

三参数幂函数回归分析

北京航空航天大学 傅惠民**

【摘要】 本文给出了一种三参数幂函数回归分析方法。根据该方法,可以通过散点试验测定高置信度、高可靠度的三参数幂函数性能曲线。它不但适用于异方差情况,而且与传统的成组试验法相比,既可节省大量试件又具有更高的精度。

主题词: 回归分析 可靠性分析 置信限

1 三参数幂函数回归分析方法

三参数幂函数是工程中常见的一类性能曲线,例如 $P-S-N$ 曲线和 $P-\epsilon-N$ 曲线^[1]。由于还经常伴随着异方差情况,所以,过去主要采用成组试验法进行测定。但成组试验需要大量试件,而且测试精度又低,特别是在测定高置信度、高可靠度的性能曲线时,其费用和误差之大都使工程上难以接受。为此,下面将给出一种既省试件、精度又高的三参数幂函数回归分析方法。

设 $y \sim N(a + b \ln(t - \epsilon), \sigma^2(t))$, 即

$$y = a + b \ln(t - \epsilon) + e \quad (1)$$

或

$$y = a + bx + e \quad (2)$$

式中

$$x = \ln(t - \epsilon), \quad e \sim N(0, \sigma^2(t)), \quad \sigma(t) = \sigma_0 [1 + \theta(\ln t - x_0)] \quad (3)$$

其中 a, b, ϵ, θ 和 x_0 均为待定参数。如果 $(t_1, y_1), (t_2, y_2), \dots, (t_n, y_n)$ 是通过 n 次独立试验得到的一个样本,那么 a, b, ϵ 和 θ 的估计量将由下列诸式给出

$$\hat{a} = \bar{y} - b\bar{x} \quad (4)$$

$$\hat{b} = l_{xy}/l_{xx} \quad (5)$$

$$E_1(\epsilon, \theta) = \frac{l_{x\epsilon}}{l_{xx}} - \frac{l_{y\epsilon}}{l_{xy}} = 0 \quad (6)$$

$$E_2(\epsilon, \theta) = l_{yy\theta} + \frac{l_{xy}^2}{l_{xx}^2} l_{xx\theta} - 2 \frac{l_{xy}}{l_{xx}} l_{xy\theta} = 0 \quad (7)$$

式中

$$x_0 = \frac{1}{n} \sum_{i=1}^n \ln t_i$$

$$I(t, \theta) = 1 + \theta(\ln t - x_0)$$

$$x_{\epsilon} = \ln(t_i - \epsilon)$$

$$x_{\epsilon} = \frac{1}{t_i - \epsilon}$$

$$\bar{x} = \sum_{i=1}^n \frac{x_i}{I^2(t_i, \theta)} / \sum_{i=1}^n \frac{1}{I^2(t_i, \theta)} \quad \bar{y} = \sum_{i=1}^n \frac{y_i}{I^2(t_i, \theta)} / \sum_{i=1}^n \frac{1}{I^2(t_i, \theta)}$$

$$\bar{x\epsilon} = \sum_{i=1}^n \frac{x_i \epsilon_i}{I^2(t_i, \theta)} / \sum_{i=1}^n \frac{1}{I^2(t_i, \theta)} \quad l_{xx} = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{I^2(t_i, \theta)}$$

$$l_{yy} = \sum_{i=1}^n \frac{(y_i - \bar{y})^2}{I^2(t_i, \theta)} \quad l_{xy} = \sum_{i=1}^n \frac{(x_i - \bar{x})(y_i - \bar{y})}{I^2(t_i, \theta)}$$

$$l_{x\epsilon} = \sum_{i=1}^n \frac{(x_i - \bar{x})(x_i \epsilon_i - \bar{x\epsilon})}{I^2(t_i, \theta)} \quad l_{y\epsilon} = \sum_{i=1}^n \frac{(y_i - \bar{y})(x_i \epsilon_i - \bar{x\epsilon})}{I^2(t_i, \theta)}$$

$$l_{xx\theta} = \sum_{i=1}^n \frac{(x_i - \bar{x})^2 (\ln t_i - x_0)}{I^3(t_i, \theta)} \quad (8)$$

$$l_{yy\theta} = \sum_{i=1}^n \frac{(y_i - \bar{y})^2 (\ln t_i - x_0)}{I^3(t_i, \theta)} \quad (9)$$

$$l_{xy\theta} = \sum_{i=1}^n \frac{(x_i - \bar{x})(y_i - \bar{y})(\ln t_i - x_0)}{I^3(t_i, \theta)} \quad (10)$$

首先可由式(6)和(7)联立求出 ϵ 和 θ , 然后再由式(4)和(5)求得 $\hat{a}$, $\hat{b}$. 式(6)和(7)可采用下述双二分法求解: 对于给定的 ϵ , 设 θ 满足 $E_2(\epsilon, \theta) = 0$, 则当 $\theta' < \theta$ 时, 有 $E_2(\epsilon, \theta') > 0$; 当 $\theta' > \theta$ 时, 有 $E_2(\epsilon, \theta') < 0$, 所以可很容易通过二分法求得 θ . 然后再计算 $E_1(\epsilon, \theta)$, 如果 $E_1(\epsilon, \theta) > 0$, 则 $\epsilon < \hat{\epsilon}$; 如果 $E_1(\epsilon, \theta) < 0$, 则 $\epsilon > \hat{\epsilon}$. 故又可用二分法并重复前一二分法求得使 $E_1(\epsilon, \theta) = 0$ 的 ϵ , 此时 $\hat{\theta}$ 也即为所要求的 θ . 如果 $E_1(\epsilon, \theta)$ 和 $E_2(\epsilon, \theta)$ 的零点不在 ϵ 或 θ 所允许的区间内, 则可取与该零点最邻近的边界点作为 ϵ 或 θ 即可. ϵ 所在的区间为 $\epsilon \in [t_{\min}, \dots]$, 其中 $t_{\min}$ 为诸 $t_i (i=1, 2, \dots, n)$ 中的最小者, 有时还要求 $\epsilon \geq 0$. θ 所在区间有三种情况: 当标准差随着 t 的增加而单调减小时, 有

$$-1/(x_{\max} - x_0) < \theta \leq 0 \quad (11)$$

当标准差随着 t 的增加而单调增大时, 有

$$0 \leq \theta < 1/(x_0 - x_{\min}) \quad (12)$$

当标准差既可以单调增也可以单调减时, 则有

$$-1/(x_{\max} - x_0) < \theta < 1/(x_0 - x_{\min}) \quad (13)$$

式中 $x_{\min}$ 和 $x_{\max}$ 分别为诸 $\ln t_i (i=1, 2, \dots, n)$ 中的最小者和最大者.

其回归方程由下式给出

$$\tilde{y} = \hat{a} + \hat{b} \ln(t - \epsilon) \quad (14)$$

$$\tilde{y} = \bar{y} + \hat{b}(x - \bar{x}) \quad (15)$$

即

$$\tilde{y} = \bar{y} + \hat{b}(x - \bar{x}) \sim N \left[a + bx, \left(1 / \sum_{i=1}^n \frac{1}{I^2(t_i, \theta)} + \frac{(x - \bar{x})^2}{l_{xx}} \right) \sigma_0^2 \right] \quad (16)$$

σ_0 的估计量 $\hat{\sigma}_0$ 为

$$\hat{\sigma}_0 = \sqrt{\frac{1}{v} \sum_{i=1}^n \frac{(y_i - \hat{a} - \hat{b}x_i)^2}{I^2(t_i, \theta)}} \quad (17)$$

标准差 $\sigma(t)$ 的估计量 $\hat{\sigma}(t) = \hat{\sigma}_0 I(t, \theta)$, 并且

$$v \hat{\sigma}^2(t) / \sigma^2(t) \sim \chi^2(v) \quad (18)$$

式中 v 为方差 $\hat{\sigma}^2(t)$ 的自由度, 当 ϵ 和 θ 均为零时, $v = n - 2$; 当 ϵ 和 θ 中只有一个为零时, 则 $v =$

$n-3$; 当 ϵ 和 θ 均不为零时, 有 $v=n-4$ 。线性相关系数为

$$r = l_{xy} / \sqrt{l_{xx} l_{yy}} \quad (19)$$

对于一般的异方差情况, 可设

$$\sigma(t) = \sigma_0 \{1 + \theta [\varrho(t) - x_0]\}^\alpha \quad (20)$$

式中 $\varrho(t)$ 为已知函数, α 为预先给定的正实数。那么只要令

$$x_0 = \frac{1}{n} \sum_{i=1}^n \varrho(t_i), \quad I(t, \theta) = \{1 + \theta [\varrho(t) - x_0]\}^\alpha$$

并将式(8)、(9)和(10)中的 $\ln(t - x_0)$ 换为 $\partial(t, \theta)/\partial$, 本文所有公式和结论都将仍然成立。其中 θ 所在的区间根据标准差 $\sigma(t)$ 随 $\varrho(t)$ 单调减、单调增或两者均可, 分别由式(11)、(12)和(13)给出, 式中 $x_{\min}$ 和 $x_{\max}$ 是诸 $\varrho(t_i)$ ($i=1, 2, \dots, n$) 中的最小者和最大者。例如当取 $I(t, \theta) = \sqrt{1 + \theta(\ln t - x_0)}$ 时, 即为方差成线性变化的情况。

更进一步, 如果式(20)中的 $\varrho(t)$ 还与 ϵ 有关, 即 $\varrho(t) = \varrho(t, \epsilon)$, 那么只要将上述中的 $\varrho(t)$ 和 $\varrho(t_i)$ 均换成 $\varrho(t, \epsilon)$ 和 $\varrho(t_i, \epsilon)$, 并将式(6)用

$$E_1(\epsilon, \theta) = l_{yy\epsilon} + (l_{xy}^2/l_{xx}^2)(l_{x\epsilon} + l_{xx}\epsilon) - (l_{xy}/l_{xx})(l_{y\epsilon} + 2l_{xy}\epsilon) = 0 \quad (21)$$

代替即可。式中

$$\begin{aligned} l_{xx\epsilon} &= \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{I^3(t_i, \theta)} \frac{\partial(t_i, \theta)}{\partial \epsilon} \\ l_{yy\epsilon} &= \sum_{i=1}^n \frac{(y_i - \bar{y})^2}{I^3(t_i, \theta)} \frac{\partial(t_i, \theta)}{\partial \epsilon} \\ l_{xy\epsilon} &= \sum_{i=1}^n \frac{(x_i - \bar{x})(y_i - \bar{y})}{I^3(t_i, \theta)} \frac{\partial(t_i, \theta)}{\partial \epsilon} \end{aligned}$$

例如, 当取 $I(t, \theta) = 1 + \theta(x - x_0)$ 时, 即为标准差与 $x = \ln(t - \epsilon)$ 成线性变化的情况。

2 百分位值的置信限曲线

2.1 母体下百分位值的单侧置信下限曲线

工程中有时只关心随机变量 y 的下百分位值的单侧置信下限, 如安全寿命、安全强度等。根据文献[2~4]和式(16)、(18)可知, 百分率为 P ($P \leq 50\%$) 的母体下百分位值 $y_P(t)$ 的置信度为 γ 的单侧置信下限 $y_{PL}(t)$ 由下式给出

$$y_{PL}(t) = a + b \ln(t - \epsilon) - k_P(t) \hat{\sigma}(t) \quad (22)$$

即

$$P\{y_{PL}(t) \leq y_P(t)\} = \gamma \quad (23)$$

式中 $k_P(t)$ 是置信度为 γ 概率为 P 的二维单侧容限系数, 当 $v \geq 2$ 时, 可由下式计算

$$k_P(t) = \frac{-u_P + u_Y \sqrt{(1 - u_Y^2/w)/n_e(t) + u_P^2/w}}{1 - u_Y^2/w} \sqrt{\frac{2v-1}{2v-2}} \quad (24)$$

$$w = 2 \left[v + u_Y - c - \frac{1}{\sqrt{v + u_Y - c}} \right] \quad (25)$$

$$n_e(t) = \frac{I^2(t, \theta)}{1/\sum_{i=1}^n \frac{1}{I^2(t_i, \theta)} + \frac{(x - \bar{x})^2}{l_{xx}}} \quad (26)$$

其中 $n_e(t)$ 是回归函数 $\hat{y}$ 的自由度, ν 为 $\hat{\sigma}^2(t)$ 的自由度, 这里 $\hat{y}$ 和 $\hat{\sigma}^2(t)$ 可以来自不同的样本。 u_P 和 u_Y 为标准正态偏量, 即 $P = \Phi(u_P)$ 和 $Y = \Phi(u_Y)$, $\Phi(\cdot)$ 为标准正态变量的分布函数。 c 值可由文献[2]查得。 当 $\nu \geq 5$ 时, 对于 $P \leq 50\%$, 可取 $c = 0.64$; 对于 $P > 50\%$, 可取 $c = 1.24u_Y$ 。 将式(24)代入式(22)即得

$$y_{PL}(t) = a + b \ln(t - \epsilon) + u_P c_0 \hat{\sigma}_0 I(t, \theta) - u_Y c_0 \hat{\sigma}_0 \sqrt{\left[1 / \sum_{i=1}^n \frac{1}{I^2(t_i, \theta)} + \frac{(x - \bar{x})^2}{l_{xx}} \right] \left(1 - \frac{u_Y^2}{w} \right) + \frac{u_P^2}{w} I^2(t, \theta)} \quad (27)$$

式中 $c_0 = \sqrt{\hat{\sigma}_0^2 (2\nu - 1) / (2\nu - 2) / (1 - u_Y^2/w)}$ 。 当 $I(t, \theta) = 1 + \theta[\mathcal{Q}(t) - x_0]$ 时, 则有

$$y_{PL}(t) = a_0 + b \ln(t - \epsilon) + u_P c_0 \hat{\sigma}_0 \theta \mathcal{Q}(t) - u_Y c_0 \hat{\sigma}_0 \sqrt{\left(1 - \frac{u_Y^2}{w} \right) \frac{x^2 - 2x\bar{x}}{l_{xx}} + \frac{u_P^2}{w} \theta^2 \mathcal{Q}^2(t) + 2\theta(1 - \theta x_0) \mathcal{Q}(t)} + c_5 \quad (28)$$

特别当 $\mathcal{Q}(t) = \ln t$, 即 $I(t, \theta) = 1 + \theta(\ln t - x_0)$ 时, 有

$$y_{PL}(t) = a_0 + b \ln(t - \epsilon) + u_P c_0 \hat{\sigma}_0 \theta \ln t - u_Y c_0 \hat{\sigma}_0 \sqrt{c_1 \ln^2(t - \epsilon) + c_2 \ln(t - \epsilon) + c_3 \ln^2 t + c_4 \ln t + c_5} \quad (29)$$

式中

$$\begin{aligned} a_0 &= \hat{a} + c_0 \hat{\sigma}_0 (1 - \theta x_0) u_P, & c_1 &= (1 - u_Y^2/w) / l_{xx} \\ c_2 &= -2\bar{x}(1 - u_Y^2/w) / l_{xx}, & c_3 &= \theta^2 u_P^2 / w, & c_4 &= 2\theta(1 - \theta x_0) u_P^2 / w \\ c_5 &= \frac{(1 - \theta x_0)^2 u_P^2}{w} + \left[1 / \sum_{i=1}^n \frac{1}{I^2(t_i, \theta)} + \frac{\bar{x}^2}{l_{xx}} \right] \left(1 - \frac{u_Y^2}{w} \right) \end{aligned}$$

2.2 母体上百分位值的单侧置信上限曲线

工程中也常常只关心随机变量 y 的上百分位值的单侧置信上限, 如安全载荷、有害物质含量等。 同样, 根据文献[2~4]和式(16)、(18)可知, 百分率为 P ($P \geq 50\%$) 的母体上百分位值 $y_P(t)$ 的置信度为 Y 的单侧置信上限 $y_{PU}(t)$ 由下式给出

$$y_{PU}(t) = a + b \ln(t - \epsilon) + k_{1-P}(t) \hat{\sigma}(t) \quad (30)$$

即

$$P\{y_{PU}(t) \geq y_P(t)\} = Y \quad (31)$$

式中 $k_{1-P}(t)$ 是概率为 $1 - P$ 、置信度为 Y 的二维单侧容限系数, 由式(24)计算。 将其代入式(30), 即得

$$y_{PU}(t) = \hat{a} + b \ln(t - \epsilon) + u_P c_0 \hat{\sigma}_0 I(t, \theta) + u_Y c_0 \hat{\sigma}_0 \sqrt{\left[1 / \sum_{i=1}^n \frac{1}{I^2(t_i, \theta)} + \frac{(x - \bar{x})^2}{l_{xx}} \right] \left(1 - \frac{u_Y^2}{w} \right) + \frac{u_P^2}{w} I^2(t, \theta)} \quad (32)$$

当 $I(t, \theta) = 1 + \theta[\mathcal{Q}(t) - x_0]$ 时, 则有

$$y_{PU}(t) = a_0 + b \ln(t - \epsilon) + u_P c_0 \hat{\sigma}_0 \theta \mathcal{Q}(t) + u_Y c_0 \hat{\sigma}_0 \sqrt{\left(1 - \frac{u_Y^2}{w} \right) \frac{x^2 - 2x\bar{x}}{l_{xx}} + \frac{u_P^2}{w} \theta^2 \mathcal{Q}^2(t) + 2\theta(1 - \theta x_0) \mathcal{Q}(t)} + c_5 \quad (33)$$

尤其当 $\mathcal{Q}(t) = \ln t$, 即 $I(t, \theta) = 1 + \theta(\ln t - x_0)$ 时, 有

$$y_{PU}(t) = a_0 + b \ln(t - \epsilon) + u_P c_0 \hat{\sigma}_0 \theta \ln t + u_Y c_0 \hat{\sigma}_0 \sqrt{c_1 \ln^2(t - \epsilon) + c_2 \ln(t - \epsilon) + c_3 \ln^2 t + c_4 \ln t + c_5} \quad (34)$$

根据式(22)和式(30)还可以对两个母体的百分位值曲线进行对比检验^[4]。

3 三参数幂函数置信限曲线

由于受 ϵ 的影响, 式(22)和式(30)给出的百分位值的单侧置信下限和置信上限虽然在试验段内有较高的精度, 但不能用于外推。如果要用于外推, 则必须将式(22)转换成下面的三参数幂函数置信下限曲线

$$y_{P\epsilon}(t) = a_P + b_P \ln(t - \epsilon_P) \quad (35)$$

式中 a_P 、 b_P 和 ϵ_P 均为待定参数, 可由下面诸式确定

$$a_P = \bar{y}_P - b_P \bar{x}_P \quad (36)$$

$$b_P = l_{Pxy} / l_{Pxx} \quad (37)$$

$$E(\epsilon_P) = (l_{Px\epsilon} / l_{Pxx}) - (l_{Py\epsilon} / l_{Pxy}) = 0 \quad (38)$$

其中

$$\bar{x}_P = \frac{1}{n} \sum_{i=1}^n \ln(t_i - \epsilon_P), \quad \bar{y}_P = \frac{1}{n} \sum_{i=1}^n y_{PL}(t_i), \quad x_{ci} = \frac{1}{t_i - \epsilon_P}$$

$$\bar{x}_c = \frac{1}{n} \sum_{i=1}^n x_{ci}, \quad l_{Pxx} = \sum_{i=1}^n [\ln(t_i - \epsilon_P) - \bar{x}_P]^2$$

$$l_{Pyy} = \sum_{i=1}^n [y_{PL}(t_i) - \bar{y}_P]^2, \quad l_{Pxy} = \sum_{i=1}^n [\ln(t_i - \epsilon_P) - \bar{x}_P][y_{PL}(t_i) - \bar{y}_P]$$

$$l_{Px\epsilon} = \sum_{i=1}^n [\ln(t_i - \epsilon_P) - \bar{x}_P](x_{ci} - \bar{x}_c), \quad l_{Py\epsilon} = \sum_{i=1}^n [y_{PL}(t_i) - \bar{y}_P](x_{ci} - \bar{x}_c)$$

式中 $y_{PL}(t_i)$ 由式(22)计算得到。如果将 $y_{PL}(t_i)$ 换成由式(30)计算得到的 $y_{PU}(t_i)$, 则式(35)将给出百分率为 P 的母体上百分位值 $y_P(t)$ 的置信度为 γ 的三参数幂函数置信上限曲线。

首先可由式(38)求得 ϵ_P , 再由式(36)和(37)得到 a_P 和 b_P 。式(38)需迭代求解。由于 $E(\epsilon_P)$ 具有如下性质: 设 $\hat{\epsilon}_P$ 为 ϵ_P 的预估值, 则当 $\hat{\epsilon}_P < \epsilon_P$ 时, $E(\hat{\epsilon}_P) > 0$; 当 $\hat{\epsilon}_P > \epsilon_P$ 时, $E(\hat{\epsilon}_P) < 0$, 所以可很容易地用二分法求得 ϵ_P 。

结束语 根据文献[3]和本文方法, 还可以求得三参数幂函数情况下的百分位值和百分率的各种预测区间、容忍区间和控制域; 可以对百分位值和百分率曲线进行检验和对比试验; 可以综合利用当前数据和经验数据, 提高预测和控制的精度或减少试件个数。

参 考 文 献

- 1 傅惠民 cN 曲线三参数幂函数公式 航空学报, 1993, 14(3): A 173~176
- 2 傅惠民 正态分布百分位值和百分率的置信限和容忍限公式 航空学报, 1994, 15(1): 94~101
- 3 傅惠民 百分回归分析 航空学报, 1994, 15(2): 142~149
- 4 傅惠民 线性异方差回归分析 航空学报, 1994, 15(3): 295~302

(责任编辑 魏星禄)