DISPLAY THE DISEASE PLACE FOR NON LINEAR MODEL WITH " ANOVA" TECHNIQUE

K. EL KOURD(1), A. AZZIZI (2), F.BOUGOURZI(3), S.HAMMOUM(4)

(1)Department of physical ,preparatory school of sience & technicalof Algiers-Algeriabc, (2,3)Electronic of engineering of Med khider of Biskra, (4)Clinical of radiology-MRI service-view Kouba-Algiers
kaouther_youcef@yahoo.fr

ABSTRACT

In this paper, we transform a nonlinear model to a linear one by using numerical analysis with ''Runger-Kutta4(RK4)''. Which is a mathematical technique to approximate solution of ordinary differential equations; this method is most popular where the step size H is working to increase the lighting of the image compared with the original picture. The new data (normal & pathological images) obtained from this method is used in the statistical study of simple regression and "ANOVA" technique to detect the tumor of MRI images. After that, we study the linear regression and "ANOVA" technique by using ANOVA statistical test (equation of ANOVA: fcal) and compare it with ANOVA table(ftab) for probability p-value =0.01 (here for area 200x200, ftab=1) and see all pixels inferior to''1'' that means the hypothesis ho is accepted. All these detail is to extract the place of the lesion on MRI ,(which contain matrix data of normal image and pathological ones), the extract the accepted ho pixels directly on the pathological image. The simulation program applied here is Matlab.

KEYWORDS: Runge kutta ,linear regression, Anova.

1   INTRODUCTION

Analysis of variance became widely known after being included in Fisher's 1925 book Statistical Methods for Research Workers[1].

In statistics, analysis of variance (ANOVA) is a collection of statistical models, and their associated procedures, in which the observed variance in a particular variable is partitioned into components attributable to different sources of variation [1].

Anova ,one-way ,is for comparing the means of two or more columns of data in the m-by-n matrix X, where each column represents an independent sample containing m mutually independent observations. The function returns the p-value for the null hypothesis that all samples in X are drawn from the same population (or from different populations with the same mean). If the p-value is near zero, this casts doubt on the null hypothesis (Ho) and suggests that at least one sample mean is significantly different than the other sample means. The choice of a critical p-value to determine whether the result is judged "statistically significant" is left to the researcher. It is common to declare a result significant if the p-value is less than 0.05 or 0.01.

The ANOVA function displays two figures. The first figure is the standard ANOVA table, which divides the variability of the data in X into two parts [1][2].

- Variability due to the differences among the column means (variability between groups).

- Variability due to the differences between the data in each column and the column mean (variability within groups).

To have an excellent result, we apply numerical analysis with Runge Kutta 4th order method then the ANOVA technique. This method, RK4, is most popular and a good choice for common purposes because it is quite accurate, stable, and easy to program. RK4 simulates the accuracy of the Taylor series of order 4 [3].

This paper is organized as follows. In section II, we review the Runge Kutta4, fitted Regression and Anova expressions. In section III, we will introduce our results. And finally we conclude in section IV.

As problematic what's the meaning of Runge kutta? What it the result of it? What's the goal regression and how do we use to Anova technique? What's the relation between theoretical result of Anova(fcal) & Anova table(ftab)?.

All these questions will be more explain in the following titles.

2 BACKGROUND

2.1 Runge kutta

In numerical analysis, the Runge–Kutta(RK) methods (German pronunciation: [k ta]) are an important family of implicit and explicit iterative methods for the approximation of solutions of ordinary differential equations. These techniques were developed around 1900 by the German mathematicians C. Runge and M. W. Kutta[4][5].

2.1.1 Description

The position of the spaceship can be described as $(x(t), y(t))$ for any time point t. Similarly, the velocity is (x (t),y (t)) so you can see, the spaceship moves in two-dimensional space. Its acceleration is described by the following system of differential equations[2]: $x'' = f(x, y, x', y')$ & $y'' = g(x, y, x', y')$, given the initial values : $x_0, y_0, u_0, v_0, u = x', w = y'$.

In the RK4 method, you calculate four intermediate approximations, k_1, k_2, k_3, k_4; the final approximation will be given by a weighted average of these intermediate approximations:

$$u_{i+1} = u_i + (1/6)(k_1 + 2k_2 + 2k_3 + k_4) \qquad (1)$$

After translate the non linear model (exponential equation) to linear one ,then study the linear regression and the fitted regression line for calculate at last Anova test as the following titles .

2.2 Regression

Description about the relationship between two variables where one is dependent and the other is independent[6][7].

2.3 Fitted Regression Line

The true regression line corresponding to eq(2) is usually never known. However, the regression line can be estimated by estimating the coefficients S_1 and S_0 for an observed data set[4][5].

$$Y = X * S + V \qquad (2)$$

The actual values of y, (which are observed as yield from the chemical process from time to time and are random in nature), are assumed to be the sum of the mean value, E(Y), and a random error term, e : The actual values of y, (which are observed as yield from the chemical process from time to time and are random in nature), are assumed to be the sum of the mean value, E(Y), and a random error term,

e:eq(3).

$$Y = E(Y) + V$$
$$= S_0 + S_1 * x + V \qquad (3)$$

The estimates, $\hat{S}_1$ and $\hat{S}_0$, are calculated using least squares. The estimated regression line, obtained using the values of, $\hat{S}_1$ and $\hat{S}_0$, is called the fitted line. The least square estimates, $\hat{S}_1$ and $\hat{S}_0$, are obtained using the following equation:eq(4).

$$\hat{S}_1 = \frac{\sum_{i=1}^{n} y_i x_i - \frac{(\sum_{i=1}^{n} y_i)(\sum_{i=1}^{n} x_i)}{n}}{\sum_{i=1}^{n} (x_i - \overline{x})^2} \qquad (4)$$

$$\hat{S}_0 = \overline{y} - \hat{S}_1 \overline{x} \qquad (5)$$

Where $\overline{y}$ is the mean of all the observed values and $\overline{x}$ is the mean of all values of the predictor variable at which the observations were taken. $\overline{y}$ is calculated using

$$\overline{y} = (1/n)\sum_{i=1}^{n} y_i$$ and $\overline{x}$ is calculated using

$$\overline{x} = (1/n)\sum_{i=1}^{n} x_i .$$

Once $\hat{S}_1$ and $\hat{S}_0$ are known, the fitted regression line can be written as:

$$\hat{y} = \hat{S}_0 + \hat{S}_1 x \qquad (6)$$

where $\hat{y}$ is the *fitted* or *estimated* value based on the fitted regression model. It is an estimate of the mean value, E(Y). The fitted value, $\hat{y}_i$, for a given value of the predictor variable, xi, may be different from the corresponding observed value, y_i. The difference between the two values is called the *residual*, ei [6][7].

$$e_i = y_i - \hat{y}_i \qquad (7)$$

To calculate the Statistic Fo , it must pass by the six titles [8][9][10][11][12][13][14][15] [16][17].

2.3.1 Total Sum of Squares (SST)

On simple linear regression that the total sum of squares, SST , is obtained using the following equation:

$$SS_T = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

$$= \sum_{i=1}^{n} y_i{}^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n}$$

(8)

The total sum of squares in matrix notation is:

$$SS_T = y'y - (\frac{1}{n})y'J_y$$

$$= y'\left[I - (\frac{1}{n})J\right]y$$

(9)

Where y is the vector of observed values, I is the identity matrix of order n; & J represents an n x n square matrix of ones.

2.3.2 Model Sum of Squares (SSR)

Similarly, the model sum of squares or the regression sum of squares, SSR, can be obtained in matrix notation as:

$$SS_R = \sum_{i=1}^{n} \hat{y}_i{}^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n}$$

$$= \hat{y}'\hat{y} - (\frac{1}{n})y'J_y$$

$$= y'\left[H - (\frac{1}{n})J\right]y$$

(10)

Where H is the hat matrix and is calculated using

$$H = X(X'X)^{-1}X'$$

(11)

2.3.3 Error Sum of Squares

The error sum of squares or the residual sum of squares, SSE, is obtained in the matrix notation from the vector of residuals, e, as:

$$SS_E = SS_T - SS_R$$

(12)

$$SS_E = e'e$$
$$= (y - \hat{y})'(y - \hat{y})$$
$$= y'(I - H)y$$

(13)

2.3.4 Mean Squares (MS_T)

Mean squares are obtained by dividing the sum of squares

with their associated degrees of freedom. The number of degrees of freedom associated with the total sum of squares, SST, is (n-1) since there are n observations in all, but one degree of freedom is lost in the calculation of the sample mean $\bar{y}$. The total mean square is:

$$MS_T = \frac{SS_T}{n-1}$$

(14)

2.3.5 Regression mean square (MSR)

The number of degrees of freedom associated with the regression sum of squares,SSR, is k. There are (k+1) degrees of freedom associated with a regression model with (k+1) coefficients, $\beta_o, \beta_1 \beta_2 \dots \beta_k$ However, one degree of freedom is lost because the deviations, $(\tilde{y}_i - \bar{y})$, are subjected to the constraints that they must sum to zero $\sum_{i=1}^{n}(\tilde{y}_i - \bar{y})^2$. The regression mean square is:

$$MS_R = \frac{SS_R}{k}$$

(15)

The number of degrees of freedom associated with the error sum of squares is: n-(k+1), as there are n observations in all, but (k+1) degrees of freedom are lost in obtaining the estimates of $\beta_o, \beta_1 \beta_2 \dots \beta_k$ to calculate the predicted values , $\tilde{y}_i$. The error mean square is[5][6][7][8][9].

$$MS_E = \frac{SS_E}{n-(k+1)}$$

(16)

The error mean square, MSE, is an estimate of the variance, σ^2 of the random error terms, ei.

2.3.6 Mean square error(MS_E)

Is estimate variance ($\tilde{\sigma}^2$) of random error ei .see (eq 15).

$$\dagger^2 = MS_E$$

(17)

2.3.7 Calculation of the Statistic F_o

Once the mean squares MSR and MSE are known, the statistic to test the significance of regression can be calculated as follows:[4]

$$F_0 = \frac{MS_R}{MS_E}$$

(18)

3   EXPRIMENTAL RESULTS

3.1   Algorithm

- Read the images.
- Choose a sample of image.
- Choose model non linear & applied it in RK4.
- Applied analysis mathematic with Runge kutta4 and display the new image for normal & for pathological image(y and x).
- Display the image for each matrix by two ways :before & after used RK4.
- Calculate the error between y & y estimate then plot it for one vector.
- Application the analysis of variance 'ANOVA' .
- Comparison the result of Anova with use Runge kutta & the technical of Anova without use RK4.

3.2   Analysis of data (the protocol radiologic)

Our protocol is for a patient aged 55 years; He made examination by MRI scan with injection of contrast medium. The machine used : type "SIEMENS", and with the field B = 1.5 Tesla. The sequences performed T1 and T2.

The result MRI scan of a tumor appear in the middle of the field on the right side and the third ventricle is to evoke the tumor.

In Figure1, we present normal image in left side & pathological ones in the right side . in the bottom of them , we have their both   samples with length 180 .
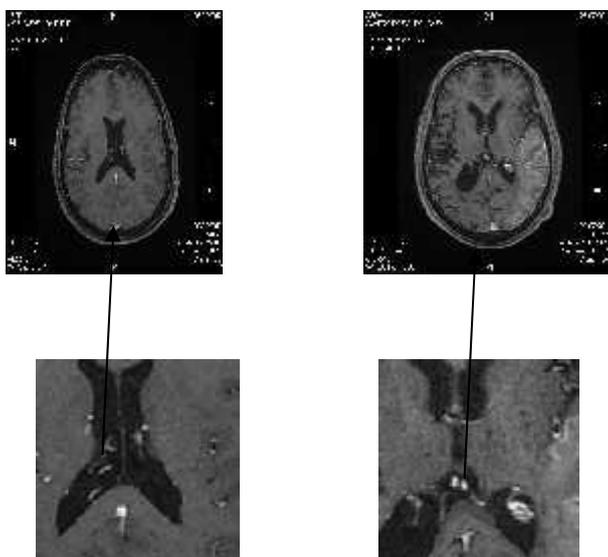
Figure 1: left .Normal image; right : Pathological image.

3.3   Analysis of RK4

3.3.1   Choose non linear model

Figure(2) presented: non linear model , the step size and the surface of image

Figure 2:   Example of model non linear, H & length (m).

3.3.2   Display the first vector of each image by two ways :before & after used RK4

Figure3, presented the form of matrix (y) before used RK4(fig 3.a) and after used RK4(fig 3.b),where we can see a few variation value for the normal image(y) . The same thing for the both matrixes in   pathological image (x).see (figure 3.c)& (figure3.d).
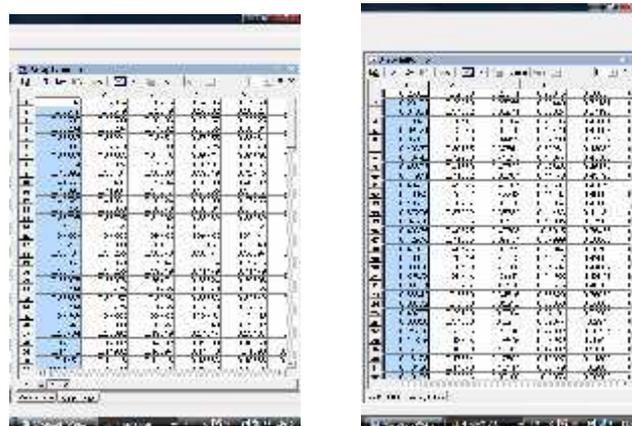
a. Vector   Y (1,:)
    before RK4

b. Vector   Y (1,:)
    after RK4
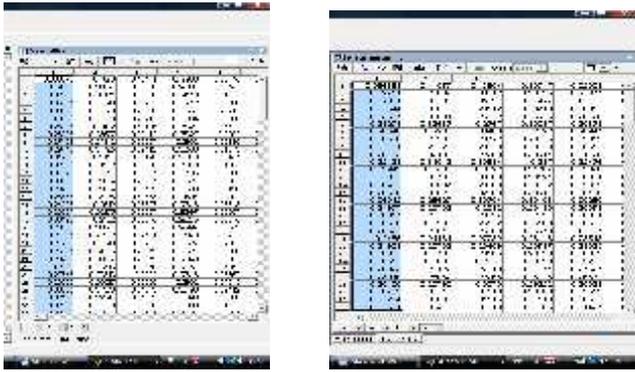
c. Vector   x (1,:)

before RK4

d. Vector   x (1,:)

after RK4

Figure 3: Choosing the first vector with the blue color of matrixes (normal& pathological image).

(a)First vector y(1,:) of matrix (y) before RK4

(b) First vector y(1,:) of matrix (y) after used RK4(b)

(c) first vector x(1,:) of matrix (x)before used RK4

(d) first vector x(1,:) of matrix (x)after used RK4

The first vector of the matrixes which displays by blue color (a,b,c,d) of figure3, presented in the following curves of y(1,:)&x(1,:) before & after used RK4. See figure4.

Where, $Y(1,:)$ & $x(1,:)$, means first vector for the length m=180.

a

b
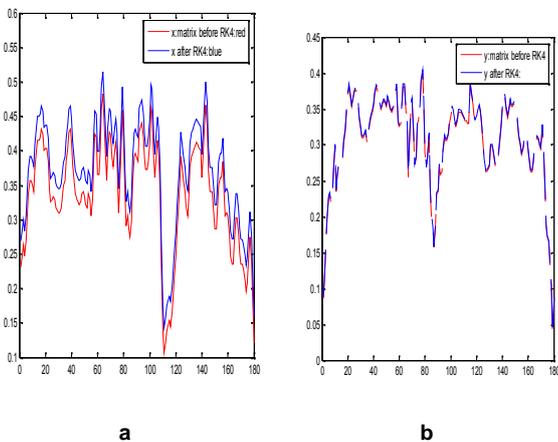
Figure 4: Plotting the first vectors of fig(2).

(a)curve of pathological image(x) before used RK4(red color)& after RK4(blue color).

(b)curve of normal image (y)before used RK4(red color)& after used RK4(blue color).

Figure (4.a) shows us two curves of the image curve x where blue is located down the red. The second figure (4.b) shows the color blue beneath red varying small. all those curves red Accomplished with RK4.

Note:

The presence of the red curve in pathological image explains why the image appears after RK4 more Lighting in front of normal image.

3.4   Display the image for each matrix by two ways :before & after used RK4

In the precedent title we have seen formes of curves of victors(x &y)and now we will show image after using RK4 which became more Lighting for the step size (H=0.04)as presented in fig(5).
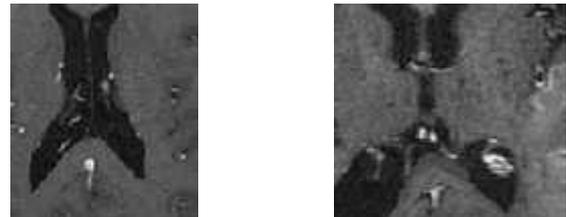
a .Normal image

before used RK4

b. Pathological image

before used RK4
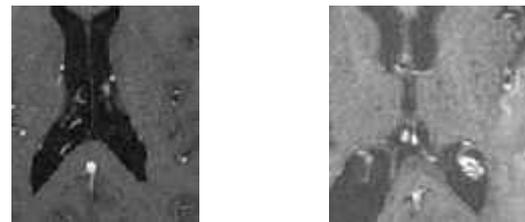
c .Normal image

after used RK4

d. Pathological image

after used RK4

Figure 5: The grayscale luminosity before used RK4 (a,b)&after used RK4 For step size H=0.04(c,d).

Note :

Whenever we have increased H we get more Lighting picture and result to be obtained   be bad   as in fig(6.a) ,where   H = 1, haven't detect   any tumors & it presented only white picture. See fig(6.b).

Example H=1

a

b

a-The grayscale luminosity after used RK4 for H increase(H=1)

b- Any detection of lesion

Figure 6: Presentation of image after used RK4 (a) and The result obtained after applied anova for H=1(b).

3.5 Error between "y" and fitted regression line

After use RK4 , we pass to calculate the error between fitted vector of y and y of surface (180X180) ,as in eq (7)where we have given an example of two vectors of the both images and detect the lesion or the error by calculate the error for one vector.

Eq(7) is : $e = y - \hat{y}$

Where: y is the vector, & $\hat{Y}$ is data estimate. See (figure7)

a   **b**
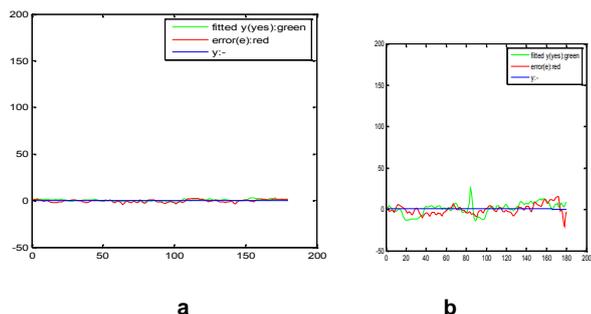
Figure 7 : Presentation of error e, y & $\hat{Y}$ for both curves, where a. curve for H=0.04; b.curve for H= 1.

We note that the error rate in the figure(7.a) lower than in Figure (7.b) and this is what explains and confirms the role of step size H in the control of picture, this is what explains the success of our method and its impact in the discovery of tumors.

After that and from (eq 18), we extract the disease for all image with calculate Fcal by two ways : directly on the pathological image and from Gaussian curve : See (figure8).

After previous studies we are going to discover tumors through the use of an equation anova (eq 18), where fig(8.a) represent pathological image , figure(8-b) present gauss curve (gaussian distribution of Anova –test) ,and from the table of Anova(fisher table) test we have :

for $\alpha = 0.01$, and from the table of fisher test we have :

*Degree of liberty: ddl:

$\nu = n$ -1= 180-1=179.

*F-tab = 1.

*Hypothesis ho will be :

If F-cal $\leq F$ -tab $\Rightarrow$ f-test with (anova) ➔ accept H_0.

If F -cal $> F$-tab $\Rightarrow$ test F-test reject H_0.(red color)

The figure(8-c) present the curve of anova technique as anova equation:F- cal=MS_R/MS_E.

Where:

Fcal present with green color its value beside zeros number.

MS_R: The within-groups variation which present with blue star color .

MS_E:The between-groups, it is present it with red color.

Figure(8-d) detect the lesion directly on the pathological image ,which presented with white color .
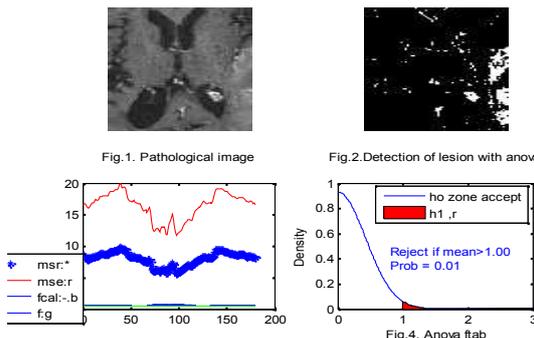
Figure 8: Detect the lesion with two ways

The next fig(9) present the hypothesis accept or reject between fcal & ftab for p-value=0.01.
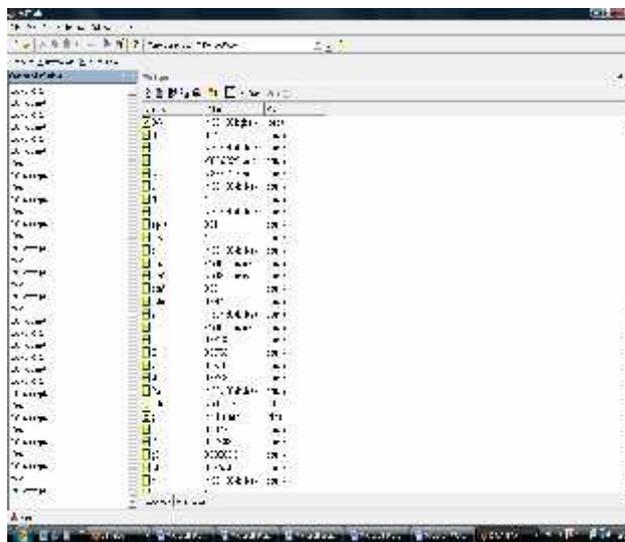
Figure 9 : Presentation of hypothesis ho or h1 as display in Figure(8-b).

3.6 Comparison between ANOVA with RK4 or without it

The result with RK4 presented in figure(10.a) is more precise in front of anova without it as in figure(10.b) because with Runge kutta 4 we have the step size H which That control step to show the image resolution.

The best result is for H minimal and here Hence shows the

superiority of our idea of using RK4.

Fig.1. Pathological image   Fig.2.Detection of lesion with anova

a   **b**
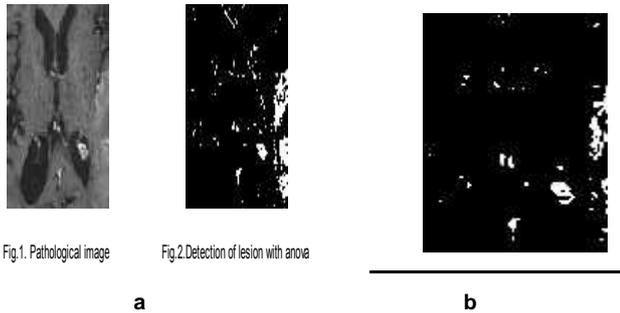
Figure 10:   Comparaison between (a& b).

a:Application Anova for RK4 on the pathological image.

b: Linearization without RK4.

4   CONCLUSION

From this study we conclude the following:

- Numeric analysis with Runge kutta is used to pass from non linear to linear model & control any image from the step size H.

- The step size H is working to increase the lighting of the image compared with the original picture.

- If decreasing H to 0 and (H>0), the result will be more precise & present an excellent solution in front of increasing H to 1.

- Detect the lesion with RK4 is best then detect it without it.

- For the step size H increased more than 0.5, we can't detect any lesion that means there is no any detection (we will obtain white image).

- The Error value will be minimal, if the step size H is very small.

- The statistical study is used to do the comparison between two models: a new and old one in different specialty.

- We propose for researcher to use always Anova technique with RK4, but for nonlinear regression and for multi-images.

REFERENCES

[1] David G. Herr.1986. On the History of ANOVA in Unbalanced, Factorial Designs: The First 30 Years. The American Statistician, 40.4. 265-270.

[2] Hogg, R. V., and J. Ledolter. 1987.Engineering Statistics. MacMillan Publishing Company.

[3] John H.Mathews, Kurtis D.Fink , 1999, Numerical Method using Matlab .3rd ed ,Prentice Hall, upper saddle river,NJ07458,.458-459.

[4] Atkinson, Kendall A. 1989, An Introduction to Numerical Analysis .2nd ed, New York: John Wiley & Sons, ISBN 978-0-471-50023-0 .

[5] Ascher, Uri M.; Petzold, Linda R. 1998, Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations, Philadelphia: Society for Industrial and Applied Mathematics, ISBN 978-0-89871-412-8.

[6] Armstrong, J. Scott 2012. Illusions in Regression Analysis.International Journal of Forecasting .28 .3: 689.

[7] York .1966. Least-squares fitting of a straight line. Canad. J. Phys. 44: 1079-1086.

[8] Uts, J. and R. Hekerd.2004. Mind on Statistics. Chapter 16 - Analysis of Variance. Belmont, CA: Brooks/Cole - Thomson Learning, Inc.

[9] T. Dudok de Wit 2010. ANALYSE NUMÉRIQUE, Licence de physique – 3ème année, Université d'Orléans – Faculté des Sciences,Université d'Orléans.

[10] Plonsky, M.2007. One Way ANOVA.Retrieved from: http://www.uwsp.edu/psych/stat/12/anova-1w.htm.

[11] K.El Kourd.A.El kourd .2013, The Detection of disease by statistic test of .Analyze of variance. Journal IEEE.ISBN 978-1-4673-5285-.

[12] Gear, C.W. 1971. Numerical Initial Value Problems in Ordinary Differential Equations (EnglewoodCliffs, NJ: Prentice-Hall).

[13] From:http://www.weibull.com/DOEWeb/introduction. ht;12/6/2012, 15:45.

[14] Research Methods I, ANOVA and Multiple Regression.

[15] Viviane Kostrubiec .Les comparaisons multiples: entre mythe et réalité. ;Laboratoire Adaptations Perceptivo-Motrices et Apprentissage (EA 3191).Université Paul Sabatier–Toulouse III.

[16] Jon Roiser and Predrag Petrovic.2006 . t-tests, ANOVA and regression.

[17] R.Henson. and W.Penny.2005 .ANOVAs and SPM.Institute of Cognitive Neuroscience,Wellcome Department of Imaging Neuroscience,University College London.