

# Detect the Tumor with Numerical Analysis and With “ANOVA” Technique for MRI Image

El Kourd Kaouther, Bougourzi Fares, El kourd Souhir  
Electronic institute, Biskra University, Algeria

**Abstract**— In this paper, we translate non linear model to linear one by using numerical analysis: “Runge kutta4 (RK4)” before pass to the statistical study where we used the linear regression then Anova technique to extract the place of the lesion on MRI image by two ways : distribution of Gaussian curve(hypothesis test of ho) and directly on the pathological image. The program used here is Matlab.

**Index Terms**— Anova, Runge kutta, linear regression.

## I. INTRODUCTION

Analysis of variance became widely known after being included in Fisher's 1925 book. [1]Statistical Methods for Research Workers. 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. In its simplest form, ANOVA provides a statistical test of whether or not the means of several groups are all equal, and therefore generalizes *t*-test to more than two groups. Doing multiple two-sample *t*-tests would result in an increased chance of committing an error. For this reason, ANOVAs are useful in comparing two, three, or more means.[1] 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.

## II. BACKGROUND

### 1. Runge kutta

In numerical analysis, the Runge–Kutta(RK) methods (German pronunciation: [ˌʁʊŋəˈkʊʦta]) 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.[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:[1][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:[1],[2]

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

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. Regression

Description about the relationship between two variables where one is dependent and the other is independent.[3]

### 3. Fitted Regression Line

The true regression line corresponding to (2) is usually never known. However, the regression line can be estimated by estimating the coefficients  $\beta_1$  and  $\beta_0$  for an observed data set.[3]

$$E(Y) = \beta_0 + \beta_1 x \quad (2)$$

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  $e$  .see (3).[4],[5],[6]

$$Y = E(Y) + \epsilon \\ = \beta_0 + \beta_1 x + \epsilon \quad (3)$$

The estimates,  $\tilde{\beta}_1$  and  $\tilde{\beta}_0$ , are calculated using least squares. The estimated regression line, obtained using the values of,  $\tilde{\beta}_1$  and  $\tilde{\beta}_0$ , is called the *fitted* line. The least

square estimates,  $\hat{\beta}_1$  and  $\hat{\beta}_0$ , are obtained using the

following equations:(4)

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n y_i x_i - \frac{(\sum_{i=1}^n y_i)(\sum_{i=1}^n x_i)}{n}}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (4)$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (5)$$

Where  $\bar{y}$  is the mean of all the observed values and  $\bar{x}$  is the

mean of all values of the predictor variable at which the observations were taken.  $\bar{y}$  is calculated using

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

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

Once  $\hat{\beta}_1$  and  $\hat{\beta}_0$  are known, the fitted regression line can be

written as:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \quad (6)$$

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,  $x_i$ , may be different from the corresponding observed value,  $y_i$ . The difference between the two values is called the residual,  $e_i$ :

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

To calculate the Statistic  $F_0$ , it must pass by the six titles:[4],[5],[6],[7],[8],[9]

### 3.1. Total Sum of Squares (SST)

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

$$\begin{aligned} SS_T &= \sum_{i=1}^n (y_i - \bar{y})^2 \\ &= \sum_{i=1}^n y_i^2 - \frac{(\sum_{i=1}^n y_i)^2}{n} \end{aligned} \quad (8)$$

The total sum of squares in matrix notation is:

$$\begin{aligned} SS_T &= \mathbf{y}'\mathbf{y} - \left(\frac{1}{n}\right)\mathbf{y}'\mathbf{J}\mathbf{y} \\ &= \mathbf{y}'\left[\mathbf{I} - \left(\frac{1}{n}\right)\mathbf{J}\right]\mathbf{y} \end{aligned} \quad (9)$$

Where  $\mathbf{y}$  is the vector of observed values,  $\mathbf{I}$  is the identity matrix of order  $n$ ; &  $\mathbf{J}$  represents an  $n \times n$  square matrix of ones.

### 3.2. Model Sum of Squares (SSR)

Similarly, the model sum of squares or the regression sum of squares,  $SS_R$ , can be obtained in matrix notation

$$\begin{aligned} SS_R &= \sum_{i=1}^n \hat{y}_i^2 - \frac{(\sum_{i=1}^n y_i)^2}{n} \\ &= \hat{\mathbf{y}}'\hat{\mathbf{y}} - \left(\frac{1}{n}\right)\mathbf{y}'\mathbf{J}\mathbf{y} \\ &= \mathbf{y}'\left[\mathbf{H} - \left(\frac{1}{n}\right)\mathbf{J}\right]\mathbf{y} \end{aligned} \quad (10)$$

as:

Where  $\mathbf{H}$  is the hat matrix and is calculated using

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \quad (11)$$

### 3.3. Error Sum of Squares

The error sum of squares or the residual sum of squares,  $SS_E$ , is obtained in the matrix notation from the vector of residuals,  $\mathbf{e}$ , as:

$$SSE = SS_T - SS_R \quad (12)$$

$$\begin{aligned} SS_E &= \mathbf{e}'\mathbf{e} \\ &= (\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}}) \\ &= \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y} \end{aligned} \quad (13)$$

### 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,  $SS_T$ , 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:[4],[5],[6],[7],[8],[9]

$$MS_T = \frac{SS_T}{n-1} \quad (14)$$

### 3.5. Regression mean square ( $MS_R$ )

The number of degrees of freedom associated with the regression sum of squares,  $SS_R$ , is  $k$ . There are  $(k+1)$  degrees of freedom associated with a regression model with  $(k+1)$  coefficients,  $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ . However, one degree of

freedom is lost because the deviations,  $(\hat{y}_i - \bar{y})$ , are

subjected to the constraints that they must sum to zero  $\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ . The regression mean square is:

$$MS_R = \frac{SS_R}{k} \quad (15)$$

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  $\beta_0, \beta_1, \beta_2 \dots \beta_k$  to calculate the predicted

values  $\hat{y}_i$ . The error mean square is:[4][5][6][7][8][9]

$$MS_E = \frac{SS_E}{n - (k + 1)} \tag{16}$$

The error mean square,  $MS_E$ , is an estimate of the variance,  $\sigma^2$ , of the random error terms,  $e_i$ .

### 3.6. Mean square error( $MS_E$ )

Is estimate variance ( $\hat{\sigma}^2$ ) of random error  $e_i$ .see (15)

$$\hat{\sigma}^2 = MS_E \tag{17}$$

### 3.7. Calculation of the Statistic $F_0$

Once the mean squares  $MS_R$  and  $MS_E$  are known, the statistic to test the significance of regression can be calculated as follows:[2][4][5]

$$F_0 = \frac{MS_R}{MS_E} \tag{18}$$

## III. EXPERIMENTAL RESULTS

### 1. Algorithm

- 1) Read the images
- 2) Choose a sample of image
- 3) Applied analyze mathematic with Runge kutta4.
- 4) Estimate the linear regression with parameters estimates (image x).
- 5) Extract the error between y & y estimate.
- 6) Application the analyses of variance 'ANOVA'.
- 7) Conception & results.
- 8) Comparison between Anova and protocol of radiologic.

### 2. Analysis of data (the protocol radiologic)

Our protocol is for a patient aged 15 years; who is present in clinical context of late stature and puberty with pituitary dysfunction. The patient is made by an MRI scan with injection of contrast medium. The machine is used radiologist type "SIEMENS", and with the field. B = 1.5 Tesla. The sequences performed T1-weighted and T2. MRI appearance of a tumor of the median line developed at the expense's floor of the third ventricle and the left sidewall .See fig1-b. Fig 1 Present an Image Normal In (Fig (1-A)) & Pathological One in (Fig (1-B))

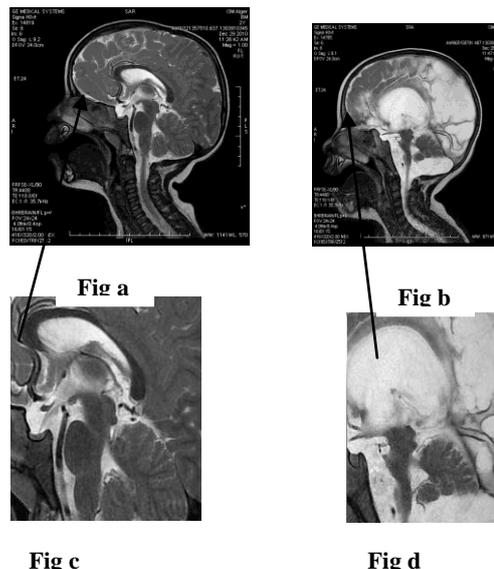


Fig.1."left .Normal image ; right : Pathologic image"

The fig(1-c) ,is a section of normal image and fig(d) is a section (120X120)choosing to analyze it and to detect it .

The error result for one vector (with length n=200: see (7) :  $e = y - \hat{y}$ ), display in fig(2)

Where: y is the vector of normal image , &  $\hat{Y}$  is data estimate (pathological image). See (fig(2))

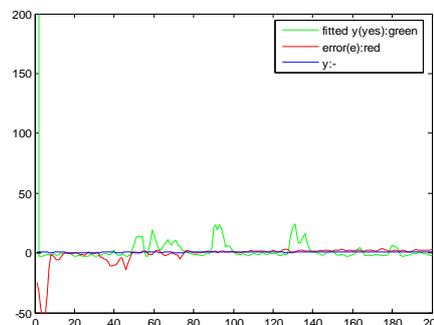


Fig.2: "Presentation error e, y &  $\hat{Y}$ ".

Y present with blue color;  $\hat{Y}$  is with green color & error is with red color where between (0 & 45 of the surface), it's present the place of disease.

To Calculate F cal :See (18) with two methods: direct and from computation: See fig (3)

From fig (3) the first image(fig3-a) is the pathological image. The next image fig (3-b), detect the lesion directly on the pathological image(see white color) .The next figure(3-c) is the curve or computation of anova technique as following:(18):F- cal= $MS_R/MS_E$

Where:

$MS_R$ : The within-groups variation.

$MS_E$ : the between-groups.

precise if we compare it with the report of radiologic for unknown places.

V. APPENDIX

We propose using Anova but for non linear regression and for data bus of images.

REFERENCES

- [1] T. Dudok de Wit, ANALYSE NUMÉRIQUE, Licence de physique – 3ème année, Université d’Orléans – Faculté des Sciences, Université d’Orléans, Septembre 2010.
- [2] Chapitre 5 - Comparaison de plusieurs moyennes pour des échantillons indépendants (ANOVA), Université Paris X – Nanterre UFR SPSE-Master1, PMP EF 205 Méthodes Statistiques pour l’analyse de données en psychologie ,2008-2009
- [3] <http://www.weibull.com/DOEWeb/introduction.htm>;12/6/201 2, 15:45
- [4] K.El Kourid , The Detection of disease by statistic test of "Analyze of variance", 978-1-4673-5285-7/13/\$31.00 ©2013 IEEE
- [5] Research Methods I, ANOVA and Multiple Regression,
- [6] UFR SPSE-Master1,Université Paris X – Nanterre;PMP EF 205;Méthodes Statistiques pour l’analyse de données en psychologie "Chapitre5. Comparaison de plusieurs moyennes pourdes échantillonsindépendants (ANOVA)"; 2008-2009.
- [7] Viviane Kostrubiec ; "Les comparaisons multiples: entre mythe et réalité" ;Laboratoire Adaptations Perceptivo-Motrices et Apprentissage (EA 3191);Université Paul Sabatier–Toulouse III
- [8] Jon Roiser and Predrag Petrovic; "t-tests, ANOVA and regression"; February 1st 2006.
- [9] R.Henson (1) and W. Penny (2), ANOVAs and SPM, Institute of Cognitive Neuroscience, Wellcome Department of Imaging Neuroscience, UniversityCollegeLondon.July12, 2005.



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

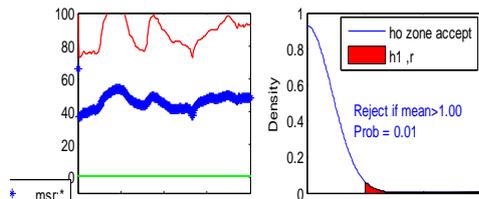


Fig c :Curve of Anova Fig d : Normal distribution

Fig3.”Detect the lesion with two ways (curve & on image)”

F-cal present with green color, MSR is with blue star color & MS<sub>E</sub> with red Color. For the last figure(fig3- d) ,it present gauss curve (gaussian distribution of Anova –test)for  $\alpha = 0.01$ ,and from the table of fisher test we have :

\*Degree of liberty: ddl:

$$v = n - 1 = 200 - 1 = 199.$$

\*F-tab = 1.

\*Hypothesis ho will be :

- If  $F\text{-cal} \leq F\text{-tab} \Rightarrow f\text{-test with (anova)} \rightarrow \text{accept } H_0.$
  - If  $F\text{-cal} > F\text{-tab} \Rightarrow \text{test } F\text{-test reject } H_0.$  (red color)
- The next figure(4) present the hypothesis accept or reject

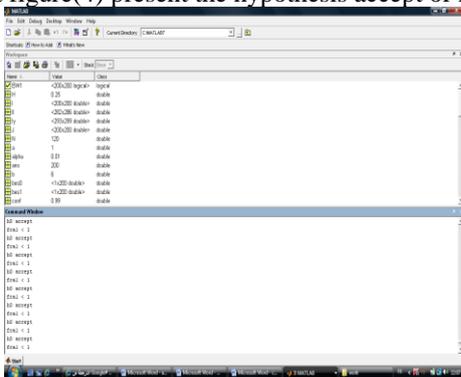


Fig4. “Presentation of Hypothesis Ho or H1 as Display in Fig (3-D)”

3. Comparison with protocol of radiologic

Here our technique extracts the lesion as report of radiologic; The disease is with white color see fig3: (fig2))

IV. CONCLUSION

From these studied we can distinct.:

- Numeric analysis with runge kutta is used to pass from non linear to linear model.
- Our study is applied for comparison between two models a new and old one in different specialty.
- The result with Anova for non linear model and for detect the lesion with precise and will be more