第1章

理论介绍

本书涉及的理论基础包括：区间规划模型、SLP 模型、模糊综合评价模型、分数阶理论以及数据挖掘等理论方法，将在本章进行详细的介绍及分析，为本书的理论研究奠定基础.

1.1　线性回归模型

回归分析是考察变量之间统计联系的一种重要方法，它在数学建模、自然科学和社会科学等许多领域中都有极其广泛的作用. 本节主要讨论一个随机变量或多个随机变量之间的关系. 回归(regression)一词是英国著名人类学家和气象学家 Francis Galton(1822—1911 年)于 1885 年引入的. 在“身高遗传中的平庸回归”的论文中，Galton 阐述了他的重大发现：虽然高个子的先代会有高个子的后代，但子代的身高并不像其父代，而是趋向于比他们的父代更加平均，就是说如果父亲身材高大，则子代的身材要比父代矮小一些；如果父亲身材矮小，则子代的身材要比父代高大一些. 因此，他用回归一词来描述子代身高与父代身高的这种关系. 而后，他的朋友、英国著名统计学家 K. Pearson 等人搜集了上千个家庭成员的身高数据，分析出儿子的身高 y 与父亲的身高 x 大致可归结为以下关系：

$$y = 0.516x + 33.73 \quad (\text{单位为英寸})$$

此式进一步证实了 Galton 的“回归定律”. 这就是回归一词最初在遗传学上的含义.

1.1.1　回归的概念

变量之间的关系大致可分为两类：一类是确定性的关系，如我们熟知的函

数关系；另一类是非确定性的关系. 对于某些非确定性的关系，如随机变量 Y 与变量 x（它可能是多维向量）之间的关系，当自变量 x 确定之后，因变量 Y 的值并不随着确定，而是按一定的统计规律（即随机变量 Y 的分布）取值. 这时我们将它们之间的关系表示为

$$Y = f(x) + \varepsilon$$

其中，$f(x)$是一个确定的函数，称为回归函数，ε 为随机项，且 $\varepsilon \sim N(0,\delta^2)$.

回归分析的任务之一是确定回归函数 $f(x)$. 当 $f(x)$是一元线性函数时，称为一元线性回归；当 $f(x)$是多元线性函数时，称为多元线性回归；当$f(x)$是非线性函数时，称为非线性回归.

1.1.2 一元线性回归

1. 一元线性回归的数学模型

假定只研究 x 与 y 的关系，可以有如下结构式：

$$y = \beta_0 + \beta_1 x + \varepsilon$$

式中，β_0 和 β_1 是未知常数，ε 表示其他随机因素对 y 获得率的影响，它服从 $N(0,\sigma^2)$分布.

取定一组不完全相同的值 $x_1, x_2, \cdots, x_n$，作独立试验得到 n 对观察结果 $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$，其中 y_i 是 $x = x_i$ 处对随机变量 y 观察的结果. 将数据点$(x_i, y_i)(i=1,2,\cdots,n)$代入，则有

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1,2,\cdots,n$$

并且假定 $\varepsilon_i \sim N(0,\sigma^2)$.

回归分析的首要任务是通过观察结果来确定回归系数 β_0, β_1 的估计 $\hat{\beta}_0, \hat{\beta}_1$，以下用最小二乘法确定回归直线方程：

$$y = \beta_0 + \beta_1 x$$

因为我们要确定一条直线，也就是要确定 β_0, β_1 的估计值，使回归直线与所有数据点都比较“接近”. 为了刻画这种“接近”程度，我们引进残差的概念，所谓残差是指观察值 y_i 与回归值$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ 的偏差 $y_i - \hat{y}_i$，很自然地可以用绝对残差和$\sum_{i=1}^{n} | y_i - \hat{y}_i |$来度量观察值与回归直线的接近程度. 绝对残差和越小，回归直线就与所有数据点越接近. 但为计算方便，一般用残差平方和：

$$Q(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \hat{y})^2 = \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_i)]^2$$

来描述所有观察值与回归直线的偏离程度.

所谓最小二乘估计,就是使残差平方和 Q 达到最小值的 β_0,β_1 作为回归系数的估计.

令 $\frac{\partial Q}{\partial \beta_0}=0$,$\frac{\partial Q}{\partial \beta_1}=0$,得

$$\hat{\beta}_1=\frac{S_{xy}}{S_{xx}},\quad \hat{\beta}_0=\bar{y}-\beta_1\bar{x}$$

其中,

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i,\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i,S_{xx}=\sum_{i=1}^{n}(x_i-\bar{x})^2,S_{xy}=\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}).$$

2. 一元线性回归的显著性检验

从前面求回归方程的过程来看,对任意样本观察值 (x_1,y_1),(x_2,y_2),…,(x_n,y_n) 做出的散点图,即使一看就知道这些点不可能近似在一条直线的附近,即 y 与 x 不存在线性关系,但是,若用最小二乘法仍然可求得 y 对 x 的线性回归方程与 $y=\beta_0+\beta_1x$,这样求得的方程是没有意义的.所以,在求 y 对 x 的线性回归方程之前,必须判断 y 与 x 的关系是否满足一元线性回归模型.

在处理具体问题时,判断 y 与 x 的关系是否满足一元线性回归模型,专业知识是重要的,在数学上,做出散点图是一种粗略的判断.下面介绍根据样本观察值 (x_1,y_1),(x_2,y_2),…,(x_n,y_n) 进行统计检验的一种做法.

由 $y=\beta_0+\beta_1x+\varepsilon$ 可知,当 $|\beta_1|$ 越大,y 随 x 的变化而变化的趋势就越明显;当 $|\beta_1|$ 越小,y 随 x 的变化而变化的趋势就不明显,特别当 $\beta_1=0$ 时,就认定 y 与 x 之间不存在线性相关关系.这样,判断 y 与 x 是否满足一元线性回归模型就转化为在显著性水平 α 下,检验假设:

$$H_0:\beta_1=0,\quad H_1:\beta_1\neq 0$$

记 $S_{yy}=\sum_{i=1}^{n}(y_i-\bar{y})^2$,$S_{xy}=\sum_{i,j=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$,$U=\sum_{i=1}^{n}(\beta_0+\beta_1x-\bar{y})^2=\frac{S_{xy}}{S_{xx}}$(称为回归平方和,$U+Q=S_{yy}$).

选取检验统计量:

$$F=\frac{U}{Q/(n-2)}$$

在 H_0 成立的条件下,$F\sim F(1,n-2)$,得到拒绝域:

$$W=\left\{F=\frac{U}{Q/(n-2)}\geqslant F_{\alpha}(1,n-2)\right\}$$

如果 $F>F_\alpha$，则否定 H_0，即变量 y 与 x 之间存在线性相关关系；否则，接受 H_0，即变量 y 与 x 之间不存在线性相关关系. 此时可能有以下几种情况：

（1）y 对 x 没有显著影响，此时应该丢掉自变量 x.

（2）y 对 x 有显著影响，但这种情况不能用线性关系来表示，应该做非线性回归.

（3）除 x 外还有其他不可忽略的变量对 y 也有显著影响，从而削弱了 x 对 y 的影响，此时对该问题的专业知识的了解往往起着重要作用.

3. 利用一元线性回归进行预测与控制

如何根据样本提供的信息来预测当变量 $x=x_0$ 时随机变量 Y_0 的值？一个自然的想法是：用预测量 $y_0=\beta_0+\beta_1 x_0$ 来代替，但是它与真值 Y_0 的差值是多少呢？预测量的优劣取决于 $|y_0-Y_0|$ 的大小，记为

$$d^2=1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}},\quad \hat{\delta}^2=\frac{Q}{n-2}$$

可以证明，当 Y_0 与 $Y_1,Y_2,\cdots,Y_n$ 相互独立时，$\dfrac{y_0-Y_0}{d\hat{\delta}}\sim t(n-2)$. 这样在显著性水平 α 下可得到 Y_0 的预测区间：

$$[y_0-t_\alpha(n-2)d\hat{\delta},y_0+t_\alpha(n-2)d\hat{\delta}]$$

当 n 较大时，预测区间的上下限近似取做 $y_0\pm 1.96\hat{\delta}$（可信度为 95%）或 $y_0\pm 2.58\hat{\delta}$（可信度为 99%）.

控制是预测的反问题，即若要随机变量 Y 落在指定的区间 (y_L,y_U) 内，变量 x 应控制在什么区间内？从方程

$$\begin{cases} y_L=\beta_0+\beta_1 x_L-1.96\hat{\delta}, \\ y_U=\beta_0+\beta_1 x_U-1.96\hat{\delta} \end{cases}$$

中解出 x_L 和 x_U，则当 $\beta_1>0$ 时，控制区间为 (x_L,x_U).

1.1.3 多元线性回归

在许多数学建模实际问题中，还会遇到一个随机变量与一组变量的相关关系问题，这要用到多元回归分析的方法来解决.

1. 多元线性回归的数学模型

设随机变量 $\boldsymbol{Y}$ 与 m 个变量 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_m$ 有关系

$$\boldsymbol{Y} = \beta_0 + \beta_1 \boldsymbol{x}_1 + \beta_2 \boldsymbol{x}_2 + \cdots + \beta_m \boldsymbol{x}_m + \boldsymbol{\varepsilon}$$

其中，$\boldsymbol{\varepsilon}$ 为随机项，且 $\varepsilon_i \sim N(0,\sigma^2)$. 记：

$$\boldsymbol{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1m} \\ 1 & x_{21} & \cdots & x_{2m} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{nm} \end{pmatrix}_{n\times(m+1)}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix}$$

则随机变量 $\boldsymbol{Y}$ 与变量 $\boldsymbol{X}$ 的关系可化为 $\boldsymbol{Y}=\boldsymbol{X\beta}+\boldsymbol{\varepsilon}$ 与一元线性回归情况类似，求回归系数 $\beta_0,\beta_1,\cdots,\beta_m$ 的最小二乘估计. 作残差平方和：

$$Q = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_{im} x_{im})^2$$

求 Q 分别关于 $\beta_0,\beta_1,\cdots,\beta_m$ 的一阶偏导数，并令它们等于零，得

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}$$

2. 多元线性回归的显著性检验

与一元回归情况类似，首先建立带检验假设：

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_m = 0$$

若能通过那个检验拒绝 H_0，则 $\boldsymbol{Y}$ 与 m 个变量 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_m$ 之间存在线性相关关系.

记 $\bar{y} = \dfrac{1}{n}\sum_{i=1}^{n} y_i$， $S_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2$， $U = S_{yy} - Q.$

选取检验统计量：

$$F = \frac{U/m}{Q/(n-m-1)}$$

在 H_0 成立的条件下，$F \sim F(m-1,n-m-1)$. 得到拒绝域：

$$W = \left\{ F = \frac{U/m}{Q/(n-m-1)} \geqslant F_\alpha(m-1,n-m-1) \right\}$$

如果 $F > F_\alpha$，则否定 H_0，即 $\boldsymbol{Y}$ 与 m 个变量 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_m$ 之间存在线性相关关系；否则，接受 H_0，即 $\boldsymbol{Y}$ 与 m 个变量 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_m$ 之间不存在线性相关关系.

在多元线性回归模型中，拒绝假设 H_0，即回归方程显著. 然而变量 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_m$ 对 $\boldsymbol{Y}$ 的影响并不都是十分重要的，人们还关心 $\boldsymbol{Y}$ 对 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_m$ 的回归中哪些因素更重要些，哪些因素不重要. 要剔除不重要的，需要采用偏 F 检验

法，即检验假设：

$$H_k: \beta_k = 0, \quad k = 1,2,\cdots,m$$

通常选取统计量：

$$F_\alpha = \frac{\beta_k^2 / a_{kk}}{Q/(n-m-1)}$$

其中，a_{kk} 是矩阵 $(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}$ 的主对角线上的第 $k+1$ 个元素.

在 H_k 成立的条件下，$F_k \propto F(1, n-m-1)$. 得到拒绝域：

$$W = \left\{ F_k = \frac{\beta_k^2 / a_{kk}}{Q/(n-m-1)} \geqslant F_\alpha(1, n-m-1) \right\}$$

如果 $F_k > F_\alpha$，拒绝 H_k，即 $\boldsymbol{x}_k$ 对 $\boldsymbol{Y}$ 的影响显著；否则，接受 H_k，即 $\boldsymbol{x}_k$ 对 $\boldsymbol{Y}$ 的影响不显著.

3. 预测问题

如何根据样本提供的信息来预测当变量 $(x_1, x_2, \cdots, x_m) = (x_{01}, x_{02}, \cdots, x_{0m})$ 时随机变量 $\boldsymbol{Y}_0$ 的值？一个自然的想法是用预测量：

$$\boldsymbol{y}_0 = \beta_0 + \beta_1 \boldsymbol{x}_{10} + \beta_2 \boldsymbol{x}_{20} + \cdots + \beta_m \boldsymbol{x}_{m0}$$

来替代. 预测量 $\boldsymbol{y}_0$ 的优劣取决于 $|\boldsymbol{y}_0 - \boldsymbol{Y}_0|$ 的大小. 记：

$$\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}, \quad l_{ij} = \sum_{k=1}^{n}(x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j), \quad i,j = 1,2,\cdots,m$$

$$\boldsymbol{L} = \begin{Bmatrix} l_{11} & \cdots & l_{1m} \\ \vdots & & \vdots \\ l_{m1} & \cdots & l_{mm} \end{Bmatrix}, \quad \boldsymbol{L}^{-1} = \begin{Bmatrix} l'_{11} & \cdots & l'_{1m} \\ \vdots & & \vdots \\ l'_{m1} & \cdots & l'_{mm} \end{Bmatrix}$$

$$d^2 = 1 + \frac{1}{n} + \sum_{i=1}^{m}\sum_{j=1}^{m} l'_{ij}(x_{0i} - \bar{x}_i)(x_{0j} - \bar{x}_j), \quad \hat{\delta}^2 = \frac{Q}{n-m-1}$$

可以证明当 $\boldsymbol{Y}_0$ 与 $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \cdots, \boldsymbol{Y}_n$ 相互独立时，$\dfrac{\boldsymbol{y}_0 - \boldsymbol{Y}_0}{d\hat{\delta}} \sim \boldsymbol{t}(n-m-1)$. 这样在显著性水平 α 下可得到 $\boldsymbol{Y}_0$ 的预测区间：

$$[\boldsymbol{y}_0 - \boldsymbol{t}_\alpha(n-m-1)d\hat{\delta}, \boldsymbol{y}_0 + \boldsymbol{t}_\alpha(n-m-1)d\hat{\delta}]$$

上面介绍了一元回归方程和多元回归方程的理论基础，下面引入与线性模型有关的 R 函数.

适应于多元线性模型的基本函数是 lm()，其形式为

```
lm(formula, data, subset, weights, na.action, method = 'qr', model = TRUE,
x = FALSE, y =  FALSE, qr = TRUE, singular.ok = TRUE, contrasts = NULL, offset, … )
```

其中 formula 为模型公式；data 为数据框；subset 为可选择向量，表示观察值的

子集；weights 为可选择向量，数据拟合的权重；返回值为线性模型结果的对象，存放在 fitted. model 中. 例如：

```
fm2<－lm(y～xl + x2, data = production)
```

适用于 y 关于 x_1，x_2 的多元回归模型(隐含着截距项).

lm()函数的返回值称为拟合结果的对象，本质上是一个具有类属性值 lm 的列表 model，coeffcients，residuals 等成员. lm()的结果非常简单，为了获得更多的信息，可以使用对 lm()类对象有特殊操作的通用函数，这些函数包括：

```
addl  coef  effects  kappa  predict  residuals
alias  deviance  family  labels  print  step
anova  dropl  formula  plot  proj  summary
```

下面简单地介绍函数的使用方法.

(1) anova()函数. 其使用格式为

```
anova (object, …)
```

其中 object 是由 lm 或 glm 得到的对象，其返回值是模型的方差分析表.

(2) coefficients()函数简写形式为 coef()，其使用格式为

```
coefficients(object, …)
coef(object, …)
```

其中 object 是由模型构成的对象，其返回值是模型的系数.

(3) deviance()函数，其使用格式为

```
deviance(object, …)
```

其中 object 是由模型构成的对象，其返回值是模型的残差平方和.

(4) formula()函数. 其使用格式为

```
formula(object, …)
```

其中 object 是由模型构成的对象，其返回值是模型公式.

(5) plot()函数，其使用格式为

```
plot (object, …)
```

其中 object 是由 lm 构成的对象，绘制模型诊断的几种图形，显示残差、拟合值和一些诊断情况.

(6) predict()函数，其使用格式为

```
predict(object, newdata = data. frame)
```

其中 object 是由 lm 构成的对象. newdata 是预测点的数据,它由数据框形式输入,其返回值是预测值和预测区间.

(7) print()函数,其使用格式为

```
print(object, …)
```

其中 object 是由模型构成的对象,其返回值是显示模型拟合的结果,一般不用 print()而直接用键入对象的名称来显示.

(8) residuals()函数,其使用格式为

```
residuals (object, type = c ( " working"," response"," deviance"," pearson",
"partial"),
```

其中 object 是由 lm 或 aov 构成的对象. type 是返回值的类型,其返回值是模型的残差,简单的命令形式为 resid (object).

(9) step()函数,其使用格式为

```
step(object, …)
```

其中 object 是由 lm 或 glm 构成的对象,其返回值是逐步回归,根据 AIC (Akaike's Information Criterion)的最小值选择模型.

(10) summary()函数,其使用格式为

```
summary(object, …)
```

其中 object 是由 lm 构成的对象,其返回值是显示较为详细的模型拟合结果.

例 1.1 根据经验,在人的身高相等的情况下,血压的收缩压 Y 与体重(千克)、年龄(岁数)有关,现收集了 13 个男子的数据,见表 1.1. 试建立 Y 关于 X_1, X_2 的线性回归方程.

表 1.1 数据表

序号	X_1	X_2	Y	序号	X_1	X_2	Y
1	76.0	50	120	8	79.0	50	125
2	91.5	20	141	9	85.0	40	132
3	85.5	20	124	10	76.5	40	123
4	82.5	30	126	11	82.0	40	132
5	79.0	30	117	12	95.0	40	155
6	80.5	50	125	13	92.5	20	147
7	74.5	60	123				

解 R 软件中的 lm()同样可以求出回归系数,并作相应的检验.

下面是 R 软件的计算过程:

```
> blood <- data.frame(
    Xl = c(76.0, 91.5, 85.5, 82.5, 79.0, 80.5, 74.5, 79.0, 85.0, 76.5, 82.0,
95.0, 92.5),
    X2 = c(50, 20, 20, 30, 30, 50, 60, 50, 40, 55, 40,40,20),
    Y =  c(120, 141, 124, 126, 117, 125, 123, 125, 132, 123, 132, 155, 147)
    )
> lm.sol <- lm(Y ~ X1 + X2, data = blood)
> summary(lm.sol)
Call:
lm(formula = Y ~ Xl + X2, data = blood)
Residuals :
   Min       1Q    Median    3Q      Max
-4.0404  -1.0183  0.4640  0.6908  4.3274
Coefficients :
            Estimate Std. Error t value Pr(>ltl)
(Intercept) -62.96336  16.99976  -3.704 0.004083 **
Xl            2.13656   0.17534  12.185 2.53e-07 ***
X2            0.40022   0.08321   4.810 0.000713 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05' .' 0.1 ' '1
Residual standard error: 2.854 On 10 degrees of freedom
Multiple R-Squared: 0.9461, Adjusted R-squared: 0.9354
F-statistic: 87.84 On 2 and 10 DF, p-value:4.531e-07
```

从计算结果可以得到,回归系数与回归方程的检验都是显著的,因此,回归方程为

$$\hat{Y} = -62.96 + 2.136X_1 + 0.4002X_2$$

```
>  source("beta.int.R")
>  beta.int(lm.sol)
                   Estimate          Left         Right
    (Intercept) -62.9633591  -100. 8411862  -25.0855320
    xl            2.1365581     1.7458709     2.5272454
    x2            0.4002162     0.2148077     0.5856246
```

其中 R 程序 beta.int.R 描述的是参数向量 β 的置信水平为 $1-\alpha$ 的区间估计程序:

```
beta. int <- function (fm, alpha = 0. 05){
```

```
A <- summary(fm) $ coefficients
        df <- fm $ df .residual
        left <- A[,i] - A[,2] * qt(1 - alpha/2, df)
        right <- A[,1] + A[,2]  * qt(1 - alpha/2, df)
        rowname <- dimnames (A)[ [1]]
        colname <- c("Estimate", "Left", "Right")
        matrix(c(A[,1], left, right), nc01 = 3, dimnames = list(rowname,
colname))
      }
```

其中 summary 是提取模型信息,返回值为一列表,其中 $ coefficients 是由回归系数、标准差、t 值和 P 值构成的矩阵,若 fm 是由 Im 计算得到回归模型,其中 $ df. residual 为模型的自由度.

1.1.4 曲线回归模型

在许多数学建模实际问题中,一个随机变量与另一个随机变量的关系不是线性关系,而是某种曲线关系.那么如何确定回归方程呢?常用的有 3 种方法,即:化为一元线性回归、多项式回归和分段回归.

1. 化为一元线性回归

在某些非线性回归方程中,为了确定其中的未知参数,一般将非线性回归方程转化为线性回归方程,通过线性回归方程的参数估计将非线性回归方程的参数估计出来.表 1.2 列出了常用的可线性化回归方程($a>0$).

表 1.2 常用的可线性化回归方程

曲线方程	变换公式	变换后的线性方程
$1/y=a+b/x$	$u=1/x, v=1/y$	$v=a+bu$
$y=ax^b$	$u=\ln x, v=\ln y$	$v=c+bu(c=\ln a)$
$y=a+b\ln x$	$u=\ln x, v=y$	$v=a+bu$
$y=a\mathrm{e}^{bx}$	$u=x, v=\ln y$	$v=c+bu(c=\ln a)$
$y=a\mathrm{e}^{b/x}$	$u=1/x, v=\ln y$	$v=c+bu(c=\ln a)$
$y=1/(a+b\mathrm{e}^{-x})$	$u=\mathrm{e}^{-x}, v=1/y$	$v=a+bu$

2. 多项式回归

在曲线回归中,比较困难的是选择合适的曲线类型.有的曲线也不一定经变