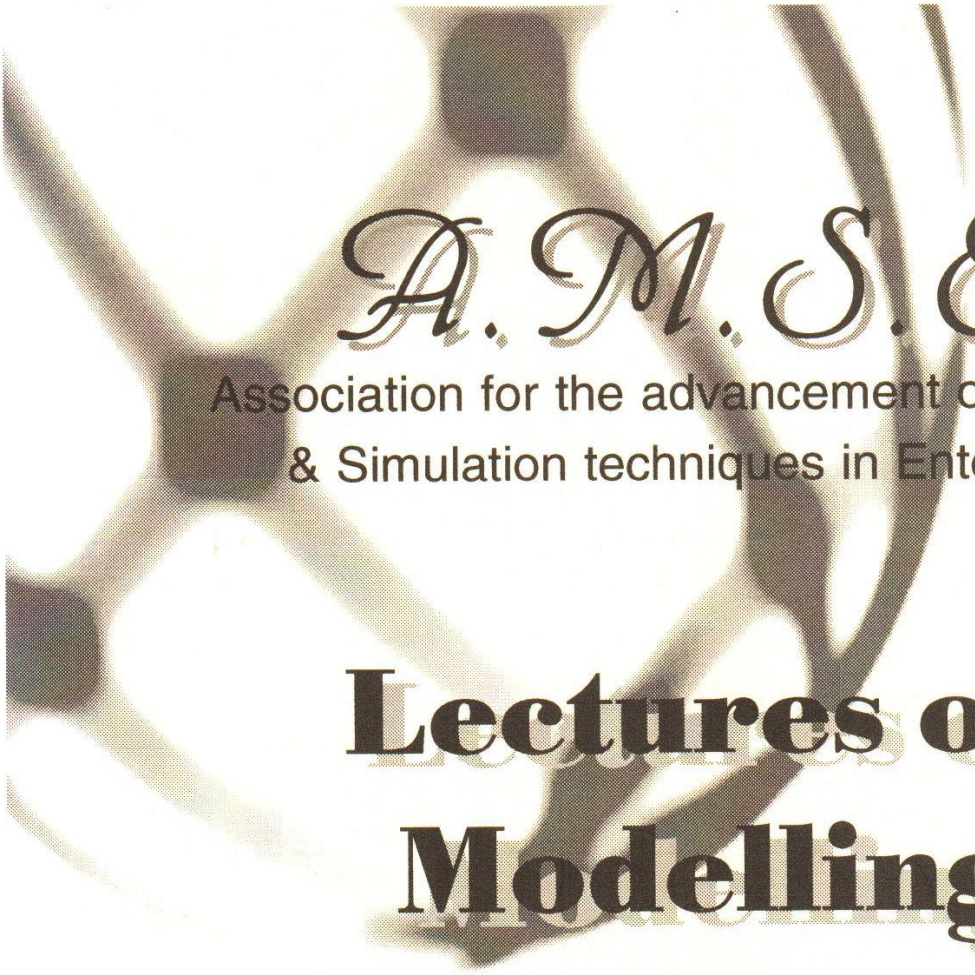


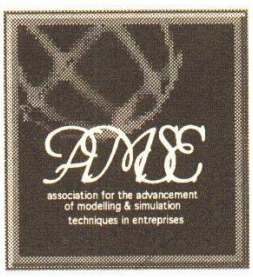
A.M.S.E.

Association for the advancement of Modelling & Simulation techniques in Enterprises

Lectures on Modelling and Simulation

A Selection From AMSE

2006



- "Defect detection in a cylindrical pipe. Application to a natural gas pipeline". M. Guellal; M. C. A. Touabti; H. Abdesselam (Algeria-Malaysia) 100

- "Fuzzy-set based methodology for assessment of seismic response of structures". P. Stemberk; J. Kruis (Czech Republic) 110

- "Modelling of cold start emission control system in IC engine". S. Sendilvelan; R. Anandanatarajan; S. Mohanamurugan (India) 120

Konya (Turkey) August 28-30, 2006

- "A performance comparison of various neural network techniques in predicting the savings deposit insurance fund transferred banks in turkey". M. A. Boyacioglu; Y. Kara (Turkey) 1

- "Simulation, manufacturing and performance of 242 watt free convection heat sink for electrical circuits". M. A. Abdulsattar; A. A. Al-Ajaj; M. K. Hamid (Irad) 10

- "Evaluation of risky investment projects through simulation and response surface methodology". Ö. Armaneri; Ö. Yalçinkaya (Turkey).....21

- "Real coded genetic algorithm based fuzzy control of a brushless DC motor". O. Aydogdu; R. Akkaya (Turkey) 31

- "On some measures of nonlinearity in regression analysis". A. Evren; E. Öztürk (Turkey) 40

- "Multifractal, genetic algorithms and neural network in cervical cell images recognition". L. Hamami; N. Lassovavi; N. Novali (Algeria) 51

- "Balancing of turkey's electricity energy until the year 2020". C. Hamzaçebi; A. Sarucan; F. Kutay ; T. Altinok (Turkey) 62

- "The thermal conductivity correlation of propane by artificial neural networks". E. Ö. Karabulut; M. Koyunku (Turkey) 69

- "Developing a modified mean value SI engine model with EGR". M. Mostofi; A. H. Shamekhi; M. Gorji-Bandpy (Iran) 80

- "Simulation of confined fluid flow witch is subjected to a perturbation". Z. Ouchiha; A. Ghezal; J. C. Loraud; M. Kessal (Algeria-France) 90

- "Improving error metric calculation techniques for 3D mesh generation". R. Samet; E. Ozsavas (Turkey) 99

- "The solution of the causal and anticausal n-dimensional bessel diamond operator and the fourier bessel transform of their B-convolution". M. Z. Sarikaya; H. Yildirim (Turkey) 109

- "Manufacturing line simulator for design and performance evalution". M. Savsar (Kuwait) 119

- "Determination of the endurance state of a specimen by processing its temperature images". M. Selek (Turkey) 129

- "Application of fuzzy logic and image processing in experiments with hardening concrete". P. Stemberk; A. Kohoutková (Czech Republic) 139

- "The investigation of usability of artificial neural networks in the determination of residential real estate prices". S. Yalpir; G. Özhan; G. Tezel (Turkey) 149

- "MEISC – the medium instruction set computers". K. Balasubramian (Turkey) 158

Bahía Blanca (Argentina) September 27-29, 2006

- "Value of the hidden influences and unnoticed factors in information-measurement systems by the method of A. Kaufmann and J. Gil Aluja". R. M. Mirsalimov (Azaerbaijan) 1

- "Dual memory network searches for dis-and associative prototypes in small and medium-sized enterprises". R. O. Dichiara; A. Claverie; B. I. Niel (Argentina) 10

- "A framework of Markov games for multi-agent reinforcement learning". F. F. Farahat (Egypt) 20

- "Proposal of a model of an enterprise development centre for the academic division of informatics and systems of the Juarez Autonomous University of

Selection at the AMSE Conference Konya/Turkey, August 27-29, 2006

On Some Measures of Nonlinearity in Regression Analysis

*A. Evren **E. Öztürk

Yıldız Technical University, Art And Science Faculty ,Statistics Department

Davutpaşa Mah.Davutpaşa Cad. 34210 Esenler/İstanbul

*aevren@hotmail.com **eozturk@yildiz.edu.tr

Abstract

In regression analysis there are some techniques to determine the nonlinearity in the variables. An F test or a comparison between linear correlation coefficient and correlation ratio may be helpful. The third method is based on entropy and mutual information measures. In this paper we studied these three tests and then compared the results obtained by each test.

Key words

Nonlinear regression, dependence measures,mutual information.

1. Nonlinear Regression

In regression analysis, the models in the form

$$Y_i = \beta_0 + \sum_{k=1}^{p-1} \beta_k X_{ik} + \varepsilon_i \quad \text{i=1,2,...,n} \qquad (1)$$

are linear in the parameters. Here the variables X_k (k=1,2,...,p-1) are explanatory variables, **Y** the explained variable, and ε_i's (i=1,2,...,n) are the stochastic variables which are called "disturbances" and generally are assumed to distribute normally with zero means and a common variance σ^2. β_k's are the parameters (k=1,2,..,p-1) whose estimates will be based on the sample information at hand.

Linear regression models including more complex models such as polynomial models, or transformably linear models can be put in the above form by appropriate transformations and hence they all can be analysed by linear regression techniques. Some models although not linear in the parameters can be transformed so that the parameters appear in the linear form. Some models which are not linear in the parameters and which cannot be linearized by suitable transformations are analysed by nonlinear regression models.

Least Squares Estimation in Nonlinear Regression

A typical nonlinear regression model is in the form

$$Y_i = f(X_i, \boldsymbol{\gamma}) + \varepsilon_i \quad ; \text{i=1, 2,..., n} \tag{2}$$

Here $\boldsymbol{\gamma}$ is generally a vector of parameters whose estimates are to be obtained by the least squares' techniques. The sum of squares function to be minimized is

$$Q = \sum_{i=1}^{n} \left(Y_i - f(X_i, \boldsymbol{\gamma})\right)^2 \tag{3}$$

and this can be done by solving the following partial differential equations

$$\frac{\partial Q}{\partial \gamma_k} = 0 \quad \text{k=0,1,...} \tag{4}$$

Gauss-Newton Method

The Gauss-Newton Method (GNM) uses a Taylor series expansion to approximate the nonlinear regression model with linear terms. The GNM begins with starting values (or initial guesses) of the parameters. Once the starting values for the parameters have been obtained, the mean responses $f(X_i, \boldsymbol{\gamma})$ are approximated by the Taylor series expansion around the starting values of the parameters:

$$f(X_i;\gamma) \cong f(X_i;g^{(0)}) + \sum_{k=0}^{p-1}\left[\frac{\partial f(X_i;\gamma)}{\partial \gamma_k}\right]_{\gamma=g^{(0)}} \left(\gamma_k - g_k^{(0)}\right) \tag{5}$$

To simplify the notation let,

$$f_i^{(0)} = f(X_i; g^{(0)}) \tag{6}$$

$$\beta_k^{(0)} = \left(\gamma_k - g_k^{(0)}\right); \tag{7}$$

$$D_{ik}^{(0)} = \left[\frac{\partial f(X_i;\gamma)}{\partial \gamma_k}\right]_{\gamma=g^{(0)}} \tag{8}$$

The Taylor approximation for the mean response for the ith case then becomes

$$f(X_i;\gamma) \cong f_i^{(0)} + \sum_{k=0}^{p-1} D_{ik}^{(0)} \beta_k^{(0)} \tag{9}$$

Then $Y_i = f(X_i;\gamma) + \varepsilon_i$ (10)

$$Y_i \cong f_i^{(0)} + \sum_{k=0}^{p-1} D_{ik}^{(0)} \beta_k^{(0)} + \varepsilon_i \quad \text{and let} \tag{11}$$

$Y_i^{(0)} = Y_i - f_i^{(0)}$ values are the residuals after the first iteration . By using matrix notation

$$\mathbf{Y^{(0)} = D^{(0)}\beta^{(0)} + \varepsilon} \tag{12}$$

Note that the model given above is in the form of a general linear regression model. **D** matrix of partial derivatives plays the same role as the **X** matrix in the linear model. The vector of

parameters $\boldsymbol{\beta}^{(0)}$ are estimated therefore by the ordinary techniques used in linear regression model:

$$b^{(0)} = \left(D^{(0)'}D^{(0)}\right)^{-1} D^{(0)'}Y^{(0)} \tag{13}$$

where $\mathbf{b}^{(0)}$ is the vector of least squares estimated regression coefficients.

By using these estimates one can obtain the revised regression coefficients $g_k^{(1)}$ by

$$g_k^{(1)} = g_k^{(0)} + b_k^{(0)} \tag{14}$$

Inferences About Nonlinear Regression Parameters

The inference procedures of parameters in nonlinear regression models are similar to those of linear models. An error term variance is required to make such inferences. The estimate of the common variance for the disturbances is the same as that for the linear regression:

$$\text{Mean Square Error (MSE)} = \sum_{i=1}^{n} e_i^2 / n - p \tag{15}$$

For nonlinear regression model, MSE is not an unbiased estimator for the common variance of the disturbances σ^2, but the bias is negligible especially when n is large. When the error terms (disturbances) ε_i are independently and normally distributed with $\mu = 0$ and σ^2, for large sample sizes, the sampling distribution is approximately normal with

$$\mathbf{E}\{\mathbf{g}\} = \boldsymbol{\gamma} \tag{16}$$

An estimate of the approximate variance-covariance matrix of the estimated regression coefficients is $\mathbf{S}^2\{\mathbf{g}\} = MSE(D'D)^{-1}$ (17)

Thus, most of the classical hypothesis testing and confidence interval estimation procedures might be based on the entries of the variance-covariance matrix mentioned above.

2. Tests For Nonlinearity

F Test For Lack of Fit

The lack of fit test assumes that the observations Y for given X are independent and normally distributed, and the distributions of Y have the same variance σ^2.The lack of fit test requires repeated observations at one or more X levels. as $X_1, X_2, ..., X_c$. And we shall denote the number of replicates for the j.th level of X as n_j. The observed value of the response variable for the i.th replicate for the j.the level of X by Y_{ij}, $i = 1,...,n_j$, $j = 1,...,c$. The mean of the Y observations at the level $X = X_j$ shall be denoted by $\bar{Y}_j$. The general linear test approach begins with the specification of the full model used for the lack of fit test. This full model is:

$$Y_{ij} = \mu_j + \varepsilon_{ij} \quad (18)$$

where μ_j are parameters j=1,…,c ; ε_{ij} are independent and $N(0,\sigma^2)$

Since the error terms have expectation zero, it follows that $E\{Y_{ij}\} = \mu_j$ (19)

It can be shown that the least squares estimators for the parameters μ_j are sample means: $\hat{\mu}_j = \bar{Y}_j$. Thus the estimated expected value for observation Y_{ij} is $\bar{Y}_j$, and therefore the error sum of squares for the full model is $SSE(F) = \sum_j \sum_i (Y_{ij} - \bar{Y}_j)^2 = SSPE$ (20)

In the lack of fit test the full model error sum of squares is called the pure error sum of squares and denoted by SSPE. The degrees of freedom associated with SSPE is

$$df_F = \sum_j (n_j - 1) = \sum_j n_j - c = n - c$$

For testing the appropriateness of a linear regression function, the alternatives are:

$$H_0 : E(Y) = \beta_0 + \beta_1 X_1 + \ldots + \beta_{p-1} X_{p-1}$$

$$H_a : E(Y) \neq \beta_0 + \beta_1 X_1 + \ldots + \beta_{p-1} X_{p-1}$$

Thus H_0 postulates that μ_j in the full model is linearly related to:

$$\mu_j = \beta_0 + \beta_1 X_1 + \ldots + \beta_{p-1} X_{p-1} \quad (21)$$

The estimated expected value for observation Y_{ij} for this reduced model is:

$$\hat{Y}_{ij} = b_0 + b_1 X_1 + \ldots + b_{p-1} X_{p-1} \quad (22)$$

Hence the error sum of squares for the reduced model is:

$$SSE(R) = \sum \sum (Y_{ij} - \hat{Y}_{ij})^2 = SSE \quad (23)$$

and the degrees of freedom associated with SSE(R) are : $df_R = n - p$.

The general linear test statistic is $$F^* = \frac{SSLF}{(n-p)-(n-c)} \div \frac{SSPE}{n-c} \quad (24)$$

SSLF the lack of fit sum of squares equals the difference SSE-SSPE.

If $F^* \leq F(1-\alpha; c-p, n-c)$, conclude H_0 that means linear regression function is applicable for this relation(Neter,J.at all,1996).

SSTO measures the variation in the observation Y_i ,,when no account of the predictor variable X is taken: $SSTO = \sum (Y_i - \bar{Y})^2$ .Similarly, SSE measures the variation in the Y_i when a regression model utilizing the predictor variable X is employed. A measure of the effect of X in reducing uncertainty in predicting Y is to express the reduction in variation (SSTO-SSE=SSR) as a proportion of the total variation. This measure is r^2 and called correlation coefficient:

$$r^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO} \tag{25}$$

Since $0 \le SSE \le SSTO$, it follows that $0 \le r^2 \le 1$.

The definition of SSLF (SSE-SSPE) indicates that we have decomposed the error sum of squares SSE into two components: SSPE and SSLF.

$$\underbrace{Y_{ij} - \hat{Y}_{ij}}_{\text{Error deviation}} = \underbrace{Y_{ij} - \bar{Y}_j}_{\text{Pure error deviation}} + \underbrace{\bar{Y}_j - \hat{Y}_{ij}}_{\text{Lack of fit deviation}}$$

when both sides of this equality are squared and summed over all observations we obtain,

$$\sum\sum (Y_{ij} - \hat{Y}_{ij})^2 = \sum\sum (Y_{ij} - \bar{Y}_j)^2 + \sum\sum (\bar{Y}_j - \hat{Y}_{ij})^2 \tag{26}$$

we can express SSLF as $SSLF = \sum_j n_j (\bar{Y}_j - \hat{Y}_j)^2$ (27)

The correlation coefficient can be rewritten as (Keeping,E.S.,1995)

$$r^2 = 1 - \frac{\sum_j \sum_i (Y_{ij} - \hat{Y}_j)^2}{\sum_j \sum_i (Y_{ij} - \bar{Y})^2} \quad i = 1,\ldots,n_j \;\; j = 1,\ldots,c \tag{28}$$

The correlation ratio η_{yx}^2 is defined by,

$$\eta_{yx}^2 = \frac{\sum_j n_j (\bar{Y}_j - \bar{Y})^2}{\sum_j \sum_i (Y_{ij} - \bar{Y})^2} = 1 - \frac{SSPE}{SSTO} \tag{29}$$

From this formula it is clear that $\eta_{yx}{}^2 \le 1$. Since $SSPE \le SSE$; the inequality $1 \ge \eta_{yx}{}^2 \ge r^2$ is valid. When the regression of Y on X is linear $\eta_{yx}{}^2 = r^2$ .A non-zero value of $\eta_{yx}{}^2 - r^2$ is thus associated with a departure of the regression function from linearity. The difference may be used to estimate the degree of departure from linearity:

$$\eta_{yx}^2 - r^2 = \frac{\sum_j n_j (\bar{Y}_j - \bar{Y})^2 - \sum_j n_j (\hat{Y}_j - \bar{Y})^2}{\sum_j \sum_i (Y_{ij} - \bar{Y})^2} \tag{30}$$

By denoting sum of squared deviations of Y on overall mean with $\sum y^2$,the following formula can be derived: $(\eta_{yx}^2 - r^2)\sum y^2 = \sum_j n_j \left[(\bar{Y}_j - \bar{Y})^2 - (\hat{Y}_j - \bar{Y})^2 \right]$ (31)

The test ratio is $(\eta_{yx}^2 - r^2)/(1 - \eta_{yx}^2)$.

If the values of y within each column are normally distributed with a variance σ^2 common to all the columns, then $\sum_i (Y_{ij} - \bar{Y}_j)^2$ divided by σ^2 for any column is distributed as χ^2 with $\sum_{j=1}^{c} n_j - 1 = N - c$ degrees of freedom. $(\eta_{yx}^2 - r^2)\sum y^2$ divided by σ^2 is distributed as χ^2 with c-2 degrees of freedom.

The ratio $F = \frac{N-c}{c-2} \cdot \frac{\eta_{yx}^2 - r^2}{1-\eta_{yx}^2}$ (32)

is distributed as Snedecor's F with c-2 and N-c degrees of freedom. A significant value of F indicates a significant departure from linearity. The test is one tailed (Keeping,E.S.,1995).

We can express this test mentioned above by using tabulation:

Source of Variation	df	Sum of Squares	Mean Squares
Explained variation by linear regression	1	$\sum_j n_j (\hat{Y}_j - \bar{Y})^2$	$r^2 \sum y^2$
Explained variation by nonlinear regression	c-2	$\sum_j n_j (\bar{Y}_j - \hat{Y}_j)^2$ (SSLF)	$(\eta_{yx}^2 - r^2)\sum y^2 / c-2$
Unexplained variation	N-c	$\sum_j \sum_i (Y_{ij} - \bar{Y}_j)^2$ (SSPE)	$(1-\eta_{yx}^2)\sum y^2 / N-c$
Total	N-1	$\sum_j \sum_i (Y_{ij} - \bar{Y})^2$ (SSTO)	$\sum y^2 / n-1$

Table.1. The Analysis of Variance

The F value is the ratio of second mean square about third mean square. If the value of F is large significantly we can say that the regression function is not linear(Weatherburn,C.E.,1946).

Entropy And Mutual Information

The concept of statistical entropy is defined as

$H(X) = -\sum_x p(x) \log p(x)$ (if **X** is discrete) (33)

$H(X) = -\int_{-\infty}^{\infty} \log f(x) f(x) dx$ (if **X** is continuous) (34)

Relative entropy is a measure of the distance between two distributions. The relative entropy $D(p/q)$ is a measure of the inefficiency of assuming that the distribution is q when the true distribution is p (Cover,T.M.,Joy,A.T.,1991).

$D(p/q) = \sum p(x) \log \frac{p(x)}{q(x)}$ (35)

Mutual Information is a measure of independence between random variables. It can be expressed in terms of entropy (Joe,H.,1989). Mutual information is nonnegative and becomes zero if and only if the random variables are independent. Mutual information measures the

general dependence while the correlation coefficient measures the general dependence and mutual information is a better quantity then the correlation function to measure the dependence (Li,Wention,1990). If X and Y are random variables with joint distribution p(x,y) and mariginal distributions p(x) and p(y) then the mutual information $M(X;Y)$ is the relative entropy between the joint distribution and the product of two marginal distributions.

M(X,Y) is the information that Y conveys about X is the reduction in uncertainty about X due to the existence of Y.

$$M(X,Y) = D(p(x,y)/p(x)p(y)) = H(X) - H(X/Y) \quad [1] \qquad (36)$$

The relationship between entropy and mutual information can be shown as:

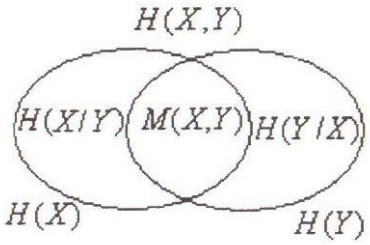


Fig.1. Entropy And Mutual Information

It can be seen that, when X and Y are independent $M(X,Y) = 0$. If Y and X are independent variables then Y can tell nothing about X (Cover,T.M.,Joy,A.T.,1991).

Testing Linearity By Using Mutual Information

This method is used for detecting and evaluating intrinsic nonlinearity and suggested by Tanaka,N., Okamoto,H., Naito,M. The mutual information measures statistical dependence but does not discriminate between linear and nonlinear dependence. This method is based on comparing the mutual information between the original data series with that between new ones obtained by removing linear dependence from the original ones (new ones are equivalent to residuals from linear regression). The point is to remove linear dependence from the original data and then to compute mutual information. For continuous variables $M(x,y)$ is given by,

$$M(x,y) = \int\int p(x,y)\log_2 \frac{p(x,y)}{p(x)p(y)} dxdy \qquad (37)$$

p(x,y) denotes the joint probability density related to x and y. p(x) denotes the probability density of x and p(y) denotes the probability density of y.

After removing linear dependence from the original data set a new data series $\{\xi_i\}$ (i=m, m+1,...,N) generated by $\xi_i = y_i - b_0 - \sum_{j=0}^{m-1} a_j x_{i-j}$ is considered to include no linear dependence on

[1] H(X/Y) denotes the conditional entropy of X given Y.

x. The parameters are estimated by maximum likelihood method. And m is the order of the model. a_j is given by the correlation coefficient. The mutual information $M(x,\xi)$ can be computed as well as $M(x,y)$. And $M(x,\xi)$ can be regarded as a measure of intrinsic nonlinearity present in the mutual dependence between x and y. To evaluate the degree of nonlinearity included in the dependence the two mutual information measures can be compared. When $M(x,\xi) \square M(x,y)$ it can be interpreted that the dependence is based on nonlinearity. If the value of $M(x,\xi)$ is nearly zero this indicates that nonlinearity in the dependence between x and y is negligible. For comparing different types of dependence measures (correlation coefficient, correlation ratio, mutual information) it is necessary to normalize the dependence measures. Normalized mutual information between x and y can be given by the following formula.

$$\rho(x,y)=\frac{M(x,y)}{H(x,y)} \qquad (38)$$

Similarly, $\rho(x,\xi)$ can be defined. $\rho(x,y)=0$ then means that the two variables are statistically independent, and $\rho(x,y)=1$ means that x and y are statistically identical. By using this measure we can compare the results among different types of dependence. $\rho(x,y)$ and $\rho(x,\xi)$ can be compared to decide the shape of the relation between x and y. If the two measures are close to each other the dependence is nonlinear and $\rho(x,y)>\rho(x,\xi)$ condition suggests the dependence is linear (Tanaka,N.,at all,2000).

3. Application

In this paper first we use the F test for deciding whether a linear regression function is a good fit for the data(Yüce,M.,2004). Under the information about F test mentioned above we arranged the table,

Source of Variation	Df	Sum of Squares	Mean Squares
Explained variation by linear regression	1	$\sum_j n_j(\hat{Y}_j-\bar{Y})^2$ =84717.97	$r^2\sum y^2$ =84717,97
Explained variation by nonlinear regression	c-2 22-2=20	$\sum_j n_j(\bar{Y}_j-\hat{Y}_j)^2$ =32648.8 (SSLF)	$\frac{(\eta_{yx}^2-r^2)\sum y^2}{c-2}$ =1632,4
Unexplained variation	N-c=192	$\sum_j\sum_i(Y_{ij}-\bar{Y}_j)^2$ =2067.892(SSPE)	$\frac{(1-\eta_{yx}^2)\sum y^2}{N-c}$ =10.7704
Total	N-1 214-1=213	$\sum_j\sum_i(Y_{ij}-\bar{Y})^2$ =119434.7	=560,7263

For testing the appropriateness of a simple linear regression function, the null and alternative hypothesis are:

$H_0 : E(Y) = \beta_0 + \beta_1 X$

$H_a : E(Y) \neq \beta_0 + \beta_1 X$

The general linear test statistic for simple regression model is:

$$F^* = \frac{SSLF}{(n-2)-(n-c)} \div \frac{SSPE}{n-c} = \frac{1632.44}{10.77043} = 151.5668$$

$F^* > F(0,95;20,192)$ so we can conclude H_a with 0,0000 significance value. This means linear regression function is not a good fit for the data.

The dependence measures are, $r^2 = 0,709325$ and $\eta_{yx}^2 = 0,982686$. A non-zero value of $\eta_{yx}^2 - r^2$ is associated with a departure of the regression function from linearity. Calculated value of this difference is $\eta_{yx}^2 - r^2 = 0,273361$, implying that the functional relationship is not linear. To test the significance of this difference. The alternatives are

H_0 : The relationship between x and y is linear

H_1 : The relationship between x and y is curvilinear

The test statistic is $F = \frac{N-c}{c-2} \cdot \frac{\eta_{yx}^2 - r^2}{1-\eta_{yx}^2} = 151.569$ and distributed as Snedecor's F with c-2 and N-c degrees of freedom. A significant value of F indicates a significant departure from linearity. By comparing the F table value a quite significance derived from. This result is the same as the one from the F test. The results about entropy and mutual information are tabulated as below,

$M(x,y)$ =0,654	$M(x,\xi)$ =0,461	$\rho(x,y)$ =0,462	$\rho(x,\xi)$ =0,309

We see that $M(x,\xi)$ is far from zero. Hence we can't say the relationship between x and y is purely linear. Similarly. $\rho(x,y)$ and $\rho(x,\xi)$ are close to each other so the dependence is nonlinear.

4. Conclusion

One technique to check nonlinearity is based on an F test. Second, a comparison between linear correlation coefficient and correlation ratio may be helpful. The difference between these dependence measures can be tested. Besides, there is a method for detecting and evaluating the intrinsic nonlinearity, based on the mutual information measures. All these tests have been performed on the data. Similar results have been obtained.

References

1. Bates,D.M., Watts, D.G.,*Nonlinear Regression Analysis and Its Applications*, John Wiley &Sons,1988
2. Elif Öztürk, *Regresyon Diyagnostikleri ve Sapan Değerler*", Y.T.U. , Master Thesis, İstanbul,2005
3. Fazlollah M.Reza, *An Introduction to Information Theory*, Dover Publications, 1994
4. Joe, H., *Relative Entropy Measures of Multivariate Dependence*, Journal of The American Statistical Association, March 1989, Vol.84, No:405
5. Keeping,E.S., *Introduction to Statistical Inference*,Dover Publications,1995
6. Li,Wention, *Mutual Information Versus Correlation Functions*,Journal of Statistical Physics,1990, 60(5-6)
7. Mehmet Yüce, *Doğrusal Olmayan Regresyonda Değişen Varyans ve İlgili Uygulamalar*,Y.T.U , Master Thesis, İstanbul,2004.
8. Merih İpek, *Betimsel İstatistik*,Beta Yayınevi,1994
9. Neter,J.,Kutner,M.H.,Nachtsheim, .J.,Wasserman,W., *Applied Linear Regression Models*,Third Edt.Irwin,1996
10. Seber,G.A.F.,Wild,C.J., *Nonlinear Regression*, John Wiley &Sons,2003
11. Tanaka,N.,Okamoto,H.,Naito,M., *Detecting and Evaluating Intrinsic Nonlinearity Present in The Mutual Dependence Between Two Variables*, Physica D, 147, 2000
12. Thomas M.Cover, Joy A.Thomas, *Elements of Information Theory*,John Wiley &Sons,1991
13. Weatherburn,C.E., *A First Course in Mathematical Statistics*,Cambridge University Press,1946
14. Zehra Zeynep Şahinbaşoğlu, *Doğrusal Olmayan Regresyonda Bazı Eğrisellik Ölçüleri* , Y.T.U., Master Thesis, İstanbul,2005