

## Full Length Research Paper

**Runge kutta to precise the detection of lesion for magnetic resonance imaging (MRI) image**

K. El kourd\* and S. El kourd

Electronic institute, Biskra University, Algeria.

Received 17 July, 2013; Accepted 27 October, 2014

In this paper, we translate non linear model to linear one and used numerical analysis: "Runge kutta4 (RK4)" is mathematical solutions to study the approximation's solutions of ordinary differential equations, then pass to the statistical study for multiple regression to applied anova technique and with it, we extract the place of the lesion on medical image MRI by two ways: Distribution of gaussian curve (hypothesis test of  $h_0$ ) and directly on the pathologic image. The logiciel applied here is Matlab.

**Key words:** Runge kutta, linear regression, anova.

**INTRODUCTION**

Analysis of variance became widely known after being included in Fisher's 1925 book, *Statistical Methods for Research Workers* (David, 1986). 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 (David, 1986).

**Background****Runge kutta**

In numerical analysis, the Runge–Kutta (RK) methods [German pronunciation: (ʁuŋə'kʊtə)] 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 (Atkinson, 1989; Ascher et al., 1998).

**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 the spaceship moves in two-dimensional space. Its acceleration is described by the following system of differential equations (Atkinson, 1989):  $x''=f(x, y, x', y')$  and  $y''=g(x, y, x', y')$ , 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,  $k_1, k_2, k_3$  and  $k_4$ ; the final approximation will be given by a weighted average of these intermediate approximations.

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

After translating the non linear model (exponential equation) to linear one, the linear regression and the fitted regression line was calculated and anova test was done under the following:

\*Corresponding author. E-mail: kaouther\_youcef@yahoo.fr, el\_kourd@yahoo.com

Author(s) agree that this article remain permanently open access under the terms of the [Creative Commons Attribution License 4.0 International License](http://creativecommons.org/licenses/by/4.0/)

**Regression:** Description about the relationship between two variables where one is dependent and the other is independent (Armstrong, 2012; York, 1966).

**1. Fitted regression line:** The true regression line corresponding to equation (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 (Armstrong, 2012; York, 1966):

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

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: 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, Equation (3):

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

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,  $\tilde{\beta}_1$  and  $\tilde{\beta}_0$ , are obtained using the following equation:

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

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

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

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

Once  $\tilde{\beta}_1$  and  $\tilde{\beta}_0$  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* (Armstrong, 2012; York, 1966):

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

To calculate the Statistic  $F_0$ , it must pass by the six titles (Uts and Hekerd, 2004; Dudok, 2010; Plonsky, 2007; El Kourd and El kourd, 2013; Gear, 1971; Jon and Predrag, 2006; Henson and Penny, 2005; <http://www.weibull.com/DOEWeb/introduction.ht>;12/6/2012, 15:45; Research Methods I, ANOVA and Multiple Regression; Viviane Kostrubiec. Les comparaisons multiples: entre mythe et réalité).

**Total sum of squares (SST):** On simple linear regression that the total sum of squares,  $SS_T$ , is obtained using the following equation:

$$SS_T = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n} \quad (8)$$

The total sum of squares in matrix notation is:

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

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

**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 as:

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

Where H is the hat matrix and is calculated using:

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

**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, e, as:

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

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

**Mean squares (MS<sub>T</sub>):** 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:

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

**Regression mean square (MS<sub>R</sub>):** The number of degrees of freedom associated with the regression sum of squares,  $SS_R$  is k.

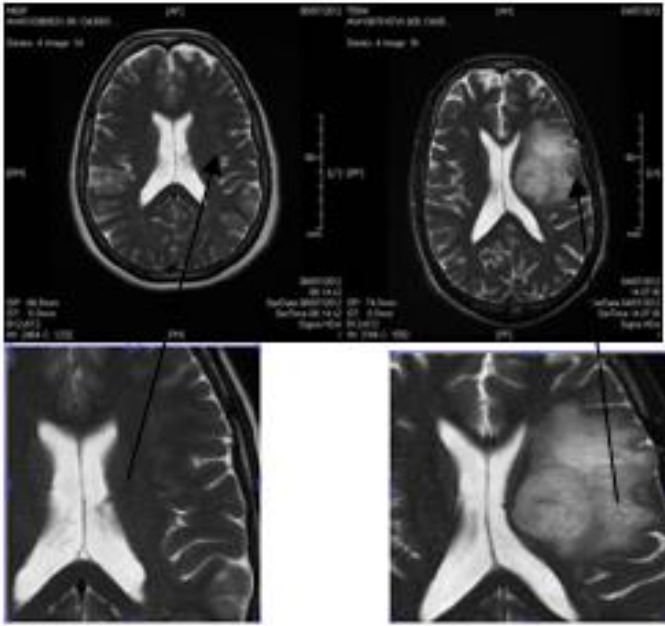


Figure 1. Normal image (left) and pathological image (right).

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,  $(\tilde{y}_i - \bar{y})$ , 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:

$$MS_R = \frac{SS_R}{k} \tag{15}$$

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,  $\tilde{y}_i$ . The error mean square is (Ascher et al., 1998; rmstrong, 2012; York, 1966; Uts and Hekerd, 2004; Dudok, 2010):

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

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

**Mean square error (MS<sub>E</sub>):** MS<sub>E</sub> is estimate variance ( $\hat{\sigma}^2$ ) of random error  $e_i$ .

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

**Calculation of the statistic F<sub>0</sub>:** Once the mean squares MS<sub>R</sub> and MS<sub>E</sub> are known, the statistic to test the significance of regression can be calculated as follows (Atkinson, 1989):

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

## EXPERIMENTAL RESULTS

### Algorithm

- (1) Read the images
- (2) Choose a sample of image
- (3) Calculate the error between y and y estimate then plot it for one vector .
- (4) Applied analyse mathematic with runge kutta4.
- (5) Estimate the linear regression with parameters estimates (image x).
- (6) Application the analyses of variance 'ANOVA' .
- (7) Compeer the result of Anova with Runge kutta and the technical of Anova with linear model

### Analysis of data (the protocol radiologic)

Our protocol is for a patient aged 55 years; He made 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 appears in the middle of the field on the left side and the third ventricle is to evoke the tumor (Figure 1).

The left bottom is a section of normal image and the right one is the section of image; it has surface (200x200), our analysis was then applied, but before that the lesion or the error between the both images was detected by calculating the error for one vector such as:

$$e = y - \hat{y} \tag{19}$$

Where: y is the vector, and  $\hat{y}$  is data estimate (Figure 2). Figure 2 from length (0 and 45) of image present the

place of disease (red color). Y is with blue color and  $\hat{Y}$  is with green one. From (Equation 18), we extract the disease for all image with calculate Fcal by two ways directly on the pathological image and from Gaussian curve (Figure 3). Figure 3a present the pathological image. Figure 3b detects the lesion directly on the pathological image, which were represented with white color. Figure 3 present the curve of anova technique as anova equation:

$$F\text{-cal} = MS_R / MS_E$$

Where: Fcal present with green color its value beside zeros number. MS<sub>R</sub>: The within-groups variation which present with blue star color. MS<sub>E</sub>: The between-groups, it is with red color. The last Figure 3d 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:

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

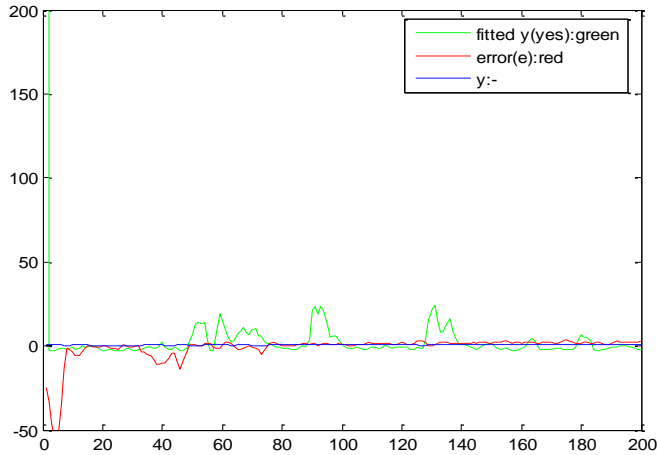


Figure 2. Presentation error e, y and  $\hat{Y}$ .

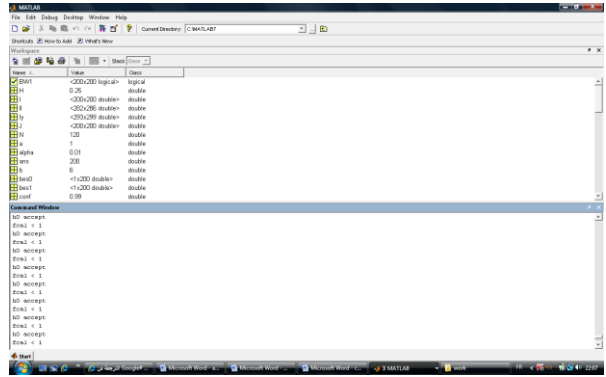


Figure 4. Presentation of hypothesis  $h_0$  or  $h_1$  as display in Figure 3d.

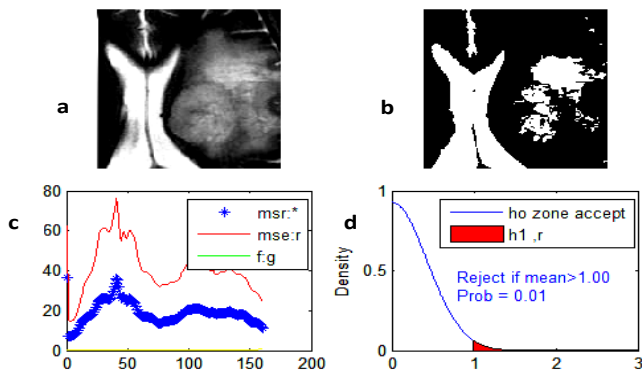


Figure 3. Ways of detecting lesion. (a) Pathological image; (b) Detection of lesion with anova; (c) Computation of anova; (d) Anova  $F$ -tab.

\* $F_{-tab} = 1$ .

\*Hypothesis  $h_0$  will be:

$$\begin{cases} \text{If } F_{-cal} \leq F_{-tab} \Rightarrow \text{f-test with (anova)} \rightarrow \text{accept } H_0. \\ \text{If } F_{-cal} > F_{-tab} \Rightarrow \text{test } F_{-test} \text{ reject } H_0 \text{ (red color)} \end{cases}$$

Figure 4 present the hypothesis accept or reject.

**Comparison between two methods of ANOVA**

The result with model non linear presented in Figure 5a is more precise in front of anova with linear model (Figure 5a)

**Conclusion**

From these study we can distinct:

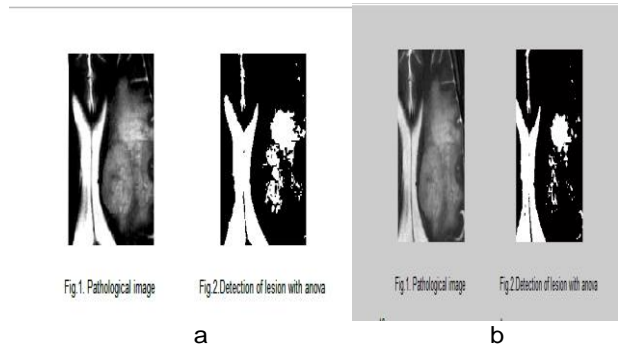


Figure 5. Comparison between Application Anova for non linear model on the pathological image (a) and linearization with anova simple (b).

1. Numeric analysis with Runge kutta is used to pass from non linear to linear model.
2. Runge kutta used for approximate the solutions of ordinary differential equations; from here come our idea. It can precise results and achieve to the complex places in image, where doctors cannot see it clearly.
3. The statistical study is used to do the comparison between two models: A new and old one in different specialty.
4. In this article we have used a medical image with MRI scan.
5. The result with Anova pass by, "non linear model" then do the linearization, is more precise then used directly the linear model of Anova.

We propose for researcher to use always Anova technical, but for non linear regression and for multi-images.

**Conflict of Interest**

The authors have not declared any conflict of interest.

## REFERENCES

- David GH (1986). On the history of ANOVA in unbalanced, factorial designs: The First 30 Years. *Am. Statist.* 40(4):265-270. <http://dx.doi.org/10.2307/2684597>
- Atkinson KA (1989). *An introduction to numerical analysis* (2nd ed.), New York: John Wiley & Sons, ISBN 978-0-471-50023-0.
- Ascher Uri M, Petzold LR (1998). *Computer methods for ordinary differential equations and differential-algebraic equations*, Philadelphia: Soc. Indust. Appl. Math. ISBN 978-0-89871-412-8. <http://dx.doi.org/10.1137/1.9781611971392>
- Armstrong JS (2012). Illusions in regression analysis. *Int. J. Forecast.* 28(3):689. <http://dx.doi.org/10.1016/j.ijforecast.2012.02.001>
- York D (1966). Least-squares fitting of a straight line. *Canad. J. Phys.* 44:1079-1086. <http://dx.doi.org/10.1139/p66-090>
- Uts J, Hekerd R (2004). *Mind on Statistics. Chapter 16 - Analysis of Variance*. Belmont, CA: Brooks/Cole-Thomson Learning, Inc.
- Dudok de Wit T (2010). *Analyse Numérique, Licence de physique – 3ème année, Université d'Orléans –Faculté des Sciences, Université d'Orléans.*
- Plonsky M (2007). One Way ANOVA. Retrieved from: <http://www.uwsp.edu/psych/stat/12/anova-1w.htm>.
- El Kourid K, El kourid A (2013), The detection of disease by statistic test of analyze of variance. *J. IEEE.* ISBN 978-1-4673-5285.
- Gear CW (1971). *Numerical initial value problems in ordinary differential equations* (EnglewoodCliffs, NJ: Prentice-Hall).PMCID:PMC1411756 <http://www.weibull.com/DOEWeb/introduction.ht;12/6/2012>, 15:45.
- Research Methods I, ANOVA and Multiple Regression, Viviane Kostrubiec. *Les comparaisons multiples: entre mythe et réalité. Laboratoire Adaptations Perceptivo-Motrices et Apprentissage (EA 3191). Université Paul Sabatier–Toulouse III.*
- Jon R, Predrag P (2006). *t-tests, ANOVA and regression*. Henson R, Penny W (2005) .ANOVAandSPM.InstituteofCognitiveNeuroscience,WellcomeDepartm entofImagingNeuroscience,UniversityCollegeLondon.

