ESTIMATION OF LOG X BASED ON REGRESSION TECHNIQUE

M. VIDYA BHALCHANDRA,

Abstract: In this paper, the authors have developed a regression method to estimate values of log x where x > 1; log x denotes logarithm of x to the base e. The method takes advantage of the high correlation coefficient that exists between x and log x. In fact, the developed method approximates the logarithmic function by a linear function.

It is believed that the paper would help the computer scientists to develop simple program(s) for estimating values of log x, provided x > 1.

Keywords: Regression equation, correlation coefficient, logarithmic function.

Theorem 1 Let $x \sim U(n, n+1)$.

Then,

$$\rho(x, \log x) = \frac{2n(n+1)\left[\log \frac{n}{n+1}\right] + (2n+1)}{\sqrt{\frac{4}{3}\left[1 - n(n+1)\left\{\log\frac{n+1}{n}\right\}^2\right]}}$$

Proof:

Note that $EX = n + \frac{1}{2}$ and ... (1)

$V(X) = \frac{1}{12}$ ... (2)

Now, $E \log x = \int_n^{n+1} \log x \, dx = (n+1)\log(n+1) - n\log n - 1$ ... (3)

$E(\log x)^2$

$= \int_n^{n+1} (\log x)^2 \, dx$

$= (n+1)[\log(n+1)]^2 - n[\log n]^2 - 2[(n+1)\log(n+1) - n\log n - 1]$ ... (4)

Hence, $V[\log x] = E(\log x)^2 - [E\log(x)]^2$

$= 1 - n(n+1)\left[\log\frac{n+1}{n}\right]^2$ ... (5)

(Using (3) and (4))

Now, $Ex.\log x = \int_n^{n+1} x.\log x \, dx = \frac{1}{2}[(n+1)^2\log(n+1) - n^2\log n] - \frac{1}{4}(2n+1)$ ... (6)

So, $Cov(X, \log X) = Ex.\log x - Ex.\ E\log x$

$= \frac{n(n+1)}{2}\left[\log\frac{n}{n+1}\right] + \frac{2n+1}{4}$ ... (7)

(Using (1), (3) and (6))

Finally,

$$\rho(x, \log x) = \frac{Cov(X, \log X)}{\sqrt{V(X).\ V(\log X)}}$$

$$= \frac{2n(n+1)\left[\log \frac{n}{n+1}\right] + (2n+1)}{\sqrt{\frac{4}{3}\left[1 - n(n+1)\left\{\log\frac{n+1}{n}\right\}^2\right]}}$$

(Using (2), (5) and (7))

Proof is complete.

Note 1: Table(1) gives values of $\rho(x, \log x)$ for some values of n.

N	ρ(x, log x)
1	0.996065
2	0.998638
3	0.999312
5	0.999723
10	0.999924
50	0.999997
100	0.999999

Remark 1: Table (1) indicates that the correlation coefficient between x and log x is "close" to 1; further, ρ(x, log x) increases with n.
This encourages us to proceed to regression analysis from the point of view of estimating log x for a given value of x.
Theorem 2 The regression equation of log x on x is given by

$$\log x = \left[6n(n+1)\log\frac{n}{n+1} + 3(2n+1)\right].x - \left[6n(n+1)\log\frac{n}{n+1} + 3(2n+1)\right].\left(n + \frac{1}{2}\right) + \left[(n+1)\log(n+1) - n\log n - 1\right]$$

Proof: The regression equation of log x on x is given by

$$[\log x - E\log x] = b[X - EX], \quad \ldots \quad (8)$$

where b (the regression coefficient of log x on x)

$$= \frac{Cov(x,\ \log x)}{\sigma x^2}$$

$$= \frac{\frac{n(n+1)}{2}\log\frac{n}{n+1} + \frac{2n+1}{4}}{\frac{1}{12}} \quad \ldots \text{ (using (2) and (7))}$$

$$= 6n(n+1)\ \log\frac{n}{n+1} + 3(2n+1) \quad \ldots \quad (9)$$

Using (1) (3) and (9) the equation (8) reduces to

$$\log x = \left[6n(n+1)\log\frac{n}{n+1} + 3(2n+1)\right].x - \left[6n(n+1)\log\frac{n}{n+1} + 3(2n+1)\right].\left(n + \frac{1}{2}\right) + \left[(n+1)\log(n+1) - n\log n - 1\right]. \quad \ldots \quad (10)$$

Proof is complete.
Note 2:
In the following, the performance of the regression equation (10) is assessed through illustrations.
Illustration 1:
Let X ~ U(2, 3).
The regression equation (10) simplifies to
log x = 0.403256108.x – 0.098597765.
Table (2) is self explanatory. In the following tables, $\log x^*$ represents the regression estimates that are obtained by using (10).

Table(2)

x	log x	log x*	% relative error
2.1	0.741937344	0.748240061	0.8495
2.2	0.78845736	0.788565672	0.0137
2.3	0.832909122	0.828891283	0.4824
2.4	0.875468737	0.869216894	0.7141
2.5	0.916290731	0.909542505	0.7365
2.6	0.955511445	0.949868115	0.5906
2.7	0.993251773	0.990193726	0.3079
2.8	1.029619417	1.030519337	0.0874
2.9	1.064710737	1.070844948	0.5761

It may be observed that the relative errors in the regression estimates are less than 1%.

Illustration 2:
Let X ~ U(5, 6).
The regression equation (10) simplifies to log x = 0.182119776.x + 0.701708485.

Table(3)

x	log x	log x*	% relative error
5.1	1.62924054	1.630519343	0.0785
5.2	1.648658626	1.64873132	0.0044
5.3	1.667706821	1.666943298	0.0458
5.4	1.686398954	1.685155275	0.0737
5.5	1.704748092	1.703367253	0.0810
5.6	1.722766598	1.721579231	0.0689
5.7	1.740466175	1.739791208	0.0388
5.8	1.757857918	1.758003186	0.0083
5.9	1.774952351	1.776215163	0.0711

It may be observed that the relative errors in the regression estimates are less than 0.09%.

Illustration 3:
Let X ~ U(50, 51). The regression equation (10) simplifies to log x = 0.019802373.x + 2.921937161.

Table(4)

x	log x	log x*	% relative error
50.1	3.914021008	3.914036048	0.00038425
50.2	3.916015027	3.916016286	0.00003215
50.3	3.918005077	3.917996523	0.0002183
50.4	3.919991175	3.91997676	0.0003677
50.5	3.921973336	3.921956998	0.0004166
50.6	3.923951576	3.923937235	0.0003655
50.7	3.925925911	3.925917472	0.0002149
50.8	3.927896355	3.927897709	0.0000345
50.9	3.929862924	3.929877947	0.0003823

It may be observed that the relative errors in the regression estimates are less than 0.0005%.

Remark 2:
It has been observed that as n increases, % relative error decreases.
However, it may be noted that the performance of regression equation is not satisfactory if n is 1 or near 1, as demonstrated below.

The regression equation when n = 1 is, log x = 0.682233833. x – 0.637056388. The corresponding calculations are shown in Table (5).

Table(5)

x	log x	log x*	% relative error
1.1	0.095310179	0.113400827	18.9808
1.2	0.182321556	0.181624211	0.3825
1.3	0.262364264	0.249847594	4.7707
1.4	0.336472236	0.318070978	5.4689
1.5	0.405465108	0.386294361	4.7281
1.6	0.470003629	0.454517744	3.2948
1.7	0.530628251	0.522741128	1.4864
1.8	0.587786664	0.590964511	0.540642
1.9	0.641853886	0.659187894	2.700621

It may be noted that the relative percent error in the estimates are large, as large as 19% (corresponding to x = 1.1).
The performance of the regression equation may be improved by estimating the % relative error, and then incorporating the same in the original regression equation [1] or by fitting a regression equation through multiple regression technique [2]. However, this has not been attempted in this paper.

Conclusion:
Under the assumption that X ~ U(n, n+1), log x may be estimated reasonably well , except in some neighborhood of 1.

References:

1. A statistical Method To Study The Exponential Function. Vidya Bhalachandra and M.S. Bhalachandra 2012.
2. Estimating Exponential function (e^x) by Multiple Regression Analysis. M.S. Bhalachandra and Vidya Bhalachandra, 2012.
3. Estimation of a^x Using Regression Technique by M. S. Bhalachandra, 2013.

***

Student,Second year B.sc.,Nashik – 422002,India.
E Mail: msbhalachandra@yahoo.co.in