

# 当logistic回归遇到log-binomial回归

作者：张华 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/6169.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！

## 当logistic回归遇到log-binomial回归

。当结局发生率较大时，再使用OR来估计RR时会不准确，建议当结局发生率大于10%时，使用log-binomial回归方法替代logistic回归，下面简单给大家介绍一下log-binomial回归。

我们知道logistic回归的模型是

$$\ln\left(\frac{p}{1-p}\right) = B_0 + B_i X_i$$

，而log-binomial回归只将模型作一点变化，即将 $p/(1-p)$ 换成 $p$  ( $p$ 为结局发生的概率)。模型很容易理解，但模型等号前的 $\ln p$ 是个负数，而等号后面的 $B_0 + B_i X_i$ 可以为正也可以为负，因此在回归时需要加一个限制条件即 $B_0 + B_i X_i \geq 0$ ，因此log-binomial回归是一个有偏估计。

现在我们模型一个数据看看两个回归的结果。我们设定某病的基线患病率为10%，分别模拟RR=1, 2, 3的三个数据，总人数为1000，暴露组和非暴露组各500例。如下表。当RR=1时，暴露组和非暴露组各有50人患病。

RR 值	暴露组 (病例/对照)	非暴露组 (病例/对照)
RR=1	50/450	50/450
RR=2	100/400	50/450
RR=3	150/350	50/450

分别使用logistic回归和log-binomial回归进行统计分析，得到的结果如下表。

RR	logistic 回归		Log-binomial 回归	
	OR	OR95%置信区间	PR	PR95%置信区间
RR=1	1.00	0.66, 1.51	1	0.69, 1.45
RR=2	2.25	1.56, 3.24	2	1.47, 2.77
RR=3	3.86	2.72, 5.47	3	2.25, 4.07

可以看出，虽然log-binomial回归是有偏估计，但估计值与预设的RR完全一样，而且置信区间也比较小，也即更加精确。因此当结局发生率较高，应该用log-binomial回归尝试进行统计分析。

在进行log-binomial回归时，自变量如果是连续变量时，可能造成模型不收敛，因此我们需要将数据进行加权，方法是将原数据复制后，复制数据其它不变，将结局变量0与1对换，给予原数据较高的权重(一般0.99)以上，给予复制数据较低的权重(一般0.01以下)，再进行回归分析。log-binomial回归尚不能在SPSS菜单中实现，现把程序给大家，拿走不谢。

```
proc genmod data=a;

ods select ParameterEstimates;

ods output ParameterEstimates=para0;

weight w;

model y=x/D=BIN link=LOG INterCEPT=-1LRCL;

run;

proc print data=para0;

run;

data para;

set para0;

PR=EXP(Estimate);

LPR=EXP(LowerLRCL);

UPR=EXP(UpperLRCL);

keep Parameter PR LPR UPR;
```

---

```
proc print data=para;
```

```
run;
```

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