CO_2 EMISSION RANDOMNESS OCCURANCE DUE TO POLYNOMIAL REGRESSION USING RUN TEST

B.SAROJAMMA, A.SRINIVASULU, S.VENKATRAMANA REDDY, P.RAMAKRISHNA REDDY

Abstract: We know that CO_2 causes very much disorder to human beings and animals. These emissions are increasing year to year. In this paper we fitted a 7^{th} degree polynomial to data, and regression coefficients are tested using t-test and coefficient of determination tell that 99.53 percent of CO_2 emission in February is through time and run test is performed for signs of residual such that for identification is residual occur randomly '+' and '-' sides of the fitted polynomial regression model.

Introduction: Regression is various types

1. Simple linear Regression.
2. Multiple linear Regression.
3. Polynomial Regression.
4. Piece wise Regression etc.

Simple linear Regression:The model with a single regressor x that has a relationship with a response y and is a straight line. The simple linear Regression model is

$$Y = \beta_o + \beta_1 x + \varepsilon_t$$

where the intercept is β_o and slope β_1 are unknown constants and ε_t is a disturbance

Multiple linear Regression:

The model with a multiple regressors $x_1, x_2, \ldots\ldots, x_n$ that has a relationship with a response y. The multiple regression model has the relationship is

$Y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots + \beta_n x_n + \varepsilon$

where $\beta_o, \beta_1, \ldots\ldots, \beta_n$ are unknown constants.

$x_1, x_2, \ldots\ldots, x_n$ are independent variables.

y is dependent variable.

Polynomial Regression:

This model with a only regressor i.e. x with powers that has a relationship with y. the n^{th} degree polynomial regression model is

$Y = \beta_o + \beta_1 x + \beta_2 x^2 + \ldots\ldots + \beta_n x^n + \varepsilon_t$

where $\beta_o, \beta_1, \ldots\ldots, \beta_n$ are unknown constants

x is independent variable

y is dependent variable

ε_t is disturbance term

Piece wise Regression:

It is a different form of regression, we have to divide the range of x into segments and fit an appropriate curve in each segment.

A cubic spline with h knots, $t_1 < t_2 < \ldots. < t_n$, continuous, first and second derivatives can be written as

$$\mathbf{E(y) = S(x) = \sum_{j=o}^{3} \beta_{oj} x^j + \sum_{i=1}^{h} \beta_i (x - t_i)^3}$$

where $\mathbf{(x - t_i)_+ = (x - t_i)}$ **if** $\mathbf{x - t_i > 0}$

$= o$ **if** $\mathbf{x - t_i \le 0}$ (i)

Adichie, J.N given "Estimates of regression parameters based on rank tests" Andreews, D.F, has been discussed upon the "Significance tests based on residuals". The choice of variable in observational studies was given by Cox D.R and E.J Snell. "Solving least squares problems" was discussed by Lawson, C.R and R.J. Hanson. Lindley D.V has discussed about "Regression lines and the linear functional relationship".

CO_2 emission causes increase of heat in atmosphere. These CO_2 emission is mainly due to usage of electronic goods like A.C, Refrigirators, Motor bikes etc. In this paper we are fitted 7^{th} degree polynomial and residual are distributed randomly in positive and negative side and are tested using run test.

Methodology: Generally by fitted regression models to data, we tell that how much percent is the time variables, explains time series value using coefficient of determination R^2. But Mean Square Error (MSE) criteria on comparison of different regression models, explain which regression model is Best as compared with others.

In this paper we fitted a 7^{th} degree polynomial regression model i.e

$$U_t = b_o + b_1 t + b_2 t^2 + b_3 t^3 + b_4 t^4 + b_5 t^5 + b_6 t^6 + b_7 t^7$$

where U_t = time series value i.e CO_2 emission for t_h month of February for 55 years data

t = time

$b_o, b_1, b_2, b_3, b_4, b_5, b_6, b_7$ are constants.

Those constants are estimated by the method of ordinary least squares. The fitted Polynomia Regression Model is

$$\mathbf{\widehat{U_t} = \widehat{b_0} + \widehat{b_1} t + \widehat{b_2} t^2 + \widehat{b_3} t^3 + \widehat{b_4} t^4 + \widehat{b_5} t^5 + \widehat{b_6} t^6 + \widehat{b_7} t^7} \quad \text{(ii)}$$

The difference between original value and estimated value gives the error of disturbance value.

Error = original value - estimated value

$= Ut - \widehat{U}_t$ (iii)

Run test for testing random by occurrence of errors after fitting 7^{th} polynomial regression model procedure is as follows.

Upon ignoring values of errors i.e we take only sign i.e '+' and '-' signs of errors

Runs are calculated such that a sequence of '+' signs followed by '-' signs or '-' signs followed by '+' signs

As usually we perform run test

Z = { U – E(U)} / V(U)
where U = number of runs

$$E(U) = \frac{2n_1n_2}{n_1 + n_2} + 1$$

$$V(U) = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}$$

where n_1 = no: of + signs & n_2 = no: of - signs
(iv) After computation of Z calculated value with z critical value i.e 1.96 we conclude the result
(v) conclusion : the positive and negative values are distributed randomly for the fitted 7^{th} degree polynomial model.
Emperical Investigtion: CO_2 is mainly used for fire extinguisher, year to year CO_2 emission is going in an increase manner. It is Global warming, because it produces Heat. The only remedy for decrease of CO_2 emission is increasing plantation of trees and usage of air conditioners and refrigerators causes the emission of CO_2.

In this paper, we fitted 7^{th} degree polynomial regression model to data i.e the fitted model is

$\hat{U}_t = 1.1E^{+12} - 4.3E^{+9}t + 7254809.9t^2 - 6710.036t^3 + 3.69439t^4 - 0.0012t^5 + 2.2E^{-7}t^6 - 1.7E^{11}t^7$

The Regression coefficients are tested by using t-test by using formula

$t = b_j / S.E(b_j)$, j=0,1,2,....7

bj	t_{cal}	p-value
b_o	0.5145	0.6093
b_1	0.5163	0.6081
b_2	0.5182	0.6068
b_3	0.5201	0.6054
b_4	0.5221	0.604
b_5	0.5242	0.6026
b_6	0.5264	0.6011
b_7	0.5285	0.5997

Coefficients of determination (R^2) : 99.53% of CO_2 emission value explains by time 't' , where as the remaining 0.47% by any other factors.
RUN TEST: Run test performed for signs of Residual and is the Residuals are Randomly occur around the fitted polynomial Regression model.
Upon seeing the Residual, we obtain 28 negative signs (n_2) and 27 positive signs (n_1). The total of negative and positive signs i.e $n_1+n_2 = n$ (total no: of years taken for calculation).
Z – Statistic for run test is
Z = {U – E(U)} / V(U)
where U = no; of runs
E(U) = 28.4183
V(U) = 3.6625
Upon substituting U, E(U) and V(U), we obtain
Z – calculated value = 1.7524
Compared with table value i.e 1.96 for 5% L.O.S for large sample test
1.7524 < 1.96
We conclude that the residual values are randomly occur on the positive and negative side of the fitted polynomial regression model.
Conclusion: CO_2 emission is increasing year to year for the month of February, this emission is may be mainly due to deforestation, rapid increase of usage of air conditioners, refrigerators and many more. In this paper we fitted a 7^{th} degree polynomial regression model.
After calculation of residual, we perform run test by taking '+' signs and '-' signs for the randomly occur on the positive and negative side of the fitted polynomial regression model.

References

1. Adichie, J.N (1967),"Estimates of regression parameters based on rank tests," Ann Math. Stat., 38, 894-904.
2. Andrews, D.M (1971), "Significance tests based on residuals," Biometrika,58,139-148.
3. Cox, D.R. and E.J. Snell (1974),"The choice of variables in observational studies, "Appl. Stat.,23,51-59.
4. Lawson, C.R. and R.J. Hanson (1974), Solving Least Squares Problems, Prentice-Hall, Englewood Cliffs, N.J.
5. http://scrippsco2.ucsd.edu/data/atmospheric_co2.html.

***

Department of Statistics, S V University, Tirupati-517502, Andhra Pradesh.
Department of Physics, S V University, Tirupati-517502, Andhra Pradesh.
Corresponding author e-mail : drsvreddy123@gmail.com