Journal of mathematics and computer science 7 (2013), 213-220

Contents list available at JMCS

Journal of Mathematics and Computer Science

Journal Homepage: www.tjmcs.com

Using Eta (η) correlation ratio in analyzing strongly nonlinear relationship between two Variables in Practical researches

Ahmad Hedayatpanah Shaldehi

Technical- Professional University (Rasht Dr. Mo'en Girls College)

a_hedayatpanah@yahoo.com

Article history:
Received April 2013
Accepted May 2013
Available online May 2013

Abstract

This paper main aim is Eta (η) correlation ratio in analyzing strongly nonlinear relationship between two Variables in breeding area researches. In researches on a relationship that leads two –Variables correlation to determine correlation coefficient if two Variables would Continual (interval, ratio) and parametric uses Pearson torque correlation and if discrete (ordinal, nominal) or nonparametric uses Spearman Brown. And to determine regression line writes Predicting equation as linear relation and if Variables would be more than two uses multivariable regression. But in many cases relationship between Variables is nonlinear that is useful to analyze data use minimizing practice method. This method could expand in many nonlinear senses. But if relationship between two Variables [both continual] would be strongly nonlinear using minimizing is not suitable. That in this correlation as researcher uses Eta (η) correlation ratio instead of correlation coefficient obtains efficient result to specify meaningful difference. This paper refers to common methods inefficiency and defects with an example then refers to (η) advantages, efficiency and how calculating and how using it in practical researches.

Keywords: Eta, ratio, strongly nonlinear, practical researches, regression, correlation

1. Introduction

In practical research, the major object is not scientific discovery, but it is testing and analyzing the application of Knowledge and it is selected its subject in education and from daily problems of education, [5]. It increases our understanding of special problems, [12]. Most of the researchers are related to education and practical Researchers [2]. Researcher separates data to principle parts for achieving the answer of research questions and testing of hypothesis , then explain this analaysis [10], and the object is

analyzes of data in an understandable and explainable from that can study and test the relationship of researching problems[10]. One of the analysis type in quantitive researches is using of correlation coefficient [11] that is line grade of regression. This is the index that show the relationship between variables , quantively[6]. At least , 10 methods has discovered to calclulate the correlation coefficient (r) between two variables. Variables that use in correlation research is place in one of the follow categories: 1-connected numbers . 2-ranking . 3- true two-valued.4- simulated two –valued .5-categorization[9]. If both variales have been continous , we

$\frac{SP_{xy}}{\sqrt{SS_x SS_y}}$= should use Pearson moment correlation with r

$\frac{6\sum d^2}{n(n-1)(n+1)}$ r = 1 - And if both of them have been ranking , we use Spearman correlation

and also fee correlation , agreement , Kendal ,tettrachoric ,expanded two-valued and poin coefficient that are use in linear data. But nonlinear , eta is more practical[6,9]. If it has been predicted that there is a common moment correlation between quantities,ý

and observed quantity of Y are calculated and multivariable correlation coefficient R is optained from (1) :

$$R=\frac{\sum \dot{Y}\hat{Y}}{\sum \hat{Y}^2 \sum \dot{Y}^2} \quad (1)$$

Where R is the highest correlation between linear compound of square independent variables and connected variables[10]. The method that is compounded by predicted variables , is called multivariate regression[5]. Multiple regression is refered to plans that independed variables are connected in them [7] and the major equation regression is (2) $= a+b_1x_1 +\ldots+b_kx_k \hat{Y}$

The purpose of determining bs in equation(2) is finding of 6 that decrease the square sum of surpluses[10] the

2.Polynominal Fitting

If we consider the equation of straight line of showing regression as

(3) Y=a+bx

Y in this equation will be true amount with deviation [8] . that is y-a-bx=e and by this

$$\sum(y - a - b\,x)^2 = e^2$$

is minimum . For achieving this linear equation , by lagrange low:

$$\frac{de^2}{da} = -2 \sum(y - a - b\,x) = 0$$

$$\frac{de^2}{db} = -2\,x \sum(y - a - b\,x\ ) = 0$$

And by solving this equation ,the coefficionts if a and b and also the linear regression equation (3) is achieved [3]

In equation:

$f(x_k) = y_k + x_k$ (4) ,how can we find the best approximate line (3) that is gonne over the mentioned points (no always out of them) ? To answer this question :

y_k (5) $\neq f(x_k) = e_k \quad k \leq N \leq 1$

} (6) $I\ y_k - f(x_k)\ I \max\{ = f) (E_\infty$

Erroe average

(7) $I y_k - f(x_k)\ \Sigma I\ \frac{1}{N}) = f (E_1$

Average square error

(8) $^{/5}]\ ^{2}\ I y_k - f(x_k)\ \Sigma I\ \frac{1}{N}[=)(f\ E_2$

For example , for linear approximation

Y=-1/6x+8/6

)10 ، 1) (-9 ،0) (7 ، 1) (5 ، 2(

)4 ، 3) (3 ، 4) (0 ، 5) (1 ،- 6(

$E_\infty = \max\{.2,.4,.0,.4,.2.8,.6,.0\}$

325) =/6/2($\frac{1}{8}$) = f((E_1

41833= /$^{5/}\frac{1/4}{8}$)) = ((f E_2

The best fitted line is found by minimizing of the (5),(6) and (7) amounts , so , there are 3 best fitted line that can be find (4).

2-1 Method of curve fitting for nonlinear equation

The Method of curve fitting for least squsres can be extend to many nonlinear cases. For example , we start from

y_k , (x_k ,y_k) by N point of data and consider the follow exponential fitting: $\neq f(x_k) = e_k$

Y= Cexp(Ax) (9)

It is necessary to find minimum of

$E(A,C)=\sum_{k=1}^{N}[C \exp(Ax_k) - y_k]^2$ (10)

To metod of least squars.

Slight derivations E(C.A) in proproation to C and A are :

$$(11)\ \frac{\partial \mathrm{E}}{\partial \mathrm{A}}=2\sum_{k=1}^{N}[\mathrm{C}\,\mathbf{exp}(\mathrm{Ax}_k)-\mathrm{y}_k][\quad \mathbf{exp}(\mathrm{Ax}_k)\,Cx_{\,k}\ ]$$

$$\frac{\partial \mathrm{E}}{\partial \mathrm{c}}=2\sum_{k=1}^{N}[\mathrm{C}\exp(\mathrm{Ax}_k)-\mathrm{y}_k][\,\mathbf{exp}(\mathrm{Ax}_k)\ ]\quad (12)$$

Normal equatoions (11) and (12) are nonlinear in proportion of unknown C and A and we can solve the equation by using of Newton method . Another method to find the minimum of E(A,C) directy, is using of Nelder mead simplex algorithms (4).

3. Correlation Ratio η

In correlation researches , hypothesis is that the predition or relationship about the research is linear . Therefore , sometimes the relationship is not linear and this is in the condition when two variables are connected and in this case , the coefficient ratio **η** is more appropriate . One course , sometimes the linear relationship is recognized after drowing scattergram [9].

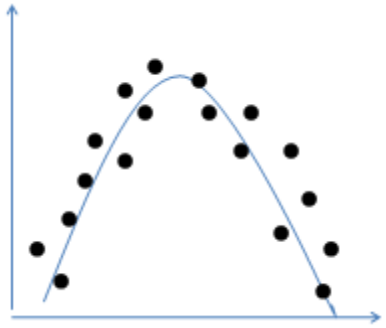

Figure (1) scattergram

Since , a curve line explaines the relationship between two variables better than a straight line ; therefore we can achieve better prediction for y numbers axis, by using of x numbers axis. For example , some reseachers assume that the relationship between anxiety and intelligence operation is a curved line relationship. If the research data of scattergram shows the nonlinear relationship between two variables, researcher should canculate the correlation ratio **η** (Eta) instead of correlation coefficient .

Advantages of correlation ratio is that since the linear relationship recognizes , then it provides the more precious index of relationship between two variables, in proportion to other correlation statistics. Other types of correlation coefficient estimate the relationship between two variables lessly when the relationship is nolinear[9].example: The average for data are 3 and 4 in two groups for learning X_1:4,5,3,2,6 , X_2:3,1,5,2,4 . Analysis of variance calculated , but the established difference is not significant in two groups efficiency .

SD=1/58)Deviation (, V=2/5)variance (

SE_m=/705)mean of standard error (SV_m=/5 ,)mean of standard variance(

The question is that the difference if 4 is more than chance differences? i.e . :

$H_1: M_{A1} > M_{A2}$ where (major hypothesis)

$M_{A1} = M_{A2}$ H_o :(zero hypothesis)

Average of differences in standard erroe $SE_{MA1} - SE_{MA2} = /997 \cong 1$ and t=1

Since the evaluated difference is not bigger than error index ,So it is not significant. Numerator and denominator of t are equal. The difference of 4-3=1 is one the clear differences that is to happen by coincidence numbers.b Therefore , there is no relationship between independent variable (methods) and dependent variable (development). If we are done the analysis of variance data by **η** method , the relationship between methods and development is as follow:

$ss_b = 2/5$ $22/5 = ss_t$ $32 / \sqrt{\frac{2/5}{22/5}} = = \eta$

Because the deta is not significant , so (F=1) , ($F=t^2$) and η are not significant on the other hand , η=132 is so near to zero. If ther is no significance between means, then η=0 and if $ss_b=ss_t$,then η =1 . This happens when all grades are the same in one group and all other grades of group are equal and nevertheless they are different from first group , and it is so far. For example, if numbers are A_1=4,4,4,4,4 and A_2=3,3,3,3,3, then $ss_b= ss_t$=2/5 and

$$1 = \sqrt{\frac{2/5}{2/5}} = \eta$$

It is dear that in this case there is no variance in groups and also it is so far . For example in follow data :x_1:6,7,5,4,8 x_1 :3,1,5,2,4

We added 2 numbers to each data of x_1 ,from previous example , s0 the variance doesn't change. But the means are 3 and 6 and the difference is no significant because F=9 . calculating of η:

Source	df	ss	ms	F	
	2/5	22/5	22/5	k-1=1	Between groups
2/5	20		n-k=8		In groups
42/5	n-1=9		total		

$(\frac{ss}{df} = ($ ms , /73 , df =degree free $= \eta = \sqrt{\frac{22/5}{42/5}}$

We should consider the main increase of η , because F is significant and also η=/73 is significant ; so , there is a strong relationship between methods and development . If we squared η .i.e. η^2 showes the same the same variance of independent and dependent variables like r^2. From scientific and practical point of

view , we must emphasize that η^2 or other indexes of relationship to calculate and submit . Just ratios report of F and if they are significant or not , are not enough . We must know that how these relationships are strong. Besides, when Ns are being great sufficiently , then the ratios of F and t are been significant statistically . Since, We should pay-attention about their effects, specially when they are weak , most of the times[10].Example: Consider two cases , strong relationship and zero relationship that are recognized whth numbers of independed variables by high , medium and low grades.

Table(1) strong relationship between training method and development

Mean	dependent variable(development)	Independent variable(training method)
9 (high)	9-9-10-8	Method A1
(medium)7	7-7-7-7	Method A2
Low)4(	3-4-4-5	Method A3

Example of table(1) competes with expection of chance result .If there is no relationship between methods and development , development means don't change like the same of methods. That means are equal approximately.

To show this point , we wrote 12^{th} development greads of table (1) on separate papers and mixed them in a container and selected 4 of them each time . The first four selection allocated to A_1 , second ones to A_2 and third selections to A_3 ; The results are shown in table(2).

And low grades.

Table(2) The zero relationship between training method and development

Mean	dependent variable(development)	Independent variable(training method)
7/25 (high)	10-8-4-7	Method A1
(medium)5/25	9-4-5-3	Method A2
Low)7/5(	9-7-7-7	Method A3

The means are different in this case , but no so much . Certainly , the relationship between methods and development grades (and mean) are not clear as previous.But we must be sure. Analysis of variance is done is two sets of data. F ratio of data in table (1) [strong relationship] is 57/59 and significant, whereas the F ratio of data in table (2) [zero or weak relationship] is 1/29 and not significant .Statistical tests confirmed our facial understanding (disply). Now we known , there is a relationship between methods and development , in table (1) , but it isn't in table (2) . However , the issue is showing a significant relationship between tests like F test and correlation method . We can do this in different ways like drawing or statistical . Now , consider the issue from statistical view . For these types of data , we can calculate the correlation coefficient . If somebody has done variance analysis , the can achieve a simple coefficient (but not satisfactory) by using follow formula:

$= \sqrt{\frac{ss_b}{ss_t}}$ where ss_b and ss_t are the sum of squares between groups total , respectively. η

To calculate this coefficient , we can choose the sum of these squares from variance analysis table . Generaly , η is called correlation ratio and it is a total or index relationship of coefficient and often , uses fo data that are nonlinear . Here we just interested in application of it in variance analysis and its ability in explaining of relationship between dependent method of Eta ration , specially when there one several predicted varialdes in research [9].

4. Conclution

In relationship researches , the connected and discrete data use correlation , in term of scale that is first priority of each researcher in diagnosing of correlation coefficient . Generally , the linear and regression subject use predicated relationship , linear and nonlinear . In this essay , consideration is on nonlinear relationships .When nonlinear relationships are strong , we use Eta ration (η) instead of correlation coefficient and achieve better answer . The major point in mentioned issue is that the similarity of these indexes with other indexes , relationship or correlation is figured out the most important , the similarity in principle and structure of variance analysis and correlation methods are cleared.

References:

[1] E. Biabangard, research method and ststistic , sanjesh and sanjesh takmili publication, **(2004).**

[2] B. Jan, research of method in educayional and behavioral science, translators: Pasha sharifi and Narges Talegani – Roshd publication**,(1998).**

[3] P. Tagdari and A N manouchehr, introductory of statistical in Eta publication**,(1990).**

[4] J. H Maew, Numeral calculation methods, Translator : Faeze Totonian-Khorasan publication**,(2000).**

[5] A. Delavar, Research method in psychology and educational sciences Payam Noor University publication**,(2005).**

[6] A. Delavar, application and contingencies-Roshd publication**,(2011).**

[7]. David R-Boty phibs, plan of tests and statistical methods. Translators; Talebi, Alamatsaz and Mousavi- Esfahan university publication**,(2001).**

[8] H. Ranjbaran, statistic and contingency University book publication**,(2008).**

[9] G. Borg. G, quantitative and qualitative methods of research in educational and psychological sciences. First and second edition – Translators: Ahmad Reza Nasr ana others shahid Beheshti University and (samt) publication,**(2006).**

[10] K,Serarching principles in behaviorial sciences. First and second volumes. Translators: Pasha Sharifi, Nagafi Zand , Ave Noor publication**,(2004).**

[11] A. hedayatpanah,How research? Darya Danesh publication**,(2009).**

[12] H. Terry-E. B,Reg , Translattors: Mohammad Arabi,Davoud Ezadi-Caltural researches office**,(2000).**