V1

第12章
非线性回归模型

方匡南 朱建平 姜叶飞

非线性回归分析是线性回归分析的扩展，由于非线性回归的参数估计涉及非线性优化问题，计算比较困难，因此在计算机诞生前较少研究。20世纪七八十年代以来，随着计算机技术的发展，非线性回归的参数估计计算困难得到了克服，统计推断和预测分析技术也有很大发展。非线性回归分析也开始受到更多的重视，现在已经成为统计学、计量经济学研究的热点之一。

本章主要可线性化的非线性回归估计，不可线性化非线性回归的估计、应用以及R语言的实现。

01 问题的提出

12.1 问题的提出

我们知道线性回归模型的变量关系只是一种特例，现实中的变量关系大多是非线性的。关于非线性回归，不同的教材有不同的定义，有些教材定义只要因变量和自变量之间的关系是非线性就称为非线性回归，比如模型 $y_t = ae^{bx_t + u_t}$ 中 y_t 与 x_t 是非线性关系，称该模型为非线性回归；有些教材定义参数求解是非线性才称为非线性回归，比如模型 $y_t = ae^{bx_t + u_t}$ 中 y_t 与 x_t 是虽然非线性关系，但是通过变换后 $\ln y_t = \ln a + bx_t + u_t$ ，此时对参数 $\ln a$ 和 b 的求解是线性的，因此从参数求解的定义来看，该模型被归为线性回归，另外比如形如 $\mathbf{Y} = A\mathbf{K}^{\alpha}\mathbf{L}^{\beta} + \boldsymbol{\varepsilon}$ 的非线性回归，就没法通过数学变换转换为线性回归，因此只能通过迭代算法等进行参数求解。

本书为了让数据分析初学者更加全面地了解非线性回归，因此本书分两部分来讲解，先讲解可线性化的非线性回归，然后再讲解不可先线性化的非线性回归。

12 非线性回归

- Cobb-Douglas生产函数
- 多项式方程模型
- 指数函数模型

可线性化的非线性回归

02

12.2.1 Cobb-Douglas生产函数

经济中著名的 Cobb-Douglas 生产函数为

$$Q = KL^{\alpha}C^{1-\alpha}$$

其中，Q 表示产量；L 表示劳动力投入量；C 表示资本投入量；K 是常数；$0 < \alpha < 1$。

这种生产函数是美国经济学家柯布和道格拉斯根据 1899-1922 年美国关于生产方面的数据研究得出的。更习惯的表达形式是

$$y_t = \beta_0 x_{t1}{}^{\beta_1} x_{t2}{}^{\beta_2} e^{u_t} \qquad (12\text{-}1)$$

12.2.1 Cobb-Douglas生产函数

模型（12-1）中的 y_t 与 x_t 是非线性关系，因此无法用 OLS 法直接估计，但可先作线性化处理。对（12-1）式的两边同取对数，得：

$$\ln y_t = \ln \beta_0 + \beta_1 \ln x_{t1} + \beta_2 \ln x_{t2} + \mu_t \qquad (12\text{-}2)$$

令 $y_t^* = \ln y_t$ ，$\beta_0^* = \ln \beta_0$ ，$x_{t1}^* = \ln x_{t1}$ ，$x_{t2}^* = \ln x_{t2}$ ，有 $y_t^* = \beta_0^* + \beta_1 x_{t1}^* + \beta_2 x_{t2}^* + \mu_t$

上式为线性模型，可用 OLS 法估计后，再将参数代入到原模型。若回归参数 $\beta_1 + \beta_2 = 1$ ，称模型为规模报酬不变型；$\beta_1 + \beta_2 > 1$ ，称模型为规模报酬递增型；$\beta_1 + \beta_2 < 1$ ，称模型为规模报酬递减型。

12.2.1 Cobb-Douglas生产函数

例 12-1：道格拉斯（Cobb-Douglas）生产函数

假设随机道格拉斯生产函数的模型为：$y_t = \beta_0 x_{t1}{}^{\beta_1} x_{t2}{}^{\beta_2} e^{u_t}$ ，$t = 1,\ldots,T$ ，其中 y_t 为产出，x_{t1} 为劳动投入，x_{t2} 为资本投入，μ_t 为随机干扰项。显然随机道格拉斯生产函数模型为非线性模型，需对方程两边取对数转化为线性模型 $\ln y_t = \ln \beta_0 + \beta_1 \ln x_{t1} + \beta_2 \ln x_{t2} + \mu_t$。收集了 1958 - 1972 年中国台湾地区农业部门的数据研究台湾农业部门的生产函数，详见表 12-1。

表12-1 中国台湾地区农业部门的数据

年份	实际总产值y	劳动天数x2	实际资本投入x3
1958	16607.7	275.5	17803.7
1959	17511.3	274.4	18096.8
1960	20171.2	269.7	18271.8
1961	20932.9	267	19167.3
1962	20406	267.8	19647.6
1963	20831.6	275	20803.5
1964	24806.3	283	22076.6
1965	26465.8	300.7	23445.2
1966	27403	307.5	24939
1967	28628.7	303.7	26713.7
1968	29904.5	304.7	29957.8
1969	27508.2	298.6	31585.9
1970	29035.5	295.5	33474.5
1971	29281.5	299	34821.8
1972	31535.8	288.1	41794.3

12.2.1 Cobb-Douglas生产函数

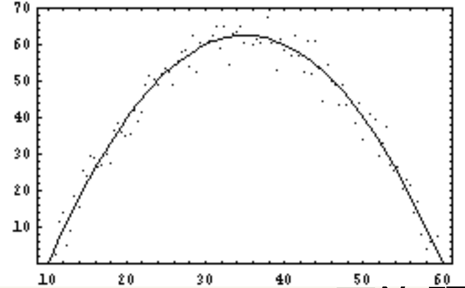
从估计的结果可以看出1958 - 1972年台湾地区农业部门产出的劳动和资本弹性分别是1.4988和0.4899。即当保持资本投入不变，劳动投入增加1%，导致产出平均增加1.4988%；当保持劳动投入不变时，资本投入增加1%，导致产出平均增加0.4899%。两个弹性相加值为1.9887，就是规模报酬参数的取值，说明是规模报酬递增的。

代回去可得 $\beta_0 = e^{-3.3385} = 0.0355$

可得柯布-道格拉斯（Cobb-Douglas）生产函数回归模型为

$$\hat{y}_i = 0.0355 X_{2i}^{1.4988} X_{3i}^{0.4899}$$

```
>dat=read.csv(file="douglas.csv")
> lm1=lm(log(y)~log(x2)+log(x3),data=dat)
> summary(lm1)
Call:
lm(formula = log(y) ~ log(x2) + log(x3), data = dat)
Residuals:
     Min       1Q   Median       3Q      Max
-0.15920 -0.02914  0.01179  0.04087  0.09640
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -3.3385     2.4495  -1.363 0.197939
log(x2)       1.4988     0.5398   2.777 0.016758 *
log(x3)       0.4899     0.1020   4.800 0.000433 ***
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.07481 on 12 degrees of freedom
Multiple R-squared: 0.889,     Adjusted R-squared: 0.8705
F-statistic: 48.07 on 2 and 12 DF,  p-value: 1.867e-06
```



项式方程模型

二次项多项式方程的表达形式是

$$y_t = b_0 + b_1 x_t + b_2 x_t^2 + \mu_t \qquad (12\text{-}3)$$

其中$b_1 > 0$，$b_2 > 0$和$b_1 < 0$，$b_2 < 0$情形的图形分别见图 12-1 和 12-2。令$x_{2t} = x_t^2$

上式线性化为$y_t = b_0 + b_1 x_t + b_2 x_{2t} + \mu_t$。经济学中的边际成本曲线、平均成本曲线与图 12-1 相似。

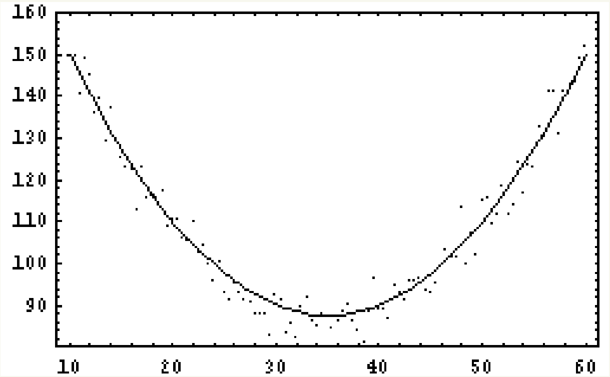


图 12-1 式（12-3）曲线图

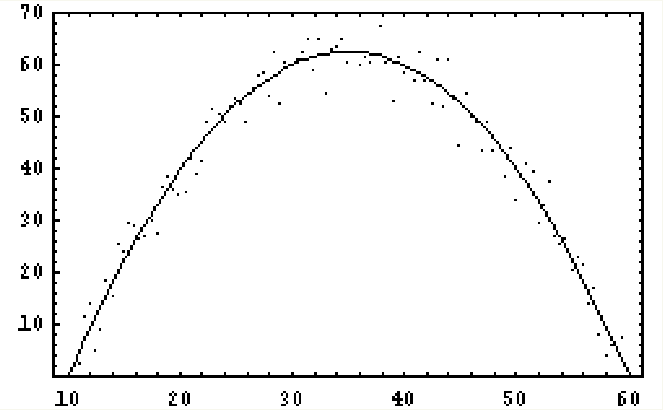


图 12-2 式（12-3）曲线图

12.2.2　多项式方程模型

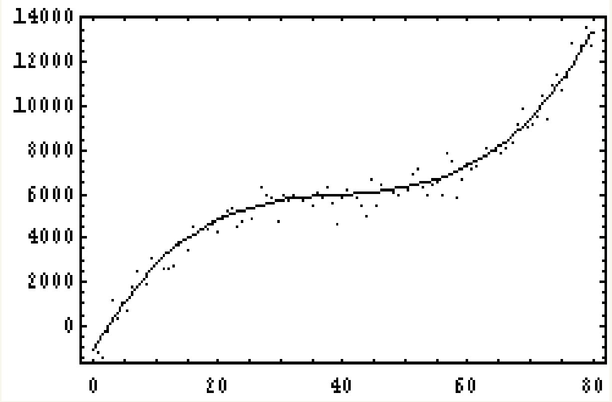


图 12-3 式（12-4）的系数为正时曲线图

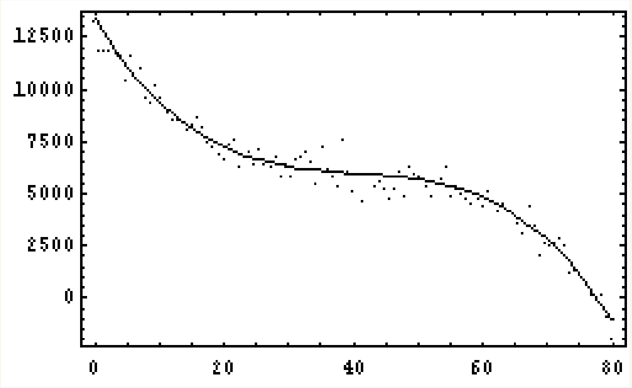


图 12-4 式（12-4）系数为负时的曲线图

三次项多项式方程的表达形式是

$$y_t = b_0 + b_1 x_t + b_2 x_t^2 + b_3 x_t^3 + \mu_t \qquad (12\text{-}4)$$

其中 $b_1 > 0$，$b_2 > 0$，$b_3 > 0$ 和 $b_1 < 0$，$b_2 < 0$，$b_3 < 0$ 情形的图形分别见图 12-3 和 12-4。令 $x_{2t} = x_t^2$，$x_{3t} = x_t^3$，上式变为 $y_t = b_0 + b_1 x_t + b_2 x_{2t} + b_3 x_{3t} + \mu_t$，这是一个三元线性回归模型。如经济学中的总成本曲线与图 12-3 相似。

例12-2 多项式回归

1609年，伽利略证明了一个物体在一个水平力的作用下，其下落轨道为一抛物线。为了验证这一事实，他做了一项实验并度量了两个变量：高度和距离，数据如下：

高度	100	200	300	400	600	800	1000
距离	253	337	395	451	295	534	574

12.2.2　多项式方程模型

通过数据描点，伽利略显然看到数据分布呈抛物线，且在数学上证明了它。在现代的眼光看来，如果确信为抛物线，我们可用二次回归模型得到那些系数。

```
> x = c（100,200,300,450,600,800,1000）
> y = c（253, 337,395,451,495,534,574）
> lm.1 = lm（y ~ x）                # 一次模型 y=a+bx
> lm.2 = lm（y ~ x + I（x^2））      # 二次模型 y=a+bx+cx²
> lm.3 = lm（y ~ x + I（x^2） + I（x^3））  # 三次模型 y=a+bx+cx²+dx³
```

下面用summary给出三种模型的拟合结果：

```
> summary（lm.1）$coef
              Estimate   Std. Error  t value  Pr（>|t|）
（Intercept）  269.4661    24.18421   11.142   0.0001015
x              0.3341     0.04181    7.992    0.0004951
> summary（lm.2）$coef
              Estimate    Std. Error  t value  Pr（>|t|）
（Intercept）  200.211950  1.695e+01   11.811   0.0002941
x              0.706182   7.568e-02   9.332    0.0007342
I（x^2）       -0.000341  6.754e-05   -5.049   0.0072374
> summary（lm.3）$coef
              Estimate   Std. Error  t value  Pr（>|t|）
（Intercept）  1.555e+02  8.182e+00   19.003   0.0003182
x             1.119e+00  6.454e-02   17.332   0.0004185
I（x^2）      -1.254e-03 1.360e-04   -9.220   0.0026986
I（x^3）       5.550e-07 8.184e-08   6.782    0.0065519
```

12.2.2　多项式方程模型

注意必须使用结构I（x^2），函数I()允许我们使用通常的幂符号。可尝试用二次和三次曲线去拟合数据，绘出图形

```
> plot（x,y）
> lines（x,fitted（lm.1）,lty=1）
> lines（x,fitted（lm.2）,lty=2）
> lines（x,fitted（lm.3）,lty=3）
> legend（700,400,c（"直线","二次曲线","三次曲线"）,lty=1:3）
```

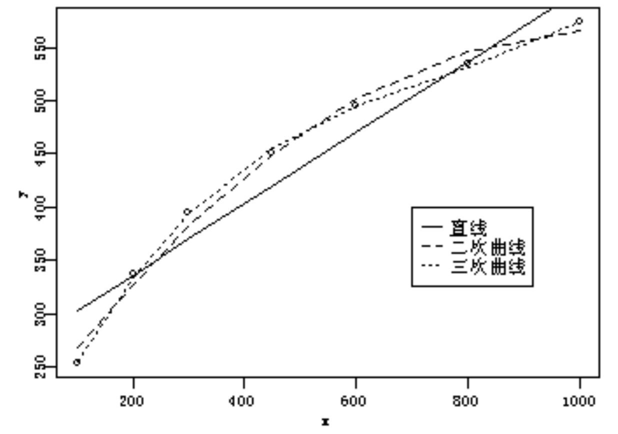


图12-5 伽俐略数据的一次、二次和三次拟合

两条曲线似乎拟合得都不错，选择哪一条呢？可以比较它们的可决系数（R^2），哪个最大，哪个模型相对来说最好。

```
> summary（lm.1）$r.squared
  [1] 0.9274
> summary（lm.2）$r.squared
  [1] 0.9902
> summary（lm.3）$r.squared
  [1] 0.9994
```

从中可以看出，相对来说，三次曲线的拟合效果最好。

12.2.3 指数函数模型

模型 $y_t = ae^{bx_t + u_t}$ 的因变量和自变量之间的关系是一种指数函数关系，其中， $b>0$ 和 $b<0$ 两种情形的图形分别见图 12-6 和图 12-7。对上式等号两侧同取自然对数，得 $\ln y_t = \ln a + bx_t + u_t$。令 $y_t^* = \ln y_t$， $a^* = \ln a$，则模型变换为 $y_t^* = a^* + bx_t + u_t$，对于参数 a^* 和 b 的求解可以通过线性回归的 OLS 得到，然后求 $a = e^{a^*}$ 得到。

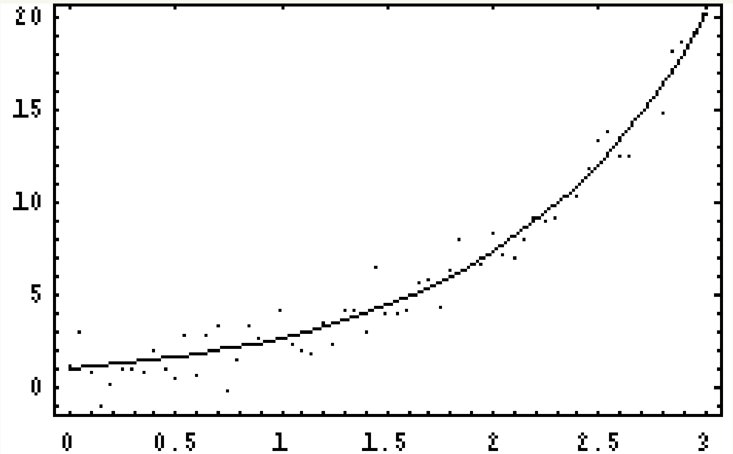


图 12-6　$y_t = ae^{bx_t + u_t}, (b>0)$

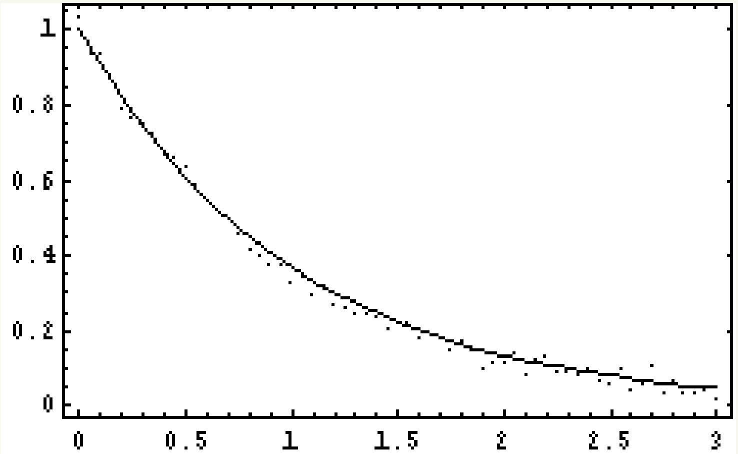


图 12-7　$y_t = ae^{bx_t + u_t}, (b<0)$

除此之外，还有半对数模型 $y_t = a + b\ln x_t + \mu_t$ ，双曲线模型 $1/y_t = a + b/x_t + \mu_t$ 等，这些模型的共同特点都是可以通过数学变换后转换为线性回归，然后我们可以利用线性回归方法求解参数，所以从参数求解上来看，本质上还是线性回归。

- 非线性模型的参数估计与迭代算法
- 初始值选取
- 收敛性

03

不可线性化的非线性回归

12.3.1 非线性模型的参数估计与迭代算法

参数估计是非线性回归分析的核心。非线性回归分析参数估计的方法也有多种，主要方法有非线性最小二乘估计(nonlinear least squares estimator, NLS)和非线性最大似然估计(Nonlinear maximum likelihood estimator, NMLE)。当非线性模型的随机误差项$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n)$时，除了误差方差的估计以外，非线性回归的最大似然估计与最小二乘估计也是相同的。

实际上，非线性最小二乘估计和非线性极大似然估计都是非线性优化的问题，非线性优化有多种算法，比如有格点搜索法、二次爬坡法、高斯牛顿法和牛顿拉夫森法等。

12.3.1 非线性模型的参数估计与迭代算法

（1）高斯－牛顿法

高斯－牛顿法（Gauss-Newton）的基本思路是：非线性最小二乘估计的问题在于最小二乘函数 $S(\boldsymbol{\beta})$ 中的 f ，也就是回归模型

$$\mathbf{Y} = f(\mathbf{X}, \boldsymbol{\beta}) + \boldsymbol{\varepsilon}$$

的趋势部分不是参数向量 $\boldsymbol{\beta}$ 的线性函数，因此最优化问题

$$\min_{\hat{\beta}_1, \cdots, \hat{\beta}_p} S(\boldsymbol{\beta}) = [\mathbf{Y} - f(\mathbf{X}, \hat{\boldsymbol{\beta}})]'[\mathbf{Y} - f(\mathbf{X}, \hat{\boldsymbol{\beta}})] \qquad (12\text{-}17)$$

的求解存在计算上的困难。当 f 是连续可微时，可以在某组参数初始值处作一阶泰勒级数展开，得到 f 的线性近似，把这个线性近似函数代入最小二乘函数得到参数的二次函数，克服参数估计计算的困难。但一阶泰勒级数展开得到的近似函数与原函数是有差异的，用上述级数展开近似的方法得到的参数估计也有偏差，偏差程度与泰勒级数展开的初始值与参数真实值的偏差相关。提高参数估计准确程度的途径是改进泰勒级数展开的初始值，方法是把已经得到的参数估计作为新的参数初始值，重新进行泰勒级数展开和参数估计。这种方法可以反复运用，直到得到比较理想的参数估计值。这种计算非线性回归参数估计的迭代算法称为“高斯 - 牛顿法”。

（2）牛顿—拉夫森法

牛顿—拉夫森（Newton - raphson）法的基本思想也是利用泰勒级数展开近似，通过迭代运算寻找最小二乘函数最优解的数值解法。

牛顿—拉夫森法的迭代运算，相当于在前一个参数估计向量的基础上，按单位移动幅度（通常称为“步长”）搜索更好的参数估计值，因此牛顿—拉夫森法也是一种搜索法。牛顿—拉夫森法的优点是搜索方向和步长的确定比较科学，因此找到满足精度要求最优水平的搜索次数一般要小一些。

牛顿—拉夫森方法的缺点是迭代运算中需要反复计算梯度向量，特别是海塞矩阵的逆矩阵，因此计算工作量很大。事实上，人们在实际应用中常常并不按照牛顿一拉夫森法进行搜索，而是根据一些简单法则确定搜索的步长，如“双向线性搜索法”就是其中常用的方法之一。

12.3.1 非线性模型的参数估计与迭代算法

R里面优化的基本函数是optimize()和optim()，可参考本书第3章。此外，R的stats包里提供专门做非线性最小二乘法的函数nls()。该函数的用法是nls(formula, data, start, control, algorithm, trace, subset, weights, na.action, model, lower, upper, ...)

nls()函数的默认迭代算法是高斯-牛顿算法。Start设定初始值，参数trace=T时，会把参数迭代过程返回来。

也可以使用极大似然法进行估计，maxLik包里提供了maxLik（ ）函数，但是需要先写出对数似然函数。其用法是

maxLik(logLik, grad = NULL, hess = NULL, start, method,constraints=NULL, ...)

12.3.2 初始值选取

在利用迭代算法进行非线性回归参数估计时，初始值的选择是一个值得重视的问题，如果我们想要得到较好的结果和提高工作效率，必须认真对待参数估计值的选择。但参数初始值的选择并没有一般法则。尽量接近参数真实值或最终估计值，最好是参数真实值的一致估计，是正确的初始值选择原则。但该原则的实用价值不大，因为参数真实值不可能知道，而一致估计量正是我们要求出的最小二乘估计量。在实践中，人们常常运用的是如下的经验方法：

1、利用参数的经济意义。

2、模型函数在特定点的性质。

3、降维法。

12.3.3 收敛性

理论上，非线性优化的迭代运算应该在梯度向量等于 0，也就是满足最优化的一阶条件处终止。但实际上这通常做不到，因为函数和导数的计算都有累积的舍入误差。因此，迭代算法一般是以某种收敛标准作为终止迭代的信号，而不是真正满足一阶条件。

判断收敛和终止迭代并没有一致接受的共同标准。常用的标准主要有：

（1） 目标函数(最小二乘函数)的改进已小于给定的小正数 ,即$|S(\boldsymbol{\beta}^{i+1})-S(\boldsymbol{\beta}^{i})|\le\varepsilon$ ，ε 即任意小正数；

（2） 第二，参数值的变化小于给定的小正数。当模型只有一个参数时即$\left\|\boldsymbol{\beta}^{i+1}-\boldsymbol{\beta}^{i}\right\|\le\varepsilon$ ；

（3） 第三，梯度向量的模小于给定的小正数，即$\left\|\mathbf{g}(\boldsymbol{\beta}^{i+1})\right\|\le\varepsilon$。

- 可决系数
- 参数显著性检验
- 似然比检验

非线性回归评价和假设检验

04

12.4.1 可决系数

由于反映线性回归模型的可决系数 $R^2 = 1 - \frac{\sum e_i^2}{\sum (Y_i - \bar{Y})^2}$ 和调整的可决系数 $\bar{R}^2 = 1 - \frac{n-1}{n-k}(1-R^2)$。不涉及参数估计量的分布性质，也不需要做以这些分布性质为基础的假设检验，因此非线性导致的问题并不影响该统计量在评价回归方程拟合度方面的作用，仍然是评价非线性模型合理程度的基本指标。它们在非线性回归分析中的使用方法仍然是与在线性回归分析中相同的。

12.4.2 参数显著性的F检验

除了对高斯 - 牛顿法非线性回归可以利用最后一次线性近似函数线性回归的 t 检验以外，检验非线性模型参数的显著性还有多种其他方法，下面这个渐近 F 分布的统计量就是其中的一种方法，即

$$F(g,n-k)=\frac{[S(\boldsymbol{\beta}_R)-S(\boldsymbol{\beta})]/g}{S(\boldsymbol{\beta})/(n-k)}$$

这个统计量分子、分母中的 $\boldsymbol{\beta}$ 是未对非线性模型参数施加约束时的参数估计，$\boldsymbol{\beta}_R$ 则是对模型的某些参数施加 0 假设约束后的参数估计，$S(\boldsymbol{\beta})$ 和 $S(\boldsymbol{\beta}_R)$ 分别是对应两种参数估计的残差平方和，g 是 0 约束参数的数量。

很显然，如果施加 0 约束的参数本身对模型的影响没有显著性，那么上述 F 统计量的数值会很小，如果这些施加 0 约束的参数对模型的影响是明显的，那么该统计量的数值会较大，就会有显著性。因此，我们可以通过检验该统计量的显著性来判断模型参数的显著性。

虽然上述 F 统计量与线性回归模型的 F 统计量形式是相似的，但因为模型是非线性的，因此 $S(\boldsymbol{\beta})$ 和 $S(\boldsymbol{\beta}_R)$ 并不服从 χ^2 分布，该统计量并不严格服从 F 分布，只是近似服从 F 分布。在样本容量较大时，该统计量的分布与 F 分布很接近。我们可以利用 F 分布检验该统计量的显著性，但检验结果论的准确程度会受到一定影响，运用时应该加以注意。

12.4.3 似然比检验

似然比检验与 F 检验在本质上一致的另一种非线性模型参数显著性检验。似然比检验的统计量为

$$\lambda = -2(\ln L(\boldsymbol{\beta}_R) - \ln L(\boldsymbol{\beta})) = -2\ln\frac{L(\boldsymbol{\beta}_R)}{L(\boldsymbol{\beta})}$$

式中 $\boldsymbol{\beta}$ 与 $\boldsymbol{\beta}_R$ 的含义与上述 F 检验统计量中同，$L(\boldsymbol{\beta})$ 与 $L(\boldsymbol{\beta}_R)$ 则分别是它们各自对应的非线性模型被自变量的似然函数值。似然函数即随机变量得到特定观测值序列的联立分布概率密度函数。

12.4.3 似然比检验

我们假设非线性模型的误差项服从均值为 0 的正态分布，那么上述统计量中的对数似然函数为

$$\ln L(\boldsymbol{\beta}) = -\frac{n}{2}\left[1+\ln(2\pi)+\ln\left(\frac{\mathbf{e}'\mathbf{e}}{n}\right)\right]$$

式中，$\mathbf{e}$ 是残差向量。如果参数估计采用的是最大似然估计，那么其中的 $\frac{\mathbf{e}'\mathbf{e}}{n}$ 实际上就是误差方差的估计。相应的有约束时，模型的对数似然函数为

$$\ln L(\boldsymbol{\beta}_R) = -\frac{n}{2}\left[1+\ln(2\pi)+\ln\left(\frac{\mathbf{e}_R'\mathbf{e}_R}{n}\right)\right]$$

因此，统计量 λ 为

$$\lambda = n\left(\ln\left(\frac{\mathbf{e}'\mathbf{e}}{n}\right)-\ln\left(\frac{\mathbf{e}_R'\mathbf{e}_R}{n}\right)\right)=n\ln\left(\frac{\mathbf{e}'\mathbf{e}}{\mathbf{e}_R'\mathbf{e}_R}\right)$$

12.4.3 似然比检验

对于大样本来说，该统计量渐近服从自由度为约束数量 g 的 χ^2 分布，因此可以根据 χ^2 分布检验 λ 的显著性。当该统计量比给定显著水平的 χ^2 分布临界值大时，拒绝 0 假设，认为所检验的参数是显著的，否则认为检验的参数是不显著的。

除了上述检验方法以外，非线性回归还有其他一些检验方法，如沃尔德检验和拉格朗日乘数检验。由于多数方法在本质上都是相似的，因此我们不再一一介绍。

Thank You