中华流行病学杂志1987年第8卷第1期

# 江苏启东五种主要恶性肿瘤年龄别死亡率Logistic曲线拟合的理论研究

南京医学院　叶本法　陈大灵　李绍奎　潘宏斌
启东肝癌所　李文广　陈建国　柳　标

随着统计学及计算机技术的发展，各类数学模式的应用已越来越受到国内外学者的重视[1~4]。本文仅对江苏启东五种主要恶性肿瘤年龄别死亡率进行Logistic曲线拟合检验，为流行病学和病因学研究提供方法和理论依据。

## 资料来源

本资料来源于江苏启东肝癌防治研究所1980～1984年的恶性肿瘤死亡卡。经逐年复核，病例可靠；肿瘤确诊单位85%为县级以上医院，一、二级诊断率达93%。总人口数及各年龄组人口数由公安部门年终报表所得。

## Logistic回归曲线模式简介

方程式：$\hat{y}-d=\frac{k}{1+ae^{bt}}$

式中，t：自变量，为原资料自变量(T)移动座标原点所得；$\hat{y}$：因变量，即年龄组理论死亡率；d：曲线下渐近线之座标值；k：曲线上、下渐近线之距；a、b：拟合曲线的系数；e：自然对数的底。

曲线的特点是略呈“S”形状，即曲线上、下两端在接近某一限度时，上升或下降的幅度均逐渐缓慢，并形成两条相互平行的渐近线。若资料的检验值与此分布相似，即可进行Logistic曲线的拟合检验。

曲线拟合度以相关指数($R^2$)表示。公式为

$$R^2=1-\frac{\Sigma(y-\hat{y})^2}{\Sigma y^2-\frac{(\Sigma y)^2}{n}}=1-\frac{\Sigma(y-\hat{y})^2}{\Sigma(y-\bar{y})^2}$$

式中，y为各年龄组死亡率，$\hat{y}$为相应年龄组的理论死亡率，n为年龄组数。$R^2$愈接近于1，说明实际数与理论数愈接近，曲线拟合得愈好。

## 方法举例

以男性胃癌年龄别死亡率的Logistic曲线拟合为例。

先将启东县1980～1984年男性胃癌年龄别死亡率检验数据在方格纸上绘图，各点连线略呈“S”形。目测上、下渐近线，并尽量使$\frac{[k-(y-d)]}{(y-d)}=Z$的对数值线化，以确定合适的的k与d值，得：k=450，d=0。当Z=1时，相当T值为61.5。令t=T-61.5，从目测直线上取便于读数的两点：$P_1$(76,0.1)，$P_2$(325,100)。以t值代替T值，求直线坡度m。

$$m=\frac{\log Z_2-\log Z_1}{t_2-t_1}=\frac{2-(-1)}{-29-14.5}=-0.0690$$

$$b=\log e10.m=2.30258\times(-\frac{3}{43.5})=-0.1588$$

$$a=\text{Antilog}[\log Z_2-t_2m]$$
$$=\text{Antilog}[2-(-29)(-\frac{3}{435})]=1$$

将k=450，d=0，a=1，b=-0.1588代入方程式，解开方程，得各年龄组死亡率$\hat{y}$。

进行方差分析(表1)。表1中：

$$\Sigma(y-\bar{y})^2=\Sigma y^2-\frac{(\Sigma y)^2}{n}$$
$$=637997.76-\frac{3735986.3}{17}=418233.85;$$

$$\Sigma(y-\hat{y})^2=3729.47$$

**表 1** 启东县1980～1984年男性胃癌年龄别死亡率的方差分析

| 变异来源 | df | SS | MS | F | P |
|---|---|---|---|---|---|
| 总$\Sigma(y-\bar{y})^2$ | 16 | 418233.85 | | | |
| 估计误差 | | | | | |
| $\Sigma(y-\hat{y})^2$ | 15 | 3729.47 | 248.63 | | |
| 减少的误差 | 1 | 414504.38 | 414504.38 | 1667.15 | <0.01 |

即拟合曲线后y方面误差从$\Sigma(y-\bar{y})^2$降至$\Sigma(y-\hat{y})^2$有非常显著的意义。

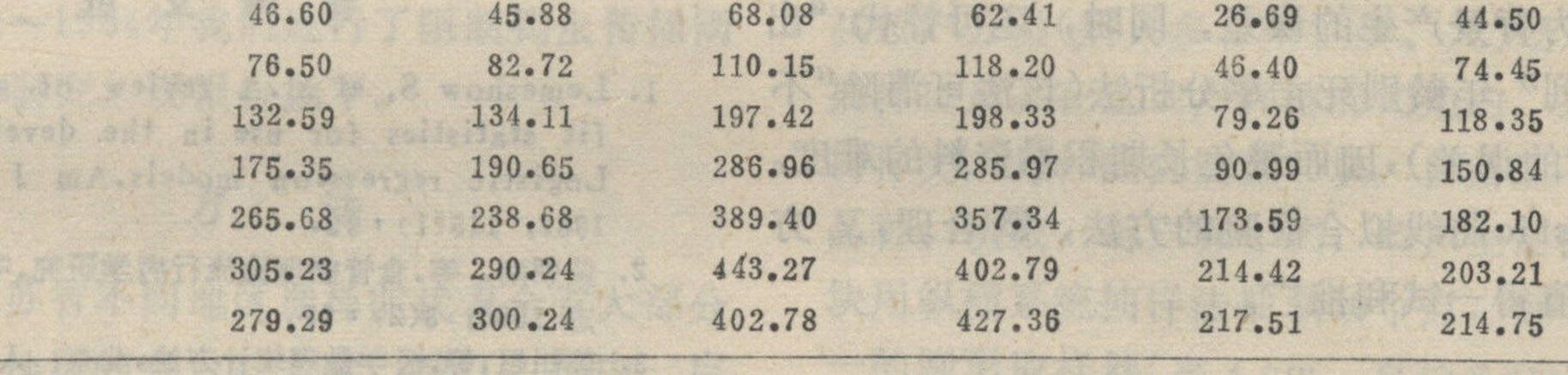
$$R^2=1-\frac{3729.47}{418233.85}=0.99$$

## 结果与讨论

一、启东县1980～1984年胃癌、肺癌、食管癌年龄别死亡率Logistic曲线拟合见表 2 和附图。

**表 2** 启东县1980～1984年胃癌年龄别死亡率Logistic曲线拟合

| 年龄组（岁） | 男＋女 | | 男 | | 女 | |
|---|---|---|---|---|---|---|
| | 实测值 | 理论值 | 实测值 | 理论值 | 实测值 | 理论值 |
| 0～ | 0.30 | 0.03 | 0.59 | 0.03 | | |
| 5～ | 0 | 0.07 | 0 | 0.06 | | |
| 10～ | 0.15 | 0.15 | 0.30 | 0.13 | | |
| 15～ | 0.45 | 0.30 | 0.59 | 0.28 | 0.30 | 0.44 |
| 20～ | 0.54 | 0.64 | 0.37 | 0.62 | 0.70 | 0.87 |
| 25～ | 1.30 | 1.33 | 1.89 | 1.36 | 0.73 | 1.73 |
| 30～ | 2.91 | 2.77 | 2.45 | 3.01 | 3.36 | 3.43 |
| 35～ | 5.58 | 5.75 | 4.98 | 6.59 | 6.18 | 6.74 |
| 40～ | 14.35 | 11.80 | 13.56 | 14.33 | 15.13 | 13.07 |
| 45～ | 24.52 | 23.74 | 32.88 | 30.53 | 16.84 | 24.68 |
| 50～ | 46.60 | 45.88 | 68.08 | 62.41 | 26.69 | 44.50 |
| 55～ | 76.50 | 82.72 | 110.15 | 118.20 | 46.40 | 74.45 |
| 60～ | 132.59 | 134.11 | 197.42 | 198.33 | 79.26 | 118.35 |
| 65～ | 175.35 | 190.65 | 286.96 | 285.97 | 90.99 | 150.84 |
| 70～ | 265.68 | 238.68 | 389.40 | 357.34 | 173.59 | 182.10 |
| 75～ | 305.23 | 290.24 | 443.27 | 402.79 | 214.42 | 203.21 |
| 80～ | 279.29 | 300.24 | 402.78 | 427.36 | 217.51 | 214.75 |
| 公式 | $\hat{y}=\frac{310}{1+1.0049e^{-0.1479t}}$ | | $\hat{y}=\frac{450}{1+e^{-0.1588t}}$ | | $\hat{y}=\frac{220}{1+1.0471e^{-0.1382t}}$ | |
| P值 | <0.01 | | <0.01 | | <0.01 | |
| $R^2$ | 0.98 | | 0.99 | | 0.94 | |




附图 启东县1980～1984年胃癌、肺癌、食道癌年龄别死亡率Logistic拟合曲线

本资料Logistic曲线拟合结果表明：启东胃癌、肺癌、食管癌各年龄别死亡率男性均高于女性。从理论曲线的上升趋势及速度看，各肿瘤男女性死亡率基本上均以50～75岁组呈直线峻陡上升，而男性尤剧。曲线的拟合度都在0.92～0.99之间，且方差分析具有高度显著差异，说明其理论曲线对这几种肿瘤年龄别死亡率具有一定的代表性，为今后有关肿瘤病因调查研究，建立起理论分析的基础。

二、肝癌年龄别死亡率无“S”型分布趋势，故未进行Logistic曲线拟合检验。鼻咽癌死亡率虽有随年龄增长而升高的趋势，但拟合度不佳（$R^2_{男}=0.60$、$R^2_{女}=0.47$、$R^2_{男+女}=0.82$），这可能与样本数较少有关，有待今后继续研究。

三、通过分析比较，Logistic曲线不仅比一般频数曲线的描述方法更为合理，可推算出任一年龄组的理论死亡率，还可避免在横断面资料年龄别分析时，尽管用“标化”处理，解决了“同代”分布的误差，却仍存在着“不同代”为背景产生的误差。同时，还可替代“出生队列”年龄别死亡率分析法（该法可消除“不同代”的误差），因而避免长期积累资料的难度。Logistic曲线拟合检验的方法，既合理，又方便，值得一试和推广。

## 摘　要

本文通过对启东县1980～1984年胃癌、肺癌、食管癌年龄别死亡率Logistic曲线的拟合检验，明确显示了这三种恶性肿瘤死亡率的年龄变动趋势，为今后对这三种癌症的流行病学和病因学研究，建立了理论基础，并揭示了Logistic曲线拟合比一般频数描述的方法更为合理和方便。本文同时指出了肝癌年龄别死亡率不合于Logistic曲线分布，鼻咽癌因例数少，Logistic曲线拟合不佳，有待后证。

**A Theoretical Study on Goodness-fit of Logistic Curve of Age-mortality of Five Main Cancer in Qidong, Jiangsu** *Ye Benfa, et al., Nanjing Medical College*

This paper described tests to goodness-fit of logistic curve of ge-mortality of stom ch cancer, lung cancer and esoph geal cancer in 1980-1984 years, Qidong Cornty. We found the tendency of age-variable of mortality in these cancers, and expounded goodness-fit of logistic curve which was simple and fit method building the theoretical foundertion for study on the epidemiologic l and etiological investigation of cancer. This data also pointed out that the age-mortality of liner c ncer was not appropriated the distribution of logistic curve and goodness-fit of logistic curve w s not good for a few of c ses o f nosopharingeal cancer.

## 参 考 文 献

1. Lemeshow S, et al. A review of goodness of fit statistics for use in the development of Logistic regression models. Am J Epidemiol 1982; 115(1)：92.
2. 黄朝南，等.食管癌理论流行病学研究.中华流行病学杂志 1983; 3(2)：97.
3. 郭祖超，等.医学数理统计方法.北京：人民卫生出版社，1963：452.
4. 上一医卫生统计教研组. 医学统计方法. 上海：上海科技出版社，1979：98.

**《腹泻病专辑》（二）征订启事**

“腹泻病专题委员会”和本刊再度联合编辑出版《腹泻病专辑》（二）。

该书辑入本刊收到和向有关专家征集的论著、技术方法、综述，还有全国人兽腹泻病病原学学术讨论会的部分材料，共计六十余篇。全书约20万字，售价2.0元（含邮费），预计于1987年3～4月出版，一律从邮局汇款：北京昌平流字五号　中华流行病学杂志编辑部。注明“购腹泻病专辑(二)”，征订截至1987年2月底止。过期不保证供书。

中华流行病学杂志编辑部