137	.595
	Grade			34.746	2	.000			
	Grade(1)	-1.554	1.075	2.087	1	.149	.211	.026	1.741
	Grade(2)	-3.646	1.052	12.009	1	.001	.026	.003	.205
	Stage			5.304	2	.071			
	Stage(1)	.529	.414	1.631	1	.202	1.696	.754	3.818
	Stage(2)	1.559	.680	5.262	1	.022	4.754	1.255	18.014
	Score	-.002	.015	.014	1	.906	.998	.970	1.028
	Constant	1.448	1.799	.648	1	.421	4.256		

a. Variable(s) entered on step 1: Age, Sex, Grade, Stage, Score.

仙桃结果（基本一致，当置信区间计算方法选择 Wald 方法，置信区间也会一样）

多因素Logistic				
变量	系数β	OR	置信区间	p值
Age	0.0091419	1.009	0.969 - 1.051	0.6576
Sex				
Male		Reference		
Female	-1.2552	0.285	0.137 - 0.595	0.0008
Grade				
0		Reference		
1	-1.5536	0.211	0.026 - 1.740	0.1486
2	-3.6458	0.026	0.003 - 0.205	0.0005
Stage				
Stage1		Reference		
Stage2	0.52855	1.696	0.754 - 3.818	0.2015
Stage3	1.5591	4.754	1.255 - 18.014	0.0218
Score	-0.0017502	0.998	0.970 - 1.028	0.9056

## 方法学

统计分析和可视化均在 R 4.2.1 版本中进行

涉及的 R 包:

rms[6.4.0]用于模型拟合

ResourceSelection[0.3-5]用于拟合优度检验



## 如何引用

生信工具分析和可视化用的是 R 语言，可以直接写自己用 R 来进行分析和可视化即可，可以无需引用仙桃，如果想要引用仙桃，可以在致谢部分 (Acknowledge) 致谢仙桃学术 ([www.xiantao.love](http://www.xiantao love))。

方法学部分可以参考对应说明文本中的内容以及一些文献中的描述。



## 常见问题

### 1. logistic 回归分析中样本量多少才算够?

答：一般认为，每一个自变量至少要 10 例结局保证估计的可靠性。这里是结局例数，而不是整个样本例数。（如果你有 7 个自变量，那至少需要 70 例研究结局，否则哪怕你有 1000 例，而结局的例数只有 10 例，依然显得不足。）

### 2. 为什么有的变量 OR 值特别大或者特别小?

答：可能有以下几个原因：

- 1) 如果某个分类变量中出现某个分类数量极少的情况，则可能会导致 OR 值非常大；若是连续型自变量，可能是样本量较小或者这个自变量的变异较小（如都在 0.1-0.2 范围）。
- 2) 多重共线性问题，多重共线性会产生大的标准误。
- 3) 空单元格，如性别与疾病的关系，所有男性都发生了疾病或都没有发生疾病，这时候可能会出现 OR 值无穷大或为 0 的情形。

### 3. 为什么现在的 logistic 回归结果的 OR 值的置信区间和原来的不一样了?

答：

目前仙桃 logistic 相关的计算 OR 值的置信区间已经从原来的传统方法 `exp(summary(model)$coefficients[2,1]+1.96*summary(model)$coefficients[2,2])` 替换成了 Wald 方法来计算置信区间，并且提供了置信区间计算参数的方法（当前默认会选择 Wald 方法）（Wald 方法计算得到的结果和 SPSS 中 logistic 计算得到的置信区间是一致的），如果计算得到旧的置信区间，可以在置信区间参数中选择传统方法。



#### 4. 如何更改结局二分类中的顺序（参考组）？

**主要结果** 补充结果 方法学

**二分类Logistic**

- Logistic回归: 二分类模型, 用于因变量是二分类变量时分析自变量与因变量的关联分析。
- 模型对应二分类结局: 1 vs. 0 (其中参考组: 0)[影响OR值和置信区间]
- P值阈值: 0.1

答:

结局变量二分类的顺序按照数据中第 1 列中分类的出现顺序来, 比如下面的数据先出现的是 0, 然后是 1, 所以参考组就是取的第 1 个出现的组。

	A	B	C	D	E	F	G	H
1	outcome	Age	Weight loss	Sex	Grade	Stage	Score	
2	0	42	15	Male	2	Stage1	90	
3	1	80		Male	0	Stage2	100	
4	1	82	15	Male	0	Stage1	90	
5	1	57	11	Male	0	Stage2	60	
6	1	60	0	Male	2	Stage1	90	
7	0	74	0	Male	2	Stage2	80	

如果要调整这个顺序, 可以自己在上传数据里面把行换一下, 然后再上传数据即可:

	A	B	C	D	E	F	G	H
1	outcome	Age	Weight loss	Sex	Grade	Stage	Score	
2	1	80		Male	0	Stage2	100	
3	0	42	15	Male	2	Stage1	90	
4	1	82	15	Male	0	Stage1	90	
5	1	57	11	Male	0	Stage2	60	
6	1	60	0	Male	2	Stage1	90	
7	0	74	0	Male	2	Stage2	80	