

## (1) 在第四讲拟合算法第 10 页后补充下面两张 PPT

清风"数学建模算法讲解"全套视频, 单人购买观看仅需58元

### "线性函数"的介绍 $R^2$ 只能用于拟合函数是“线性函数”时, 拟合结果的评价

思考:  $y = a + bx^2$  是线性函数吗?

是的, 因为我们这里说的线性函数是指**对参数为线性 (线性于参数)**。

由于本书主要讨论像方程 (2.2.2) 那样的线性模型, 所以我们必须知道线性一词的真正含义, 因为对它可作两种解释。

#### □ 对变量为线性

对线性的第一种并且也许是更“自然”的一种解释是,  $Y$  的条件期望值是  $X_i$  的线性函数, 比如说, 方程 (2.2.2)。<sup>①</sup> 从几何意义上说, 这时回归曲线是一条直线。按照这种解释, 诸如  $E(Y | X_i) = \beta_1 + \beta_2 X_i^2$  的回归函数, 由于变量  $X$  以幂或指数 2 出现, 就不是线性的。

#### □ 对参数为线性

对线性的第二种解释是,  $Y$  的条件期望  $E(Y | X_i)$  是参数  $\beta$  的一个线性函数; 它可以是或不是变量  $X$  的线性函数。<sup>②</sup> 对于这种解释,  $E(Y | X_i) = \beta_1 + \beta_2 X_i^2$  就是一个线性 (于参数) 回归模型。为了看出这一点, 让我们假设  $X$  取值为 3。因此,  $E(Y | X=3) = \beta_1 + 9\beta_2$ , 显然它是  $\beta_1$  和  $\beta_2$  的线性函数。图 2—3 中所示的所有模型因此也都是线性回

参考资料: 古扎拉蒂《计量经济学基础》第五版

数学建模学习交流  
11 / 21

关注微信公众号“数学建模学习交流”获取更多优质资料

清风"数学建模算法讲解"全套视频, 单人购买观看仅需58元

### 如何判断线性于参数的函数?

在函数中, 参数仅以一次方出现, 且不能乘以或除以其他任何的参数, 并不能出现参数的复合函数形式。

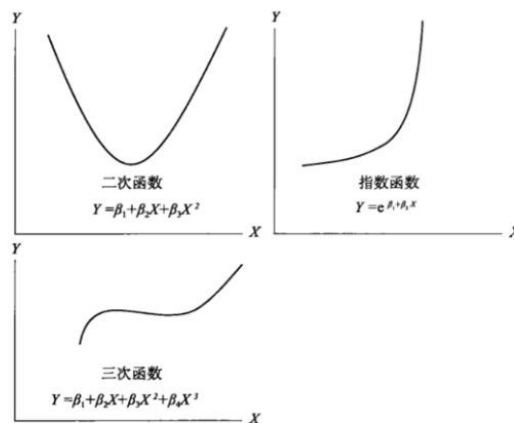


图 2—3 线性于参数的函数

$y = a + b^3 x$ 、 $y = a + bx + bcx^2$ 、 $y = a(x - b)^2$ 、 $y = a \sin(b + cx)$  都不是线性函数, 不能用  $R^2$ 。

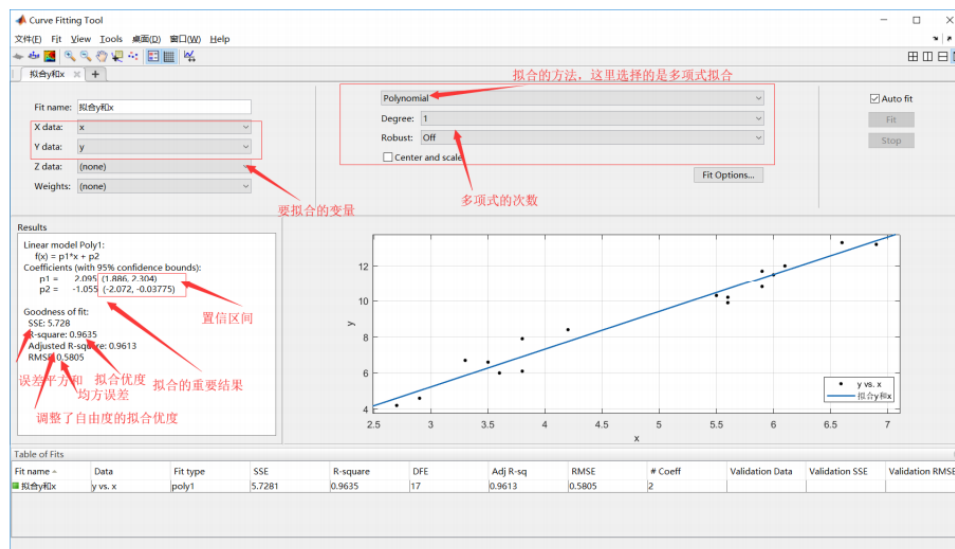
数学建模学习交流  
12 / 21

关注微信公众号“数学建模学习交流”获取更多优质资料

## 在拟合工具箱演示这页 PPT 后新增加了一个练习题

清风“数学建模算法讲解”全套视频，单人购买观看仅需58元

### 拟合工具箱演示



数学建模学习交流

关注微信公众号“数学建模学习交流”获取更多优质资料

15 / 22

清风“数学建模算法讲解”全套视频，单人购买观看仅需58元

### 利用拟合工具箱预测美国人口

population\_predict.m

下表给出了近2个世纪的美国人口统计数据（单位：百万人），请使用最下面给定的拟合函数预测后30年的美国人口。

年	1790	1800	1810	1820	1830	1840	1850	1860
人口	3.9	5.3	7.2	9.6	12.9	17.1	23.2	31.4
年	1870	1880	1890	1900	1910	1920	1930	1940
人口	38.6	50.2	62.9	76.0	92.0	106.5	123.2	131.7
年	1950	1960	1970	1980	1990	2000		
人口	150.7	179.3	204.0	226.5	251.4	281.4		

$$x(t) = \frac{x_m}{1 + \left(\frac{x_m}{3.9} - 1\right)e^{-r(t-1790)}}$$

$x_m$ 和 $r$ 是两个拟合参数， $t$ 表示年份， $x(t)$ 表示第 $t$ 年的人口

数学建模学习交流

关注微信公众号“数学建模学习交流”获取更多优质资料

16 / 22

## (2) 主成分分析的 ppt 第 10 页中少打了平方

清风"数学建模算法讲解"全套视频, 单人购买观看仅需58元

### PCA的计算步骤

假设有  $n$  个样本,  $p$  个指标, 则可构成大小为  $n \times p$  的样本矩阵  $x$ :

$$x = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} = (x_1, x_2, \cdots, x_p)$$

1. 我们首先对其进行标准化处理:

按列计算均值  $\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$  和标准差  $S_j = \sqrt{\frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n-1}}$  计算得

标准化数据  $X_{ij} = \frac{x_{ij} - \bar{x}_j}{S_j}$ , 原始样本矩阵经过标准化变为:

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix} = (X_1, X_2, \cdots, X_p)$$

关注微信公众号"数学建模学习交流"获取更多优质资料

数学建模学习交流 10/34

## (3) 更新 2 因子分析模型中语句不通顺

清风"数学建模算法讲解"全套视频, 单人购买观看仅需58元

### KMO检验和巴特利特球形检验

#### KMO检验

KMO检验是 Kaiser, Meyer 和 Olkin 提出的, 该检验是对原始变量之间的简单相关系数和偏相关系数的相对大小进行检验, 主要应用于多元统计的因子分析。

KMO统计量是取值在0和1之间, 当所有变量间的简单相关系数平方和远远大于偏相关系数平方和时, KMO值越接近于1, 意味着变量间的相关性越强, 原有变量越适合作因子分析; 当所有变量间的简单相关系数平方和接近0时, KMO值越接近于0, 意味着变量间的相关性越弱, 原有变量越不适合因子分析。

其中, Kaiser给出一个KMO检验标准:  $KMO > 0.9$ , 非常适合;  $0.8 < KMO < 0.9$ , 适合;  $0.7 < KMO < 0.8$ , 一般;  $0.6 < KMO < 0.7$ , 不太适合;  $KMO < 0.5$ , 不适合。

#### 巴特利特球形检验

巴特利特球形检验是一种检验各个变量之间相关性程度的检验方法。一般在做因子分析之前都要进行巴特利特球形检验, 用于判断变量是否适合用于做因子分析。巴特利特球形检验是以变量的相关系数矩阵为出发点的。它的原假设是相关系数矩阵是一个单位阵 (不适合做因子分析, 指标之间的相关性太差, 不适合降维), 即相关系数矩阵对角线上的所有元素都是1, 所有非对角线上的元素都为0。巴特利特球形检验的统计量是根据相关系数矩阵的行列式得到的。如果该值较大, 且其对应的  $p$  值小于用户心中的显著性水平 (一般为0.05), 那么应该拒绝原假设, 认为相关系数不可能是单位阵, 即原始变量之间存在相关性, 适合于因子分析。相反不适合因子分析。

注意: 用SPSS做因子分析时, 在查看器中若得不到KMO检验和Bartlett检验结果, 则说明你的样本量小于指标数了, 需要增加样本量或者减少指标个数再来进行因子分析。

关注微信公众号"数学建模学习交流"获取更多优质资料

数学建模学习交流 18/27