

Nonparametric Approach to Multiple Regression

NWANKWO, CHIKE H. (Corresponding author).
Department of Statistics, Nnamdi Azikiwe University, PMB 5025, Awka. Anambra State, Nigeria.
E-mail of the corresponding author: *chikeezeoke@yahoo.com*
OYEKA, I. C. A
Department of Statistics, Nnamdi Azikiwe University, PMB 5025, Awka. Anambra State, Nigeria.
E-mail: *cyprianoyeka@yahoo.com*

Absract
Many real life problems require the use of multiple regression to characterize and solve. Sometimes however the independent variables come in forms that violate the assumptions of the parametric multiple regression. This paper developed a nonparametric approach which uses ranks of both the dependent and independent variables to achieve the objectives of multiple regression. This approach accommodates data as low as the ordinal scale and robust. A prediction procedure which is by interpolation, is also presented.
Keywords: Rank, Robust, Prediction, Interpolation, Rank Correlation

1.0 Introduction

If one has sample observations on dependant variable Y and a set of independent variables X_1, X_2 … X_k, and the dependent variable satisfies the assumptions necessary for the application of parametric regression methods, then one may use these methods to estimate the regression parameters and test appropriate null hypotheses. Thus if y_i is the ith observation independently made on the dependent variable Y and x_{i1}, x_{i2}…x_{ik} are respectively the ith observation made on the independent variables X_1, X_2…X_k, for i = 1, 2 … n then we may fit the multiple regression model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + e_i \quad - \quad - \quad - \quad (1)$$

Where β_j's are partial regression coefficients and e_i's are independent error terms with $E(e_i) = 0$ for i = 1, 2 … n.
Equation 1 may alternatively be expressed in matrix form as

$$\underline{y} = X\underline{\beta} + \underline{e} \quad - \quad - \quad - \quad (2)$$

Where $\underline{y}$ is an n x 1 column vector of observations on the dependent variable Y. X is an n x r = n x (k + 1) design matrix of observations on the independent variables with full column rank r, the number of parameters in the model and k = r – 1, the number of independent variables in the model; $\underline{\beta}$ is an r x 1 = (k + 1) x 1 column vector of partial regression coefficients, and $\underline{e}$ is an n x 1 column vector of error terms, with E ($\underline{e}$) = 0 [2].
Under these assumptions the usual least squares method may be used with Equation 1 or 2 to obtain unbiased estimates of the regression parameters $\underline{\beta}$, from the normal equations [2].

$$X'Y = X'X\underline{b} \quad - \quad - \quad - \quad (3)$$

in matrix form, which when solved yields

$$\underline{b} = \hat{\underline{\beta}} = (X'X)^{-1} X'\underline{y} \quad - \quad - \quad - \quad (4)$$

Where $(X'X)^{-1}$ is the matrix inverse of X'X.
If in addition Y can be assumed to be normally distributed with constant variance, then the usual parametric F test may be used to test the null hypothesis that Equation 1 or 2 fits [2][3], and other desired hypotheses.
However, if these assumptions cannot be satisfied then the parametric tests may not be validly performed. A non-parametric approach such as the following proposed method would then be indicated.

2.0 The Proposed Method

Let y_i, x_{i1}, x_{i2}, … x_{iK} be as defined above for *i* = 1, 2 … n, where populations Y, X_1, X_2…X_k may now be measurements on as low as the ordinal scale and need not be continuous. Furthermore population Y need no longer be normally distributed for desired hypotheses to be tested.
Now instead of using the raw observations or scores in a regression model, these scores are now converted into ranks before being fitted in a regression model.
Thus let r_{iy} be the rank assigned to y_i, the ith observation for population Y, and r_{ij} be the rank assigned to x_{ij} the ith observation from population: X_{ij} for i = 1, 2 … n; j = 1, 2 …k. All tied observations in each variable are, as usual, assigned their mean ranks. With these ranks we now set up a nonparametric multiple regression model of

the ranks assigned to observations from population Y regressing on the ranks assigned to observations from the independent variables X_j's obtaining

$r_{iy} = \beta_0 + \beta_i r_{i1} + \beta_2 r_{i2} + \ldots + \beta_k r_{ik} + e_i$ - - - (5)

where the e_i's are independent error terms with $E(e_i) = 0$. Or expressed in matrix notation

$$\underline{r}_y = R\underline{\beta} + \underline{e} \quad - \quad - \quad - \quad (6)$$

$r'_y = (r_{iy}, r_{2y} \ldots r_{ny})$ is an nx1 column vector of the ranks of Y

$$R = \begin{pmatrix} 1 & r_{11} & r_{12} & r_{13} & . & . & . & r_{1k} \\ 1 & r_{21} & r_{22} & r_{23} & . & . & . & r_{2k} \\ 1 & r_{31} & r_{32} & r_{33} & . & . & . & r_{3k} \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ 1 & r_{n1} & r_{n2} & r_{n3} & . & . & . & r_{nk} \end{pmatrix}$$

is an n x r = n x (k + 1) design matrix of ranks; $\underline{\beta}' = (\beta_0, \beta_1, \beta_2 \ldots \beta_k)$ is an r x 1 = (k + 1) x 1 column vector of partial regression coefficients, not necessarily the same as those in equation 2; and $\underline{e}$ is an n x 1 column vector of independent error terms with $E(\underline{e}) = \underline{0}$. Notice that it is now no longer required that $\underline{e}$ is normally distributed.

The method of least squares [2][1] may now be applied to either equation 5 or 6 to obtain an unbiased estimate $\underline{b}$ of the regression parameters $\underline{\beta}$ from the normal equations.

$$R'\underline{r}_y = R'R\underline{b} \quad - \quad - \quad - \quad (7)$$

Solving equation 7 for $\underline{b}$ we obtain the estimated partial regression coefficient as

$$\underline{b} = (R'R)^{-1} R'\underline{r}_y \quad - \quad - \quad - \quad (8)$$

Where $(R'R)^{-1}$ is the matrix inverse of the matrix $R'R$

Using this estimate we have that the fitted or predicted nonparametric multiple regression equation is

$$\hat{r}_y = R\underline{b} \quad - \quad - \quad - \quad (9)$$

Note that equation 7 may be expressed in terms of the ranks of the sample observations as

$$\begin{pmatrix} \frac{n(n+1)}{2} \\ \sum r_{i1} r_{iy} \\ \sum r_{i2} r_{iy} \\ \sum r_{i3} r_{iy} \\ . \\ . \\ . \\ \sum r_{ik} r_{iy} \end{pmatrix} = \begin{pmatrix} n & \frac{n(n+1)}{2} & \frac{n(n+1)}{2} & . & . & . & \frac{n(n+1)}{2} \\ \frac{n(n+1)}{2} & \sum r_{i1}^2 & \sum r_{i1} r_{i2} & . & . & . & \sum r_{i1} r_{ik} \\ \frac{n(n+1)}{2} & \sum r_{i2} r_{i1} & \sum r_{i2}^2 & . & . & . & \sum r_{i2} r_{ik} \\ \frac{n(n+1)}{2} & \sum r_{i3} r_{i1} & \sum r_{i3} r_{i2} & . & . & . & \sum r_{i3} r_{ik} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ \frac{n(n+1)}{2} & \sum r_{ik} r_{i1} & \sum r_{ik} r_{i2} & . & . & . & \sum r_{ik}^2 \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ . \\ . \\ . \\ b_k \end{pmatrix} \quad - \quad - \quad - \quad (10)$$

To evaluate equation 10, we note that r_{iy} and r_{ij}'s are each ranks and values of the first n positive integers Hence

$$\sum_{i=1}^{n} r_{iy} = \sum_{i=1}^{n} r_{ij} = \frac{n(n+1)}{2} \quad - \quad - \quad - \quad (11)$$

and

$$\sum_{i=1}^{n} r_{iy}^2 = \sum_{i=1}^{n} r_{ij}^2 = \frac{n(n+1)(2n+1)}{6} \quad - \quad - \quad - \quad (12)$$

Also if ρ_{jl} is the Spearman's rank correlation coefficient [4] between X_j and X_l whose ith observations have ranks r_{ij} and r_{il} respectively, then

$$\sum_{i=1}^{n} r_{ij} r_{il} = \frac{n(n^2-1)}{12}\rho_{jl} + \frac{n(n+1)^2}{4} = \frac{n(n+1)}{2}\left(\frac{(n-1)}{2}\rho_{jl} + \frac{(n+1)}{2}\right) \quad - \quad - \quad - \ (13)$$

Using equations 11 – 13 in equation 10 yields, after simplification,

$$\begin{pmatrix} \frac{n+1}{2} \\ \frac{(n+1)}{2}\left(\frac{(n-1)}{6}\rho_{1y} + \frac{(n+1)}{2}\right) \\ \frac{(n+1)}{2}\left(\frac{(n-1)}{6}\rho_{2y} + \frac{(n+1)}{2}\right) \\ \\ \frac{(n+1)}{2}\left(\frac{(n-1)}{6}\rho_{ky} + \frac{(n+1)}{2}\right) \end{pmatrix} = \begin{pmatrix} 1 & \frac{n+1}{2} & \frac{n+1}{2} & \cdots & \frac{n+1}{2} \\ \frac{n+1}{2} & \frac{(n+1)}{2}\frac{(2n+1)}{3} & \frac{(n+1)}{2}\frac{(n-1)}{6}\rho_{12} + \frac{(n+1)}{2} & \cdots & \frac{(n+1)}{2}\frac{(n-1)}{6}\rho_{1k}\frac{(n+1)}{2} \\ \frac{n+1}{2} & \frac{(n+1)}{2}\frac{(n-1)}{6}\rho_{12}\frac{(n+1)}{2} & \frac{(n+1)}{2}\frac{(2n+1)}{3} & \cdots & \frac{(n+1)}{2}\frac{(n-1)}{6}\rho_{2k}\frac{(n+1)}{2} \\ \\ \frac{n+1}{2} & \frac{(n+1)}{2}\left(\frac{(n-1)}{6}\rho_{k1} + \frac{(n+1)}{2}\right) & \frac{(n+1)}{2}\left(\frac{(n-1)}{6}\rho_{k2} + \frac{(n+1)}{2}\right) & \cdots & \frac{(n+1)}{2}\frac{(2n+1)}{3} \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \\ b_k \end{pmatrix} \quad - \quad 14)$$

Application of a few elementary operations on the above equation further reduces the normal equations to the simpler form

$$\begin{pmatrix} \frac{n+1}{2} \\ \rho_{1y} \\ \rho_{2y} \\ \cdot \\ \cdot \\ \rho_{ky} \end{pmatrix} = \begin{pmatrix} 1 & \frac{n+1}{2} & \frac{n+1}{2} & \frac{n+1}{2} & \cdots & \frac{n+1}{2} \\ 0 & 1 & \rho_{12} & \rho_{13} & \cdots & \rho_{1K} \\ 0 & \rho_{21} & 1 & \rho_{23} & \cdots & \rho_{2K} \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & \rho_{K1} & \rho_{K2} & \rho_{K3} & \cdots & 1 \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \cdot \\ \cdot \\ b_k \end{pmatrix} \quad - \quad - \quad (15)$$

Where $\rho_{lj} = \rho_{jl}\,; \rho_{il} = 1\,; for\ l, j = 1,2,3,\ldots k$

Equations 14 and 15 may be solved simultaneously or by the use of appropriate statistical package to obtain the results in equation 8, the required estimated partial regression coefficients b. Although the regression coefficients are now easily obtained using these statistical packages, because nonparametric regression is a relatively novel concept, it may be instructive, for clearer understanding, to illustrate these calculations for cases in which k = 1, 2 and 3.

Thus for k = 1 we have from Equation 14 that

$$\frac{n+1}{2} = b_o + \frac{(n+1)}{2}b_1 \quad - \quad - \quad - \quad (a)$$

and

$$\frac{(n+1)}{2}\frac{(n-1)}{6}\rho_{1y} + \frac{n+1}{2} = \frac{(n+1)}{2}b_o + \frac{(n+1)}{2}\frac{(2n+1)}{3}b_1 \quad \text{Or}$$

$$\frac{(n-1)}{6}\rho_{1y} + \frac{(n+1)}{2} = b_o + \frac{(2n+1)}{3}b_1 \quad - \quad - \quad - \quad (b)$$

Solving (a) and (b) simultaneously yields

$$b_1 = \rho_{1y}\ and\ b_o = \frac{(n+1)}{2}(1-\rho_{1y}) \quad - \quad - \quad -(16)$$

Using these values in equation 9 yields the fitted or predicted nonparametric simple linear regression line in terms of ranks as

$$\hat{r}_{iy} = \frac{(n+1)}{2}(1-\rho_{1y}) + \rho_{1y}r_{i1} = \frac{(n+1)}{2} + \rho_{1y}\left(r_{i1} - \frac{(n+1)}{2}\right) \quad - \quad - \quad - \quad (17)$$

For k = 2 the following normal equation from equation 14 may be used to obtain the required estimates of the regression coefficients;

$$\frac{n+1}{2} = b_o + \frac{(n+1)}{2}b_1 + \frac{(n+1)}{2}b_2 \qquad - \qquad - \qquad - \qquad (c)$$

$$\frac{n(n+1)}{12}\left((n-1)\rho_{1y} + 3(n+1)\right) = \frac{n(n+1)}{2}b_o + \frac{n(n+1)(2n+1)}{6}b_1 + \frac{n(n+1)}{12}\left((n-1)\rho_{12} + \frac{3(n+1)}{2}\right)b_2 \text{ - - - } (d)$$

$$\frac{n(n+1)}{12}\left((n-1)\rho_{2y} + 3(n+1)\right) = \frac{n(n+1)}{2}b_o + \frac{n(n+1)}{12}\left((n-1)\rho_{12} + 3(n+1)\right)b_1 + \frac{n(n+1)(2n+1)}{6}b_2 \text{ - - - } (e)$$

Solving these three equations simultaneously we obtain the estimates of the regression coefficients as

$$b_1 = \frac{\rho_{1y} - \rho_{12}\ \rho_{2y}}{1-\rho_{12}^2}\ ;\ b_2 = \frac{\rho_{2y} - \rho_{12}\ \rho_{1y}}{1-\rho_{12}^2} \qquad - \qquad - \qquad - \qquad (18)$$

Substituting these values into (c) yields

$$b_0 = \frac{n+1}{2} - \frac{(n+1)}{2}\frac{\left(\rho_{1y} - \rho_{12}\ \rho_{2y}\right)}{1-\rho_{12}^2} - \frac{(n+1)}{2}\frac{\left(\rho_{2y} - \rho_{12}\ \rho_{1y}\right)}{1-\rho_{12}^2}$$

OR

$$b_0 = \frac{(n+1)}{2}\frac{\left(1-(\rho_{1y} - \rho_{2y})\right)}{1+\rho_{12}} \qquad - \qquad - \qquad - \qquad (19)$$

With these estimated regression coefficients, the fitted or estimated nonparametric bivariate regression model becomes

$$\hat{r}_{iy} = \frac{(n+1)}{2}\frac{\left(1-\rho_{1y} - \rho_{2y}\right)}{1+\rho_{12}} + \frac{\left(\rho_{1y} - \rho_{12}\rho_{2y}\right)}{1-\rho_{12}^2}r_{i1} + \frac{\left(\rho_{2y} - \rho_{I2}\ \rho_{1y}\right)}{1-\rho_{12}^2}r_{i2} \qquad - \qquad - \qquad - \ (20)$$

For i = 1, 2…n

As a last illustrative example, if there are k = 3 independent variables, then the normal equations (equation 14) may be represented as

$$\frac{n+1}{2} = b_0 + \frac{(n+1)}{2}b_1 + \frac{(n+1)}{2}b_2 + \frac{(n+1)}{2}b_3 \qquad - \qquad - \qquad - \qquad (f)$$

$$\frac{(n-1)}{6}\rho_{1y} + \frac{(n+1)}{2} = b_0 + \frac{(2n+1)}{3}b_1 + \left(\frac{(n-1)}{6}\rho_{12} + \frac{(n+1)}{2}\right)b_2 + \left(\frac{(n-1)}{6}\rho_{13} + \frac{(n+1)}{2}\right)b_3 \text{ - - } \text{-}(g)$$

$$\frac{(n-1)}{6}\rho_{2y} + \frac{(n+1)}{2} = b_0 + \left(\frac{(n-1)}{6}\rho_{i2} + \frac{(n+1)}{2}\right)b_1 + \frac{(2n+1)}{3}b_2 + \left(\frac{(n-1)}{6}\rho_{23} + \frac{(n+1)}{2}\right)b_3 \text{ - -- } (h)$$

$$\frac{(n-1)}{6}\rho_{3y} + \frac{(n+1)}{2} = b_0 + \left(\frac{(n-1)}{6}\rho_{13} + \frac{(n+1)}{2}\right)b_1 + \left(\frac{(n-1)}{6}\rho_{23} + \frac{(n+1)}{2}\right)b_2 + \frac{(2n+1)}{3}b_3 \text{ - - } (i)$$

Expectedly as k, the number of independent variables in the model increases, manual computations become increasingly cumbersome. However, a few algebraic manipulations of the above set of four equations readily yield the estimated regression coefficients.

$$b_2\frac{(\rho_{23} - \rho_{12}\rho_{13})(\rho_{3y} - \rho_{13}\rho_{1y}) - (1-\rho_{13}^2)(\rho_{2y} - \rho_{12}\rho_{1y})}{(\rho_{23} - \rho_{12}\rho_{13})^2 - (1-\rho_{12}^2)(1-\rho_{13}^2)} \qquad - \qquad - \qquad (21)$$

$$b_3\frac{(\rho_{23} - \rho_{12}\rho_{13})(\rho_{2y} - \rho_{12}\rho_{1y}) - (1-\rho_{12}^2)(\rho_{3y} - \rho_{13}\rho_{1y})}{(\rho_{23} - \rho_{12}\rho_{13})^2 - (1-\rho_{12}^2)(1-\rho_{13}^2)} \qquad - \qquad - \qquad - \ (22)$$

The remaining coefficient b_0 and b_1 may similarly be obtained by substitution. Thus

$$b_1 = \frac{(\rho_{13} - \rho_{12}\rho_{23})(\rho_{3y} - \rho_{23}\rho_{2y}) - (1-\rho_{23}^2)(\rho_{1y} - \rho_{12}\rho_{2y})}{(\rho_{13} - \rho_{12}\rho_{23})^2 - (1-\rho_{12}^2)(1-\rho_{13}^2)} \qquad - \qquad - \qquad (23)$$

$$b_0 = \frac{(n+1)}{2}\left(1-(b_1 + b_2 + b_3)\right) \qquad - \qquad - \qquad - \qquad (24)$$

Using the estimated regression coefficients from Equations 4 – 24 we have the resulting fitted non parametric trivariate regression model as

$$\hat{r}_{iy} = b_0 + b_1 r_{i1} + b_2 + r_{i2} + b_3 r_{i3} \qquad - \quad - \quad - \qquad (25)$$

For i = 1, 2…n

In all cases, interest is usually in testing whether the hypothesized regression model is adequate. In other words, interest is in testing the null hypothesis that the population regression coefficients are all equal to zero with the hope of the null hypothesis being rejected. That is interest is in the null hypothesis.

$$H_0 : \underline{\beta} = \underline{O} \text{ versus } H_1 : \underline{\beta} \neq \underline{O} \qquad - \quad - \quad - \qquad (26)$$

If H_o is rejected then further interest may be in making pair-wise comparisons of the regression coefficients to determine whether they are statistically different. The adequacy of the regression model is tested using the Fisher F test which is fairly robust while the pair-wise comparisons are conducted using the student t test.

Note that in terms of ranks the total sum of squares SST for a multiple regression model is

$$SST = \left(r_y - \bar{r}_y\right)'\left(\underline{r}_y - \bar{r}_y\right) = \underline{r}_y{}' \underline{r}_y - \frac{n(n+1)^2}{4}$$

$$= \frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)^2}{4}$$

That is

$$SST = \frac{n\left(n^2-1\right)}{12} \qquad - \quad - \quad - \qquad (27)$$

with n – 1 degrees of freedom

The regression sum of squares SSR is

$$SSR = \left(\hat{\underline{r}}_y - \bar{r}_y\right) = \left(R\underline{b} - \frac{1'}{n}R\right)'\left(R\underline{b} - \frac{1'}{n}R\right)$$

$$= \left(R\underline{b} - \frac{(n+1)}{2}\right)'\left(R\underline{b} - \frac{(n+1)}{2}\underline{b}\right)\left[R\underline{b} - \frac{(n+1)}{2}\right]$$ where 1' = (1, 1,…1) is a 1 x n row vector.

That is

$$SSR = b'\left(R'R - \frac{n(n+1)^2}{4}\right)\underline{b} \qquad - \quad - \quad - \qquad (28)$$

With r – 1 = k degrees of freedom

The error sum of squares, SSE is the difference between the above two sums of squares, namely

$$SSE = SST - SSR = \frac{n\left(n^2-1\right)}{12} - \underline{b}'\left(R'R - \frac{n(n+1)^2}{4}\right)\underline{b} \qquad - \quad - \quad - \qquad (29)$$

These results are summarized in the following analysis of variance table (Table 1)

Table 1: Analysis of Variance Table for Nonparametric Multiple Regression

Source of Variation	**Sum of Squares (SS)**	**Degree of Freedom (df)**	**Mean Sum of Squares (MS)**	**F-Ratio**
Regression	$SSR = \underline{b}'\left(R'R - \frac{n(n+1)^2}{4}\right)\underline{b}$	r – 1 = k	$MSR = \frac{SSR}{r-1}$	$F = \frac{MSR}{MSE}$
Error	$SSE = \frac{n\left(n^2-1\right)}{12} - \underline{b}'\left(R'R - \frac{n(n+1)^2}{4}\right)\underline{b}$	n – r	$MSE = \frac{SSE}{n-r}$	
Total	$SST = \frac{n\left(n^2-1\right)}{12}$	n – 1		

The results of table 1 may be used to test the null hypothesis of equation 26 that is to determine whether the hypothesized regression model is an adequate representation of the relationship between the dependent variable Y and the set of independent variables $X_1, X_2, \ldots X_k$, represented by ranks.

As noted above, the F test is robust under fairly general conditions. Hence the null hypothesis H_o of equation 26 is rejected at a specified significance level α if the F-ratio in table 1 is greater than or equal to the tabulated F-value, that is if

$$F \geq F_{(1-\alpha;\, r-1, n-r)} \qquad - \qquad - \qquad - \qquad (30)$$

Otherwise H_o is accepted.

If H_o is rejected in which case not all the β_j's are equal to zero, then one may proceed to use the usual student t test to further determine which pairs of these parameters may be statistically different.

Finally one may be interested in predicting values of a dependent variable on the basis of predicted or estimated ranks r_{iy} of equation 9. To do this, we may use the method of interpolation. Thus suppose Y_s and Y_z are the observed samples value s of the dependent variable Y with assigned ranks r_{sy} and r_{zy} respectively. Furthermore suppose the rank predicted for a value to be predicted $\hat{y}_t$ of the dependent variable Y is $\hat{r}_{ty}$ which is found to be closer in value to r_s than to r_z, then the predicted or estimated value of Y at point or condition t namely $\hat{y}_t$ is calculated as

$$\hat{y}_t = \left(\frac{\hat{r}_{ty} - r_{zy}}{r_{sy} - r_{zy}}\right) y_s + \left(\frac{r_{sy} - \hat{r}_{ty}}{r_{sy} - r_{zy}}\right) y_z \qquad - \qquad - \qquad - \qquad (31)$$

For t = 1, 2 …

Provided the ranks specified for the observations from the independent variables for which the corresponding value of the dependent variable to be estimated or predicted are within or close to the range of the ranks assigned to the original observations for respective independent variables used in the model.

3.0 Illustrative Example

A researcher is interested in determining the effects age, body weight, blood pressure and packed cell volume (PCV) level have on the heartbeat of postpartum mothers, obtaining the data shown in Table 2 on a random sample of 38 postpartum mothers in a certain hospital maternity.

Table 2: Heartbeat, Age, Body weight, Blood Pressure and PCV Level of a Random Sample of Post Partion Mothers

S/N	Pulse Rate y_i	Rank of y_i (r_{iy})	Mothers Age x_{i1}	Rank of x_{i1} (r_{i1})	Mother weight x_{i2}	Ranks of x_{i2} (r_{i2})	Systolic BP x_{i3}	Rank of x_{i3} (r_{i3})	Diastolic BP x_{i4}	Rank of x_{i4} (r_{i4})	PCV x_{i5}	Rank of x_{i5} (r_{i5})
1.	80	15.5	31	24	75	26.5	110	10.5	65	12.5	34	26.5
2.	86	32.5	30	22.5	55	1	103	3.5	63	7	31	10
3.	90	37	28	18	90	37	170	35.5	80	32	30	4.5
4.	82	24	24	10.5	60	5	102	1.5	68	20	38	33.5
5.	80	15.5	36	32.5	76	29.5	110	10.5	70	25.5	48	38
6.	86	32.5	20	5	59	3.5	110	10.5	60	2.5	35	29.5
7.	90	37	34	30	73	23.5	110	10.5	60	2.5	32	18
8.	88	36	25	12.5	83	34	110	10.5	60	2.5	37	32
9.	74	8	30	22.5	75	26.5	112	17	60	2.5	28	32
10.	82	24	24	10.5	68	10	168	33	80	32	32	18
11.	81	19	21	7	72	20.5	128	26.5	75	28	34	26.5
12.	80	15.5	26	14	78	31.5	110	10.5	65	12.5	30	45
13.	83	28	22	9	72	20.5	120	22.5	70	25.5	31	10
14.	82	24	28	18	68	10	122	24	68	20	40	35
15.	74	8	39	35	65	7	108	5	68	20	32	18
16.	82	24	28	18	61	6	103	3.5	63	7	35	29.5
17.	78	12	28	18	70	15.5	110	10.5	65	12.5	36	31
18.	74	8	18	3	72	20.5	110	10.5	65	12.5	34	26.5
19.	82	24	19	4	73	23.5	110	10.5	63	7	30	4.5
20.	90	37	21	7	78	31.5	110	10.5	65	12.5	32	18
21.	82	34	28	18	75	26.5	170	35.5	85	36.5	32	18
22.	86	32.5	28	18	68	10	112	17	63	7	31	10
23.	85	30	21	7	76	29.5	120	22.5	68	20	41	36
24.	79	13	16	2	90	37	130	28	75	28	43	37
25.	74	8	36	32.5	75	26.5	125	25	68	20	32	18
26.	72	3	38	34	70	15.5	102	1.5	68	20	38	33.5
27.	70	1	41	38	69	13	118	20.5	68	20	31	10
28.	72	3	40	36.5	59	3.5	138	30	80	32	33	24
29.	72	3	32	25.5	58	2	128	26.5	68	20	34	26.5
30.	80	15.5	33	28	70	15.5	140	31	80	32	32	18
31.	81	19	28	18	68	10	170	35.5	85	36.5	31	10
32.	73	5	32	25.5	72	20.5	150	32	80	32	30	4.5
33.	82	24	15	1	70	15.5	112	17	68	20	32	18
34.	81	19	25	12.5	82	33	170	35.5	85	36.5	30	45
35.	74	8	33	28	68	10	172	38	85	36.5	32	18
36.	77	11	33	28	88	35	118	20.5	65	12.5	30	4.5
37.	86	32.5	35	31	90	37	116	19	63	7	32	18
38.	84	29	40	36.5	71	18	133	29	75	28	32	18

We have analysed the data of Table 2 using the proposed nonparametric multiple regression method in this paper, this approach makes the analysis more general and avoids the often restrictive assumptions of parametric regression.

Thus we first rank the observations for each of the five variables from the smallest value ranked 1, to the largest value ranked 38. Tied values in each variable are assigned their mean ranks. The results of the ranking for each variable are shown beside the observations for that variable in Table 2. Pair-wise cross products of these ranks may now be taken to set up the normal equations (Equation 7) which may be solved simultaneously to obtain required estimates of the regression coefficients (equation 8) and hence the desired fitted nonparametric model (equation 9). However, manual computations are relatively difficult and cumbersome. Hence, using SPSS yields the model

$$\hat{r}_{iy} = 31.866 - 0.376 r_{i1} + 0.103 r_{i2} + 0.152 r_{i3} - 0.357 r_{i4} - 0.137 r_{i5} \qquad - \;\; -- \qquad (32)$$

The corresponding analysis of variance table is

Table 3: Analysis of Variance (ANOVA) table for the data of Table 2.

SV	SS	df	MS	F	P-Value
Regression	1205.743	5	241.149	2.207	0.078
Error	3497.126	32	109.285		
Total	4702.868	37			

The ANOVA table shows that the hypothesized regression model is adequate for a significance level of 0.1

References

[1] Draper, N. R. and Smith, H (1998): "Applied Regression Analysis" 3rd Ed." Wiley Series in Probability and Statistics. John Wiley and Sons Inc, New York.
[2] Neter, J., Wasserman, W. and Kutner, M. H. (1983): "Applied Linear Regression Models" Richard D. Irwin Inc, Illinois.
[3] Ryan, T. P. (1997): "Modern Regression Methods". Wiley Series in Probability and Statistics. John Wiley and Sons Inc, New York.
[4] Siegel, S (1956): "Nonparametric Statistics for the Behavioral Sciences". International Student Edition. McGRAW-HILL KOGAKUSHA LTD, Tokyo.