

TIÊU CHUẨN QUỐC GIA

TCVN 4554 : 2009

THỐNG KÊ ỨNG DỤNG – ƯỚC LƯỢNG VÀ KHOẢNG TIN CẬY ĐỐI VỚI CÁC THAM SỐ CỦA PHÂN BỐ WEIBULL

Applied statistics – Estimation and confidence intervals for parameters of Weibull distribution

Lời nói đầu

TCVN 4554 : 2009 thay thế cho TCVN 4554-1988;

TCVN 4554 : 2009 do Ban kỹ thuật tiêu chuẩn quốc gia TCVN/TC 69 *Ứng dụng các phương pháp thống kê* biên soạn, Tổng cục Tiêu chuẩn đo lường Chất lượng đề nghị, Bộ Khoa học và Công nghệ công bố.

THỐNG KÊ ỨNG DỤNG – ƯỚC LƯỢNG VÀ KHOẢNG TIN CẬY ĐỐI VỚI CÁC THAM SỐ CỦA PHÂN BỐ WEIBULL

Applied statistics – Estimation and confidence intervals for parameters of Weibull distribution

1. Phạm vi áp dụng

Tiêu chuẩn này quy định các quy tắc xác định các ước lượng và khoảng tin cậy cho các tham số của phân bố Weibull dựa vào mẫu ngẫu nhiên đơn giản rút ra từ tổng thể có phân bố Weibull.

2. Thuật ngữ và định nghĩa

Tiêu chuẩn này sử dụng các thuật ngữ và định nghĩa trong TCVN 8244-2 (ISO 3534-2), *Thống kê học – Từ vựng và ký hiệu – Thống kê ứng dụng*.

3. Các đặc trưng cơ bản

3.1. Đại lượng ngẫu nhiên X được gọi là có phân bố Weibull (ba tham số) nếu hàm phân bố của nó có dạng:

$$F(x; a, b, c) = \begin{cases} 1 - \exp\left[-\left(\frac{x-c}{a}\right)^b\right], & x > c \\ 0, & x \leq c \end{cases} \quad (1)$$

trong đó: $a > 0$ là tham số tỉ lệ;

$b > 0$ là tham số dạng;

c là tham số định vị.

Nếu $c = 0$ thì phân bố Weibull được gọi là phân bố Weibull hai tham số.

Khi $b = 1$, phân bố Weibull trở thành phân bố mũ.

Khi $b = 2$, phân bố Weibull trở thành phân bố Rayleigh.

3.2. Hàm mật độ của đại lượng ngẫu nhiên X có phân bố Weibull có dạng:

$$f(x; a, b, c) = \begin{cases} \frac{b}{a} \left(\frac{x-c}{a}\right)^{b-1} \exp\left[-\left(\frac{x-c}{a}\right)^b\right], & x > c \\ 0, & x \leq c \end{cases} \quad (2)$$

3.3. Kỳ vọng, phương sai, hệ số biến động, độ bất đối xứng của đại lượng ngẫu nhiên X có phân bố Weibull:

Kỳ vọng:

$$E(X) = aK_b + c \quad (3)$$

Phương sai:

$$D(X) = a^2 + g_b^2 \quad (4)$$

Hệ số biến động:

$$v_b = \frac{ag_b}{aK_b + c} \quad (5)$$

Độ bất đối xứng

$$\rho_b = \frac{r\left(1 + \frac{3}{b}\right) - 3K_b r\left(1 + \frac{2}{b}\right) + 2K_b^3}{g_b^3} \quad (6)$$

trong đó:

$$K_b = r - 1 - \frac{1}{b} \quad (7)$$

$$g_b = \sqrt{r - 1 - \frac{2}{b} - K_b^2} \quad (8)$$

và ký hiệu r chỉ hàm số sau:

$$r(y) = \int_0^y x^{y-1} e^{-x} dx \quad (9)$$

Các giá trị K_b , g_b , v_b cho trong Bảng 2.

4. Các vấn đề chung

Có mẫu ngẫu nhiên đơn giản $x_1, \dots, x_n$ và x có phân bố Weibull với ba tham số a, b, c .

4.1. Các ước lượng điểm

- Ước lượng a khi b và c đã biết;
- Ước lượng a và b khi c đã biết;
- Ước lượng a và c khi b đã biết;
- Ước lượng cả ba tham số;
- Ước lượng kỳ vọng, phương sai, hệ số biến phân.

4.2. Khoảng tin cậy

- Khoảng tin cậy của tham số a khi b và c đã biết;
- Khoảng tin cậy của tham số a khi b chưa biết và c đã biết;
- Khoảng tin cậy của tham số b khi a chưa biết và c đã biết;

4.3. Tiêu chuẩn sử dụng hai phương pháp ước lượng: phương pháp mô men và phương pháp hợp lý cực đại. Phương pháp mô men đơn giản hơn, phương pháp hợp lý cực đại chính xác hơn nhưng tính toán phức tạp hơn, nó được dùng tốt nhất trong trường hợp sử dụng máy tính.

5. Ước lượng của tham số a khi b và c đã biết

Nếu b và c đã biết, ước lượng $\bar{a}$ của a có dạng:

$$\bar{a} = \left(\frac{\sum_{i=1}^n (x_i - c)^b}{n} \right)^{\frac{1}{b}} \quad (10)$$

6. Ước lượng của các tham số a, b khi c đã biết

6.1. Ước lượng đơn giản có thể nhận được bằng cách sau:

1) Tính

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n \ln(x_i - c) \quad (11)$$

$$S_y = \sqrt{\frac{\sum_{i=1}^n [\ln(x_i - c) - \bar{y}]^2}{n-1}} \quad (12)$$

2) Tính ước lượng $\bar{b}$ của tham số b dưới dạng

$$\bar{b} = \frac{\sqrt{6}}{S_y} \quad (13)$$

3) Tính ước lượng $\bar{a}$ của tham số a dưới dạng

$$\bar{a} = \exp \left(\bar{y} - \frac{1}{\bar{b}} \right) \quad (14)$$

trong đó hằng số O'le = 0,577 226 và

$$\sqrt{6} = 1,282 55$$

Khi n bé, ước lượng $\bar{b}$ có độ chệch đáng kể. Để có ước lượng không chệch $\hat{b}$ của tham số b cần hiệu chỉnh như sau:

$$\hat{b} = M(n) \times \bar{b} \quad (15)$$

$M(n)$ được gọi là hệ số không chệch và cho trong Bảng 1.

6.2. Các ước lượng $\bar{a}, \bar{b}$ của các tham số a, b (khi c đã biết) theo phương pháp hợp lý cực đại nhận được bằng cách giải hệ phương trình (các ẩn số là a và b).

$$\frac{n}{b} - \frac{n \sum_{i=1}^n (x_i - c)^b \ln(x_i - c)}{\sum_{i=1}^n (x_i - c)^b} + \sum_{i=1}^n \ln(x_i - c) = 0 \quad (16)$$

$$a = \left(\frac{\sum_{i=1}^n \ln(x_i - c)^b}{n} \right)^{\frac{1}{b}} \quad (17)$$

Phương trình (16) có thể giải bằng phương pháp xấp xỉ liên tiếp Newton Raphson theo công thức:

$$\tilde{b}_{k+1} = \tilde{b}_k + \frac{\frac{1}{\tilde{b}_k} + \frac{s_1}{n} - \frac{s_3^{(k)}}{s_2^{(k)}}}{\frac{1}{\tilde{b}_k^2} + \frac{s_2^{(k)} s_4^{(k)} - (s_3^{(k)})^2}{(s_2^{(k)})^2}} \quad (18)$$

trong đó:

$$s_1 = \sum_{i=1}^n \ln(x_i - c) \quad (19)$$

$$s_2^{(k)} = \sum_{i=1}^n (x_i - c)^{\tilde{b}_k} \quad (20)$$

$$s_3^{(k)} = \sum_{i=1}^n (x_i - c)^{\tilde{b}_k} \ln(x_i - c) \quad (21)$$

$$s_4^{(k)} = \sum_{i=1}^n (x_i - c)^{\tilde{b}_k} \ln^2(x_i - c) \quad (22)$$

Và $\tilde{b}_k$ là xấp xỉ thứ k của nghiệm $\tilde{b}$. Có thể dùng ước lượng $\tilde{b}$ (theo 4.1) làm xấp xỉ thứ nhất. Quá trình xấp xỉ liên tiếp sẽ kết thúc khi điều kiện sau được thỏa mãn:

$$G_b^{(k)} = \left| \frac{n}{\tilde{b}_k} - \frac{n - s_3^{(k)}}{s_2^{(k)}} + s_1 \right| < \varepsilon \quad (23)$$

Ở đây ε là số dương nhỏ cho trước, chẳng hạn $\varepsilon = 10^{-5}$.

Ước lượng $\tilde{a}$ của a nhận được bằng cách thay b bằng $\tilde{b}$ vào phương trình (17).

Với $n < 120$ ước lượng không chệch của tham số b có thể nhận được bởi

$$\hat{\tilde{b}} = B(n) \tilde{b} \quad (24)$$

Hệ số hiệu chỉnh $B(n)$ cho trong Bảng 1.

7. Ước lượng của các tham số a và c khi b đã biết

7.1. Nếu b đã biết thì việc xác định các ước lượng $\bar{a}$ và $\bar{c}$ của tham số a và c theo phương pháp mô men được tiến hành theo thứ tự sau (tuy nhiên khi $b \leq 1$ nên ứng dụng trực tiếp phương pháp ở 7.3).

1) Tính các đặc trưng mẫu

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (25)$$

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (26)$$

2) Ứng với giá trị b cho trước, theo Bảng 2 tìm các giá trị g_b và K_b .

3) Ước lượng $\bar{a}$ của a được tính theo công thức

$$\bar{a} = \frac{s}{g_b} \quad (27)$$

4) Từ các giá trị mẫu $x_1, x_2, \dots, x_n$ tìm $x_{\min}$ (giá trị nhỏ nhất) và tính

$$c = \bar{x} - \bar{a}K_b \quad (28)$$

Ước lượng $\bar{c}$ của c là nhỏ nhất trong hai số c^* và $x_{\min}$.

7.2. Nếu $b > 1$, ước lượng hợp lý cực đại $\tilde{a}$ và $\tilde{c}$ của các tham số a và c được xác định bằng cách giải hệ phương trình (các ẩn số là a và c)

$$nb \frac{\sum_{i=1}^n (x_i - c)^{b-1}}{\sum_{i=1}^n (x_i - c)^b} - (b-1) \sum_{i=1}^n \frac{1}{(x_i - c)} = 0 \quad (29)$$

$$a = \left(\frac{1}{n} \sum_{i=1}^n (x_i - c)^b \right)^{\frac{1}{b}} \quad (30)$$

Phương trình (29) có thể giải bằng phương pháp xấp xỉ liên tiếp theo Newton Raphson theo công thức

$$\tilde{c}_{k+1} = \tilde{c}_k + \frac{\frac{s_4^{(k)}}{s_5^{(k)}} - \frac{b-1}{nb} s_1^{(k)}}{\frac{(b-1)s_3^{(k)}}{s_5^{(k)}} - \frac{b(s_4^{(k)})^2}{(s_5^{(k)})^2} + \frac{b-1}{nb} s_2^{(k)}} \quad (31)$$

□ trong đó

$$s_1^{(k)} = \sum_{i=1}^n \frac{1}{\tilde{c}_k - x_i} \quad (32)$$

$$s_2^{(k)} = \frac{1}{n} \sum_{i=1}^n \frac{1}{(x_i - \tilde{c}_k)^2} \quad (33)$$

$$s_3^{(k)} = \frac{1}{n} \sum_{i=1}^n (x_i - \tilde{c}_k)^{b-2} \quad (34)$$

$$s_4^{(k)} = \frac{1}{n} \sum_{i=1}^n (x_i - \tilde{c}_k)^{b-1} \quad (35)$$

$$s_5^{(k)} = \frac{1}{n} \sum_{i=1}^n (x_i - \tilde{c}_k)^b \quad (36)$$

ở đây $\tilde{c}_k$ là xấp xỉ thứ k của giá trị $\tilde{c}$.

Để làm xấp xỉ thứ nhất có thể lấy giá trị $\tilde{c}$ theo 7.1. Nếu $\tilde{c} = x_{\min}$ thì lấy giá trị nhỏ hơn một chút, chẳng hạn $\tilde{c}_1 = x_{\min} - 0,001(x_{\max} - x_{\min})$, với $x_{\max}$ là giá trị lớn nhất trong mẫu. Tương tự, để có $\tilde{c}_{k-1}$, nếu $\tilde{c}_{k-1} \geq x_{\min}$ thì lấy giá trị trong khoảng $\tilde{c}_k$ và $x_{\min}$ (ví dụ $\tilde{c}_{k-1} = 0,5(\tilde{c}_k + x_{\min})$).

Quá trình xấp xỉ liên tiếp sẽ kết thúc khi thỏa mãn điều kiện:

$$G_c^{(k)} = \left| nb \frac{s_4^{(k)}}{s_5^{(k)}} - (b-1)s_1^{(k)} \right| < \varepsilon \quad (37)$$

Với ε là số dương nhỏ cho trước, ví dụ $\varepsilon = 10^{-5}$.

Ước lượng $\tilde{a}$ của a nhận được bằng cách thay c bằng $\tilde{c}$ vào phương trình (30).

7.3. Nếu $b \leq 1$, ước lượng $\tilde{c}$ của c được nhận như sau

$$\tilde{c} = x_{\min}$$

Ước lượng $\tilde{a}$ của a nhận được bằng cách thay c bằng $\tilde{c}$ vào công thức (10) ở điều 5.

8. Ước lượng của các tham số a, b, c

Trong trường hợp này, các ước lượng bằng phương pháp mô men thu được theo cách sau:

1) Tính các đặc trưng mẫu $\bar{x}$ và s theo các công thức (25), (26) và

$$\rho_v = \frac{\frac{1}{(n-1)(n-2)} \sum_{i=1}^n (x_i - \bar{x})^3}{s^3} \quad (39)$$

2) Thay ρ_b bằng ρ_v và căn cứ vào giá trị này (có thể dùng nội suy tuyến tính) để tìm trong Bảng 2 các giá trị b , g_b , K_b tương ứng.

3) Lấy giá trị b tìm được làm ước lượng $\bar{b}$.

4) Các ước lượng $\bar{a}$ và $\bar{c}$ của a và c được tìm theo phương pháp ở 7.1 với điều kiện $b = \bar{b}$.

9. Ước lượng của kỳ vọng, phương sai, hệ số biến động

Ước lượng của kỳ vọng, phương sai, hệ số biến động của đại lượng ngẫu nhiên x có phân bố Weibull (ba tham số) được xác định như sau:

1) Ước lượng các tham số chưa biết theo các phương pháp đã nêu từ điều 5 đến điều 8;

2) Ứng dụng công thức (3) để ước lượng kỳ vọng, nếu tham số nào trong đó chưa biết thì thay bằng ước lượng của nó, còn hệ số K_b tìm trong Bảng 2;

3) Ứng dụng công thức (3) để ước lượng phương sai, nếu tham số nào trong đó chưa biết thì thay bằng ước lượng của nó, còn hệ số g_b tìm trong Bảng 2;

4) Dùng công thức (5) để ước lượng hệ số biến động, nếu tham số nào trong đó chưa biết thì thay bằng ước lượng của nó, còn hệ số K_b và g_b tìm trong Bảng 2.

10. Khoảng tin cậy của tham số a khi các tham số b và c đã biết

Khoảng tin cậy hai phía ($a_1; a_2$) của tham số a khi giá trị của các tham số b và c đã biết với mức tin cậy $1 - \alpha$ xác định bởi

$$a_1 = \left(\frac{2 \sum_{i=1}^n (x_i - c)^b}{\chi^2_{1-\frac{\alpha}{2}}(2n)} \right)^{\frac{1}{b}} \quad (40)$$

$$a_2 = \left(\frac{2 \sum_{i=1}^n (x_i - c)^b}{\chi^2_{\frac{\alpha}{2}}(2n)} \right)^{\frac{1}{b}} \quad (41)$$

trong đó $\chi^2_q(\nu)$ là giá trị q – phân vị của phân bố χ^2 với ν bậc tự do, giá trị này cho trong Bảng 6.

Khoảng tin cậy một phía với mức tin cậy $1 - \alpha$ xác định bởi a_1 nếu khoảng là $(a_1; +\infty)$ hoặc a_2 nếu khoảng là $(-\infty; a_2)$, nhưng trong công thức (40) và (41), thay $\alpha/2$ bằng α

11. Khoảng tin cậy của các tham số a và b khi tham số c đã biết

11.1. Khoảng tin cậy hai phía ($a_1; a_2$) với mức tin cậy $1 - \alpha$ của tham số a khi b chưa biết và c đã biết được xác định bởi công thức

$$a_1 = \tilde{a} \exp \left(-\frac{z_{1-\frac{\alpha}{2}}}{\tilde{b}} \right) \quad (42)$$

$$a_2 = \tilde{a} \exp \left(\frac{z_{1-\frac{\alpha}{2}}}{\tilde{b}} \right) \quad (43)$$

trong đó $\tilde{a}$ và $\tilde{b}$ là các ước lượng của a và b theo phương pháp hợp lý cực đại nêu trong 6.2.

Phân vị z_q tìm trong Bảng 3 khi $5 \leq n \leq 120$. Nếu $n > 120$ có thể dùng công thức gần đúng

$$Z_q = u_q \sqrt{\frac{1,108}{n}} \quad (44)$$

trong đó u_q là q – phân vị của phân bố chuẩn tra trong Bảng 5.

11.2. Khoảng tin cậy hai phía ($b_1; b_2$) với mức tin cậy $1 - \alpha$ của tham số b khi a chưa biết và c đã biết được xác định bởi công thức

$$b_1 = \frac{\tilde{b}}{\ell_1^{\frac{1}{2}}} \quad (45)$$

$$b_2 = \frac{\tilde{b}}{\ell_2^{\frac{1}{2}}} \quad (46)$$

trong đó $\tilde{b}$ là ước lượng hợp lý cực đại của tham số b theo 7.2. Phân vị ℓ_p tìm trong Bảng 4 nếu $5 \leq n \leq 120$. Nếu $n > 120$ thì áp dụng công thức gần đúng

$$\ell_q = 1 - u_q \sqrt{\frac{0,608}{n}} \quad (47)$$

trong đó u_q là q – phân vị của phân bố chuẩn chuẩn hóa tra trong Bảng 5.

11.3. Để xác định khoảng tin cậy một phía với mức tin cậy $1 - \alpha$ của tham số b cần xác định b_1 nếu khoảng tin cậy là $(b_1; +\infty)$ hoặc b_2 nếu khoảng tin cậy là $(-\infty; b_2)$ và trong các giá trị của phân vị phải thay $\alpha/2$ bằng α .

12. Các bảng số

Bảng 1 – Các hệ số không chệch $M(n)$, $B(n)$

n	$M(n)$	$B(n)$	n	$M(n)$	$B(n)$
5	0,738	0,669	42	0,963	0,968
6	0,778	0,752	44	0,965	0,970
7	0,806	0,792	46	0,966	0,971
8	0,831	0,820	48	0,967	0,972
9	0,848	0,842	50	0,969	0,973
10	0,863	0,859	52	0,970	0,974
11	0,875	0,872	54	0,971	0,975
12	0,884	0,883	56	0,972	0,976
13	0,893	0,893	58	0,973	0,977
14	0,900	0,901	60	0,974	0,978
15	0,906	0,908	62	0,975	0,979
16	0,912	0,914	64	0,976	0,980
18	0,921	0,923	66	0,976	0,980
20	0,928	0,931	68	0,977	0,981
22	0,934	0,938	70	0,978	0,981
24	0,939	0,943	72	0,978	0,982
26	0,943	0,947	74	0,979	0,982
28	0,947	0,951	76	0,979	0,983
30	0,950	0,955	78	0,980	0,983
32	0,953	0,958	80	0,980	0,984

34	0,955	0,960	85	0,982	0,985
36	0,957	0,962	90	0,983	0,986
38	0,959	0,964	100	0,984	0,987
40	0,961	0,966	120	0,986	0,990

Bảng 2 – Các đặc trưng cơ bản của phân bố Weibull

b	K_b	g_b	v_b	b
0,200	120,0	1901	15,84	190,1
0,210	80,36	1091	13,58	144,2
0,220	56,33	665,1	11,81	112,3
0,230	41,06	426,7	10,39	89,46
0,240	30,94	286,1	9,248	72,69
0,250	24,00	199,4	8,307	60,09
0,260	19,09	143,6	7,524	50,45
0,270	15,51	106,5	6,865	42,94
0,280	12,85	81,03	6,304	36,99
0,290	10,83	63,06	5,824	32,22
0,300	9,261	50,08	5,408	28,33
0,310	8,024	40,48	5,045	25,14
0,320	7,035	33,25	4,727	22,48
0,330	6,234	27,71	4,445	20,25
0,340	5,575	23,39	4,195	18,36
0,350	5,029	19,98	3,972	16,74
0,360	4,571	17,24	3,771	15,35
0,370	4,184	15,02	3,590	14,15
0,380	3,853	13,20	3,426	13,09
0,390	3,569	11,70	3,277	12,17
0,400	3,323	10,44	3,141	11,35
0,410	3,109	9,377	3,016	10,63
0,420	2,921	8,475	2,901	9,983
0,430	2,756	7,703	2,795	9,404
0,440	2,609	7,037	2,697	8,883
0,450	2,479	6,460	2,606	8,413
0,460	2,362	5,956	2,522	7,986
0,470	2,257	5,515	2,443	7,597
0,480	2,163	5,125	2,370	7,243
0,490	2,077	4,780	2,301	6,917

0,500	2,000	4,472	2,236	6,619
0,510	1,930	4,197	2,175	6,343
0,520	1,865	3,951	2,118	6,089
0,530	1,806	3,728	2,064	5,853
0,540	1,752	3,528	2,013	5,634
0,550	1,702	3,345	1,965	5,431
0,560	1,657	3,179	1,919	5,241
0,570	1,614	3,028	1,876	5,063
0,580	1,575	2,889	1,835	4,896
0,590	1,538	2,762	1,795	4,740
0,600	1,505	2,645	1,758	4,593
0,610	1,473	2,537	1,722	4,455
0,620	1,444	2,437	1,688	4,325
0,630	1,416	2,345	1,656	4,201
0,640	1,390	2,259	1,625	4,085

Bảng 2 (tiếp theo)

b	K_b	g_b	v_b	b
0,650	1,366	2,179	1,595	3,974
0,660	1,344	2,104	1,566	3,869
0,670	1,322	2,035	1,539	3,770
0,680	1,302	1,970	1,512	3,675
0,690	1,284	1,908	1,487	3,585
0,700	1,266	1,851	1,462	3,498
0,710	1,249	1,797	1,439	3,416
0,720	1,233	1,747	1,416	3,338
0,730	1,218	1,699	1,394	3,262
0,740	1,204	1,653	1,373	3,190
0,750	1,191	1,611	1,353	3,121
0,760	1,178	1,570	1,333	3,005
0,770	1,166	1,532	1,314	2,991
0,780	1,154	1,496	1,296	2,930
0,790	1,143	1,461	1,278	2,871
0,800	1,133	1,428	1,261	2,815
0,810	1,123	1,397	1,244	2,760
0,820	1,114	1,367	1,227	2,707
0,830	1,105	1,339	1,212	2,657