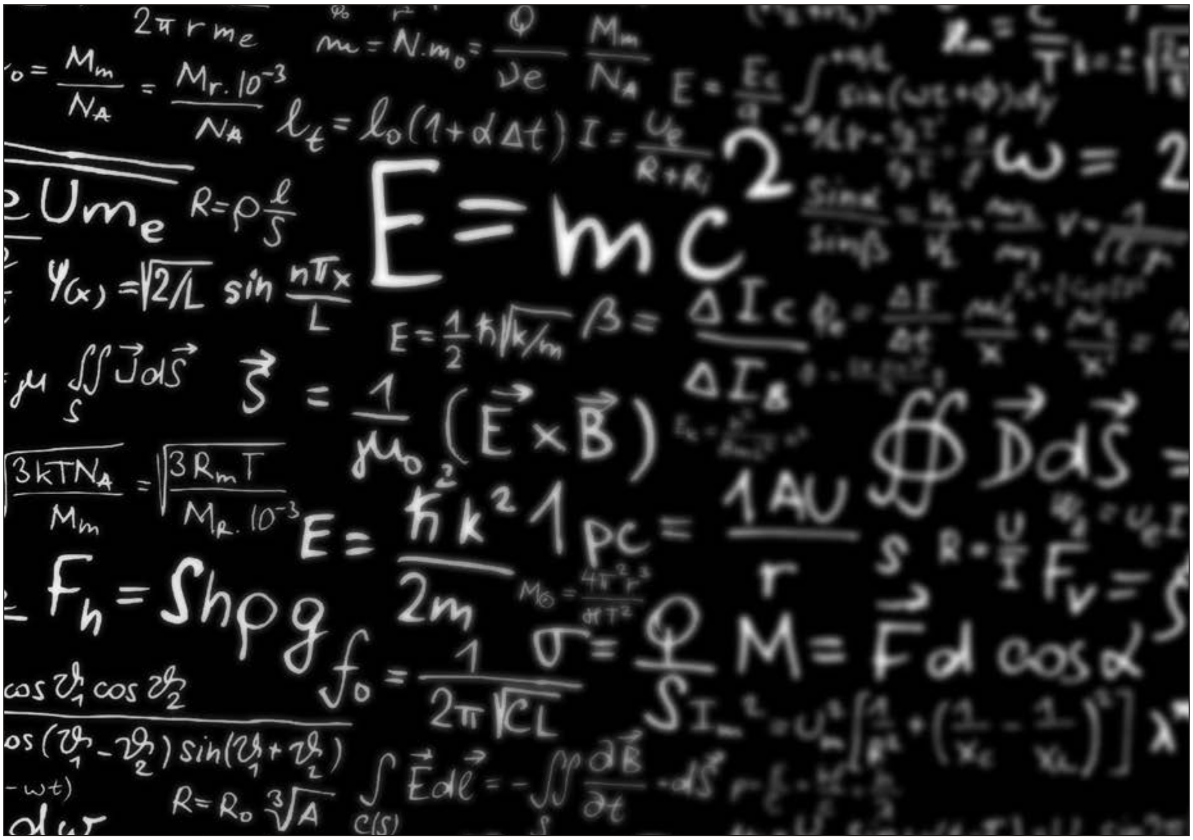

Asian Journal of Mathematics & Statistics

ISSN 1994-5418

science alert

ANSInet
an open access publisher
http://ansinet.com

Asian Journal of Mathematics and Statistics 5 (3): 93-98, 2012
ISSN 1994-5418 / DOI: 10.3923/ajms.2012.93.98

Multiple Linear Regression Formula for the Probability of the Average Daily Solar Energy using the Queue System

Mohammed Mohammed El Genidy
Department of Mathematics, Faculty of Science, Port Said University, Port-Said, Egypt

ABSTRACT

The multiple linear regression formula of the probability of the averaged daily solar energy reaching a specific location on the earth's surface in a calendar month was obtained with the assumption that the arrival process of clouds and solar energy during the day follows the exponential distribution. This formula enables any user to find out some of the required information such as knowing the maximum probability for the averaged daily solar energy and the amount of the corresponding clouds. In addition, the cumulative distribution functions of this probability was obtained.

Key words: Multiple linear regression, daily solar energy, exponential distribution, cumulative distribution function

INTRODUCTION

Jin *et al*. (2006) consider an algorithm to estimate continuous-time traffic speed using multiple regression model. Mechlouch and Brahim (2008) carried out study in a global solar radiation model for the design of solar energy systems. Khooharo *et al*. (2006) have done work a linear regression model to study the relationship of pesticide imports with agricultural productivity growth in Pakistan. Okereke (2011b) considered some consequences of adding a constant to at least one of the variables in the simple linear regression model. Safari and Gasore (2009) estimated the global solar radiation in Rwanda using empirical models. Olaomi and Ifederu (2008) carried out study in understanding estimators of linear regression model with AR(1) error which are correlated with exponential regressor. Okereke (2011a) has done work in effect of transformation on the parameter estimates of a simple linear regression model: A case study of division of variables by constants. Ekpenyong *et al*. (2008) applied the polynomial (non linear) regression method for improved estimation based on sampling. Mardikyan and Darcan (2006) have work done on a software tool for regression analysis and its assumptions.

In this study, the probability of the amounts of daily solar energy and the amount of the corresponding clouds $p(N, T)$ has three cases of events where N is a continuous random variable represents the amount of daily solar energy and T is a continuous random variable represents the amount of the corresponding clouds.

The events will be as follows:

A_1 : The event that the amount of daily solar energy is $(N = n)$ and the corresponding amount of the clouds is $(T = t)$ where n, t are positive finite quantities and not equal zero

A_2 : The event that $t \to 0$ (t is very small) and n is positive finite quantity and approaching from the maximum value of the daily solar energy

A_3 : The event that $n \to 0$ (n is very small) and t is positive finite quantity and approaching from the maximum value of the cloud amount

The all previous events are independent and the probability p(N,T) has stationary independent increments. It is clear that:

$$\begin{aligned} p(N,T) &= p(A_1)+p(A_2)+p(A_3) \\ &= \lambda e^{-\lambda n}\mu e^{-\mu t}+\mu\lambda e^{-\lambda n}+\mu\lambda e^{-\mu t} \\ &= e^{-n\lambda}\lambda\mu+e^{-t\mu}\lambda\mu+e^{-n\lambda-t\mu}\lambda\mu \end{aligned}$$

where, the arrival processes of the solar energy and the clouds follow the exponential distribution with arrival rates $\lambda>0$, $\mu>0$, respectively then:

The marginal density function of N is:

$$f_N(n)=\int_a^b p(n,t)\,dt$$

where, a = min (T) and b = max (T), then:

$$\begin{aligned} f_N(n) &= \int_a^b (e^{-n\lambda}\lambda\mu + e^{-t\mu}\lambda\mu + e^{-n\lambda-t\mu}\lambda\mu)\,dt \\ &= e^{-n\lambda-a\mu}\lambda(1+e^{n\lambda}-ae^{a\mu}\mu)-e^{-n\lambda-b\mu}\lambda(1+e^{n\lambda}-be^{b\mu}\mu) \end{aligned}$$

Similarly, the marginal density function of T is:

$$f_T(t)=\int_c^d p(n,t)\,dn$$

where, c = min (N) and d = max (N), then:

$$\begin{aligned} f_T(t) &= \int_c^d (e^{-n\lambda}\lambda\mu + e^{-t\mu}\lambda\mu + e^{-n\lambda-t\mu}\lambda\mu)\,dn \\ &= e^{-c\lambda-t\mu}(1+e^{t\mu}-ce^{c\lambda}\lambda)\mu-e^{-d\lambda-t\mu}(1+e^{t\mu}-de^{d\lambda}\lambda)\,\mu \end{aligned}$$

The average daily solar energy $E(N) = \int_c^d n f_N(n)\,dn =$

$$\begin{aligned} &\frac{1}{2\lambda}(e^{-c\lambda-a\mu-b\mu}(-2e^{a\mu}+2e^{b\mu}-2ce^{a\mu}\lambda+2ce^{b\mu}\lambda+c^2e^{c\lambda+a\mu}\lambda^2-c^2e^{c\lambda+b\mu}\lambda^2-2ae^{a\mu+b\mu}\mu+2be^{a\mu+b\mu}\mu \\ &-2ace^{a\mu+b\mu}\lambda\mu+2bce^{a\mu+b\mu}\lambda\mu))-\frac{1}{2\lambda}(e^{-d\lambda-a\mu-b\mu}(-2e^{a\mu}+2e^{b\mu}-2de^{a\mu}\lambda+2de^{b\mu}\lambda+d^2e^{d\lambda+a\mu}\lambda^2 \\ &-d^2e^{d\lambda+b\mu}\lambda^2-2ae^{a\mu+b\mu}\mu+2be^{a\mu+b\mu}\mu-2ade^{a\mu+b\mu}\lambda\mu+2bde^{a\mu+b\mu}\lambda\mu)) \end{aligned}$$

The average of the clouds $E(T) = \int_a^b t f_T(t)\, dt =$

$$\frac{1}{2\mu}\left(e^{-c\lambda-d\mu-a\mu}\left(-2e^{c\lambda}+2e^{d\lambda}-2ce^{c\lambda+d\lambda}\lambda+2de^{c\lambda+d\lambda}\lambda-2ae^{c\lambda}\mu+2ae^{d\lambda}\mu-2ace^{c\lambda+d\lambda}\lambda\mu+2ade^{c\lambda+d\lambda}\lambda\mu\right.\right.$$
$$\left.\left.+a^2e^{c\lambda+a\mu}\mu^2-a^2e^{d\lambda+a\mu}\mu^2\right)\right)-\frac{1}{2\mu}\left(e^{-c\lambda-d\lambda-b\mu}\left(-2e^{c\lambda}+2e^{d\lambda}-2ce^{c\lambda+d\lambda}\lambda+2de^{c\lambda+d\lambda}\lambda-2be^{c\lambda}\mu\right.\right.$$
$$\left.\left.+2be^{d\lambda}\mu-2bce^{c\lambda+d\lambda}\lambda\mu+2bde^{c\lambda+d\lambda}\lambda\mu+b^2e^{c\lambda+b\mu}\mu^2-b^2e^{d\lambda+b\mu}\mu^2\right)\right)$$

As before the formulas of E(N) and E(T) need many arithmetic procedures, so that we use the command FIT in mathematica program to obtain the simple formula of p(n, t) as follows:

$$\text{Fit }[\{\{t_i,n_1,p_{n_1}(t_i)\},\{t_i,n_2,p_{n_2}(t_i)\},\ldots,\{t_i,n_m,p_{n_m}(t_i)\},\{t_1,n_j,p_{n_j}(t_1)\},\{t_2,n_j,p_{n_j}(t_2)\},$$
$$\ldots,\{t_r,n_j,p_{n_j}(t_r)\},\{t_i,n_j,p_{n_j}(t_i)\},\{t_i,n_j,p_{n_j}(t_i)\},\ldots,\{t_r,n_m,p_{n_m}(t_r)\}\},\{1,T,N\},\{T,N\}]$$

where, $N = n_1, n_2, n_3, \ldots, n_m$, $T = t_1, t_2, t_3, \ldots, t_r$, $I = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, m$.

For any given data of the random variables T and N and run the previous command FIT, the formula of multi linear regression of the probability density function p(n, t) obtained as follows:

$$p(n, t) = k_1+k_2n+k_3t$$

where, k_1, k_2, k_3 are constants.

Thus the cumulative distribution function P(N<n, T<t) will take the formula:

$$P(N<n\,,\ T<t) = \int_c^n\int_a^t p(u_1,u_2)\,du_2\,du_1 =$$
$$-\frac{1}{2}c^2k_2\,(-a+t)+\frac{1}{2}k_2n^2\,(-a+t)+$$
$$\frac{1}{2}c\,(a-t)(2k_1+ak_3+k_3t)-\frac{1}{2}n\,(a-t)\,(2k_1+ak_3+k_3t)$$

Theorem: If N and T are independent random variables then:

$$E(N+T) = -\frac{1}{3}(-a+b)c^3k_2+\frac{1}{3}(-a+b)d^3k_2-\frac{1}{4}a^2(-c+d)(2k_1+ck_2+dk_2)+$$
$$\frac{1}{4}b^2(-c+d)(2k_1+ck_2+dk_2)-\frac{1}{3}a^3(-c+d)k_3+\frac{1}{3}b^3(-c+d)k_3-$$
$$\frac{1}{4}(-a+b)c^2(2k_1+ak_3+bk_3)+\frac{1}{4}(-a+b)d^2(2k_1+ak_3+bk_3)$$

Proof: The marginal density function $f_N(n)$ defined as follows:

$$f_N(n) = \int_a^b p(n,t)\,dt = \int_a^b (k_1+k_2\,n+k_3\,t)\,dt =$$
$$-\frac{a^2k_3}{2}+\frac{b^2k_3}{2}-a(k_1+k_2n)+b(k_1+k_2n)$$

Then:

$$E(N)=\int_c^d nf_N\,(n)\,dn=$$
$$-\frac{1}{3}(-a+b)\,c^3\,k_2+\frac{1}{3}(-a+b)\,d^3k_2-$$
$$\frac{1}{4}\,(-a+b)\,c^2\,(2k_1+ak_3+bk_3)+$$
$$\frac{1}{4}\,(-a+b)\,d^2\,(2k_1+ak_3+bk_3)$$

Similarly:

$$f_T(t)=\int_c^d p\,(n,t)dn=\int_c^d(k_1+k_2\,n+k_3\,t)\,dn=$$
$$-\frac{c^2k_2}{2}+\frac{d^2k_2}{2}-c\,(k_1+k_3t)+d\,(k_1+k_3t)$$

Then:

$$E(T)=\int_a^b tf_T\,(t)\,dt=$$
$$-\frac{1}{4}a^2(-c+d)(2k_1+ck_2+dk_2)+$$
$$\frac{1}{4}b^2(-c+d)(2k_1+ck_2+dk_2)-$$
$$\frac{1}{3}a^3(-c+d)k_3+\frac{1}{3}b^3(-c+d)\,k_3$$

where, N and T are independent random variables, then:

$$E(N+T)=E(N)+E(T)=$$
$$-\frac{1}{3}(-a+b)c^3k_2+\frac{1}{3}(-a+b)d^3k_2-\frac{1}{4}a^2(-c+d)(2k_1+ck_2+dk_2)+$$
$$\frac{1}{4}b^2(-c+d)(2k_1+ck_2+dk_2)-\frac{1}{3}a^3(-c+d)k_3+\frac{1}{3}b^3(-c+d)k_3-$$
$$\frac{1}{4}(-a+b)c^2(2k_1+ak_3+bk_3)+\frac{1}{4}(-a+b)d^2(2k_1+ak_3+bk_3)$$

APPLICATION WITH NUMERICAL RESULTS

Suppose the following data are given:

- $\lambda = 10$ (MJ m^{-2}) per day, $\mu = 0.1$ Oktas per day, by using the FIT command on the next data we can obtain the following result:

Fit [{0.1, 20, 0.99005}, {0.2, 19, 0.980199}, {0.3, 18, 0.970446}, {0.4, 17, 0.960789}, {0.5, 16, 0.951229}, {0.6, 15, 0.941765}, {0.7, 14, 0.932394}, {0.8, 13, 0.923116}, {0.9, 12, 0.913931}, {1, 11, 0.904837}, {1.1, 10, 0.895834}, {1.2, 9, 0.88692}, {1.3, 8, 0.878095}, {1.4, 7, 0.869358}, {1.5, 6, 0.860708}, {1.6, 5, 0.852144}, {1.7, 4, 0.843665}, {1.8, 3, 0.83527}, {1.9, 2, 0.826959}, {2, 1, 0.818813}}, {1, t, n}, {t, n}]

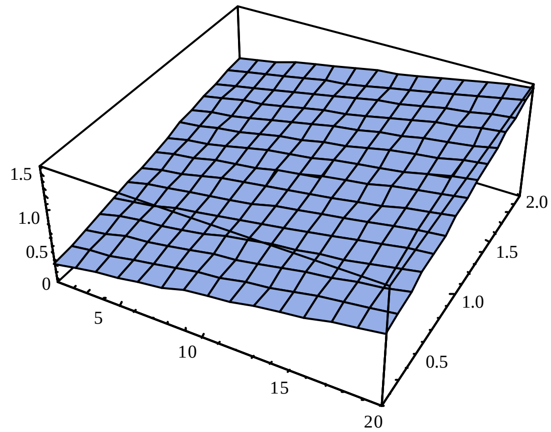

Fig. 1: of Z (t, n)

Then:

$$Z\ (t, n) = 0.15077+0.0402701\ n+0.31259\ t$$

The last formula of Z gives us good estimation for p(n,t), where Z≅p(n,t). Also we can get the Fig. 1 of Z(t,n) as follows:

Plot3D [Z, {N, 1, 20}, {T, 0.1, 2}]

In Fig. 1, there are values of T and N make Z>1 that is contradiction so we must exclude all values of T and N lead to Z larger than 1, the following program in PASCAL language solves this problem and gives us all right values of T and N:

```
Program values of solar energy and clouds_prog;
Var
k1, k2, k3, n, t, z: real;
Begin
Readln (k1, k2, k3, n, t) ;
z : = k1 + k2 * n + k3 * t;
If z<0 or z>1 then
Writeln ('Contradiction happened');
Else
Begin
Writeln (z);
Repeat
t: = t+1;
n: = n+1;
z: = k1+k2 * n+k3 * t;
Writeln (z);
Until (z<0 or z>1);
End;
End.
```

CONCLUSION

The previous program when both variables t and n increase together and if t is constant and n increases, delete the step (t: = t+1) and if n is constant and t increases, delete the step (n: = n+1). Therefore, three cases must be implemented to get all right values of t and n to satisfy the condition $0<z<1$. Thus we can know some important information such as the maximum average daily solar energy or the maximum average of the clouds.

REFERENCES

Ekpenyong, E.J., M.I. Okonnah and E.D. John, 2008. Polynomial (Non Linear) regression method for improved estimation based on sampling. J. Applied Sci., 8: 1597-1599.

Jin, X., S.K. Hong and Q. Ma, 2006. An algorithm to estimate continuous-time traffic speed using multiple regression model. Inform. Technol. J., 5: 281-284.

Khooharo, A.A., R.A. Memon and M.U. Mallah, 2006. A linear regression model to study the relationship of pesticide imports with agricultural productivity growth in Pakistan. J. Applied Sci., 6: 1209-1213.

Mardikyan, S. and O.N. Darcan, 2006. A software tool for regression analysis and its assumptions. Inform. Technol. J., 5: 884-891.

Mechlouch, R.F. and A.B. Brahim, 2008. A global solar radiation model for the design of solar energy systems. Asian J. Scientific Res., 1: 231-238.

Okereke, O.E, 2011a. Effect of transformation on the parameter estimates of a simple linear regression model: A case study of division of variables by constants. Asian J. Math. Stat., 4: 174-180.

Okereke, O.E, 2011b. Some consequences of adding a constant to at least one of variables in the simple linear regression model. Asian J. Math. Stat., 4: 181-185.

Olaomi, J.O. and A. Ifederu, 2008. Understanding estimators of linear regression model with AR(1) error which are correlated with exponential regressor. Asian J. Math. Stat., 1: 14-23.

Safari, B. and J. Gasore, 2009. Estimation of global solar radiation in Rwanda using empirical models. Asian J. Scientific Res., 2: 68-75.