American Journal of Computational Mathematics, 2018, 8, 68-77
http://www.scirp.org/journal/ajcm
ISSN Online: 2161-1211
ISSN Print: 2161-1203

Simple and Multi Linear Regression Model of Verbs in Quran

Abdelkrim El Mouatasim

Faculty of Polydisciplinary Ouarzazate (FPO), Ibn Zohr University, Ouarzazate, Morocco
Email: a.elmouatasim@uiz.ac.ma

How to cite this paper: El Mouatasim, A. (2018) Simple and Multi Linear Regression Model of Verbs in Quran. *American Journal of Computational Mathematics*, **8**, 68-77.
https://doi.org/10.4236/ajcm.2018.81006

Received: February 17, 2018
Accepted: March 13, 2018
Published: March 16, 2018

Abstract

This paper mainly presented a good simple and multi-linear regression model of verbs in the Quran book. This model, gives an analysis for the influence to frequency of words with the form (—un, ---ون) made by the frequency of plural present verbs (t—un, ت---ون) or (y—un, ي---ون), and models, and the relationship between independent variables and dependent variable by fitting a linear equation to the observed data with simple linear regression model. The matlab function is used for finding the parameters of the linear regression model and plotting the fits. The results show that the parameters of the model are one vector (1, 1) and mean of dataset is (6, 7). Its corresponding to the verb with input is frequency of the verb they enter and the frequency of enter (yadkolun يدخلون ، دخلون dakilun), also other 17 points exist in the line and in the dataset of 387 verbs and their derivate verbs in Quran. The name of Allah (الله) showed when we use tree variables and plot it in 3D with option “Show Text” for a multi regression model.

Keywords

Linear Regression, Text Mining, Quran Statistics, Matlab, Arabic Grammar, Optimization, Computation Linguistics

1. Introduction

The scripture of the Quran has been subjected to various intense mathematically based studies to reveal the protection mechanisms embedded in the composition of the Quran and to provide evidence of its credibility, authenticity and divinity see for instance [1] [2].

Therefore, the development of the mathematical theory has been highly motivated and driven by the categorical recognition of the author that Allah may have embedded varying mathematical algorithms, equations and regression models for protecting the Quran, as well as to prove its divinity and to emphatically exclude any human influence on the manufacture of the Quran. Because Allah promises that the Quran will always be preserved and protected from any corruption such as addition or deletion or relocation of any of its verses from chapter to another. Therefore, unveiling any of these algorithms would help unlock many of the Quranic secrets, particularly those related to the Quran's primary parameters like words and verbs as well as how the Quran's design is related to the fit of linear regression.

Furthermore, this work is set to statistically and numerically validate and authenticate the first drawing of the Quran (Uthmanic manuscript) related statistics such as the total number of words and verbs of the Quran.

Regression analysis describes the relationship between a dependent variable and several independent variables.

Regression analysis describes the relationship between a dependent variable and several independent variables, for the estimation of the parameters model see for instance [3] [4].

This paper is organized as follows: in Section 2, we give the initiation of linear regression; linear regression model in Quran with numerical results is given in Section 3.

2. Linear Regression Models

2.1. Simple Linear Regression

Regression analysis is a statistical technique for estimating the relationship among variables which have reason and result relation. Main focus of univariate regression is analyses the relationship between a dependent variables $X_1, \cdots, X_n$ and one independent variable Y and formulates the linear relation equation between dependent and independent variable.

The simple linear regression model is the simplest regression model in which we have only one predictor X.

This model, which is common in practice, is written as

$$Y_i = b + aX_i + \varepsilon_i,\ \ i = 1, \cdots, n,$$

where

- Y_i, X_i are the values of the response and predictor variables in the trial, respectively;
- The unknown parameters: a is called the intercept, and b is the slope of the line;
- ε_i is usually assumed to be *iid* (error) from $N\left(0, \sigma_\varepsilon^2\right)$ specially for inference purposes (see for instance [5]).

Then estimates of simple linear model's parameters should be obtained accordingly, using some method like the ordinary least squares method, which relies on minimizing the sum of square of errors $\sum \varepsilon_i^2$.

For the simple linear regression model the ordinary least squares estimations of a and b are

$$\hat{b} = \bar{Y} - \hat{a}\bar{X}$$

and

$$\hat{a} = \sum_i \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right) \Big/ \sum_i \left(X_i - \bar{X}\right)^2$$

where $\bar{X}$ and $\bar{Y}$ are the mean of the variable X and the variable Y respectively.

The goodness R^2 of fit is defined as

$$R^2 = \sum_i \left(Y_i - \bar{Y}\right)^2 \Big/ \sum_i \left(X_i - \bar{X}\right)^2 .$$

2.2. Matrix Form of Multiple Regression

Regression models with one dependent variable and more than one independent variable are called multi-linear regression (see for instance [6]).

Multivariate regression analysis model is formulated as in the following:

$$Y = \alpha_0 + \alpha_1 X_1 + \cdots + \alpha_n X_n + \varepsilon$$

where

- Y is the dependent variable,
- X_i is the independent variables,
- α_i is the parameters,
- ε is the error.

The assumptions of multi-linear regression analysis are normal distribution, linearity, freedom extreme values and having no multiple ties between independent variables [6].

The linear model can be written as

$$Y = X\alpha + \varepsilon$$

where

- $Y \in R^n$ is the vector of observations on the dependent variable,

$$Y = \left(Y_1, \cdots, Y_n\right)^{\mathrm{T}} .$$

- $X \in R^n \times R^{p+1}$ is the matrix consisting of a column of ones and p column vectors of the observations on the independent variables, the form of X is

$$\begin{matrix} \mathbf{1} & X_{11} & \cdots & X_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{1} & X_{n1} & \cdots & X_{np} \end{matrix}$$

- $\alpha \in R^{p+1}$ is the vector of parameters to be estimated,

$$\alpha = \left(\alpha_0, \alpha_1, \cdots, \alpha_p\right)^{\mathrm{T}} .$$

- $\epsilon \in R^n$ is the vector of random errors,

$$\epsilon = \left(\epsilon_0, \epsilon_1, \cdots, \epsilon_n\right)^{\mathrm{T}} .$$

The vector α is a vector of unknown constants to be estimated from the data by $\hat{\alpha}$.

The normal equations [7] are written as

$$X^{\mathrm{T}} X \hat{\alpha} = X^{\mathrm{T}} Y.$$

If $X^{\mathrm{T}} X$ has an inverse, then the unique solution of normal equations given by

$$\hat{\alpha} = \left(X^{\mathrm{T}} X\right)^{-1} \left(X^{\mathrm{T}} Y\right).$$

The vector $\hat{Y}$ of estimated means of the dependent variable Y for the values of the independent variables $X_1, \cdots, X_n$ in the dataset is computed as

$$\hat{Y} = X \hat{\alpha}.$$

However, to express $\hat{Y}$ as a linear function of Y. Thus,

$$\hat{Y} = \left[X \left(X^{\mathrm{T}} X\right)^{-1} X^{\mathrm{T}}\right] Y.$$

3. Model and Numerical Results

3.1. Definition of Variables

In this section we conceder the following integer variables:

- X_1 the frequency of plural present verbs with a form (y—un, ي---ون) or there inverse (ly—un, لي---ون);
- X_2 the frequency of plural present verbs with a form (t—un, ت---ون) or there inverse (lt—un, لت---ون);
- X_3 the frequency of the verbs in the above form without (t,ت) or (y, ي).

We use the software of Quran statistics [8] let $Y = X_1 + X_2 + X_3$ for determined the set of triple $\left(Y^i, X_1^i, X_2^i\right), i = 1, \cdots, 387$.

Let Y^7, Y^9, Y^{32} is the sum of frequency of all above derivative verbs with same form.

However, the dataset are given in Table 1.

3.2. Simple Linear Regression Model of Verbs

Since X_3 are not a frequency of plural present verbs [9] then, we let $X = X_1 + X_2$ a dependent variable and Y an independent variable for simple linear regression model. However, the simple linear regression model of verbs is:

$$Y \approx aX + b,$$

where a and b are the parameters of the model.

For estimate the parameters a and b, calculate the coefficient of correlation R, plotting a fit and test hypothesis of simple linear regression model we used the following matlab codes:

```
[r, m, b] = regression (X, Y);
plotregression (X, Y);
fitlm (X, Y);
```

Table 1. Dataset of words in Quran.

Index	(Y^i, X_1^i, X_2^i)	Index	(Y^i, X_1^i, X_2^i)	Index	(Y^i, X_1^i, X_2^i)	Index	(Y^i, X_1^i, X_2^i)
1	(143, 57, 83)	2	(141, 85, 56)	3	(94, 87, 7)	4	(91, 83, 8)
5	(74, 42, 32)	6	(74, 2, 0)	7	(71, 0, 0)	8	(66, 22, 24)
9	(51, 0, 0)	10	(48, 28, 20)	11	(45, 22, 19)	12	(43, 13, 14)
13	(41, 28, 5)	14	(38, 14, 20)	15	(37, 18, 19)	16	(36, 19, 7)
17	(32, 25, 7)	18	(29, 9, 19)	19	(29, 6, 10)	20	(28, 11, 17)
21	(26, 20, 2)	22	(25, 21, 4)	23	(25, 10, 12)	24	(25, 0, 1)
25	(23, 15, 6)	26	(23, 10, 13)	27	(23, 10, 2)	28	(22, 20, 2)
29	(22, 16, 6)	30	(21, 16, 5)	31	(21, 12, 9)	32	(21, 0, 0)
33	(20, 18, 2)	34	(20, 1, 0)	35	(18, 14, 4)	36	(17, 14, 3)
37	(16, 15, 1)	38	(16, 11, 4)	39 - 40	(16, 10, 6)	41	(15, 14, 1)
42	(15, 14, 1)	43	(14, 13, 1)	44	(14, 12, 2)	45	(14, 11, 3)
46	(14, 5, 1)	47	(13, 7, 6)	48	(13, 2, 11)	49	(12, 11, 1)
50	(12, 10, 2)	51	(12, 9, 3)	52	(12, 5, 7)	53 - 54	(12, 3, 9)
55	(11, 8, 3)	56	(11, 7, 4)	57	(11, 5, 6)	58	(10, 9, 1)
59	(10, 9, 1)	60 - 61	(10, 7, 3)	62	(10, 6, 4)	63	(10, 5, 5)
64	(10, 3, 7)	65	(9, 8, 1)	66	(9, 7, 2)	67	(9, 6, 3)
68	(9, 4, 5)	69	(9, 4, 4)	70	(9, 0, 1)	71 - 73	(8, 7, 1)
74	(8, 5, 2)	75	(8, 4, 4)	76	(8, 0, 8)	77	(8, 0, 1)
78 - 82	(7, 7, 0)	83 - 85	(7, 6, 1)	86	(7, 6, 1)	87	(7, 5, 2)
88 - 89	(7, 5, 0)	90	(7, 5, 2)	91	(7, 4, 3)	92	(7, 3, 0)
93	(7, 2, 2)	94	(7, 1, 0)	95 - 96	(6, 6, 0)	97 - 99	(6, 5, 1)
100	(6, 5, 0)	101	(6, 4, 2)	102 - 104	(6, 3, 3)	105	(6, 2, 0)
106	(6, 1, 1)	107	(6, 1, 4)	108 - 113	(5, 5, 0)	114 - 117	(5, 4, 1)
118	(5, 3, 2)	119	(5, 1, 2)	120 - 121	(5, 1, 4)	122 - 129	(4, 4, 0)
130 - 137	(4, 3, 1)	138 - 141	(4, 2, 2)	142	(4, 2, 1)	143	(4, 1, 0)
144 - 145	(4, 1, 3)	146 - 162	(3, 3, 0)	163 - 168	(3, 2, 1)	169 - 173	(3, 1, 2)
174 - 175	(3, 1, 1)	176	(3, 1, 0)	177 - 178	(3, 0, 1)	179 - 212	(2, 2, 0)
213 - 217	(2, 1, 0)	218 - 230	(2, 1, 1)	231 - 235	(2, 0, 2)	236	(2, 01)
237 - 338	(1, 1, 0)	339 - 387	(1, 0, 1)				

The results of regression m function is given as the parameters $a = 1$ and $b = 1$ the coefficient of correlation $R = 0.91441$, see also **Figure 1**. And the results of test hypotheses by fitlm. m function are:

```
>>

Linear regression model:
    y ~ 1 + x1

Estimated Coefficients:
                   Estimate      SE        tStat       pValue
                   ________    ________    ______    ___________

    (Intercept)    1.0073      0.34185     2.9468     0.0034064
    x1             0.99878     0.022534    44.324    2.7071e-153


Number of observations: 387, Error degrees of freedom: 385
Root Mean Squared Error: 6.18
R-squared: 0.836,    Adjusted R-Squared 0.836
F-statistic vs. constant model: 1.96e+03, p-value = 2.71e-153
```

The mean point $G = (\bar{X}, \bar{Y})$ is determined as follow:

```
xb=sum(X)/length(x)
yb=sum(Y)/length(x)
```

we find

```
>>

xb =

     6

yb =

     7
```

then $G = (6,7)$, for searching a correspondence point in dataset we use this matlab commends:

```
a=find(X==6);
b=find(Y(a)==7);
a(b)
```

we find

```
>>

ans =
    86
```

$\left(Y^{86}, X^{86}\right)$ is frequency of verbs of enter (yadkolun يدخلون ، دخلون dakilun).

Also there exist 17 point in dataset in the line fit their equation is $Y = X + 1$, when we use the commend *find* $(Y = X + 1)$ there indexes in dataset are:

86
(Y^{86}, X^{86}) is frequency of verbs of enter (yadkolun يدخلون ، دخلون dakilun).

Also there exist 17 point in dataset in the line fit there equation is $Y = X + 1$, when we use the commend *find(Y=X+1)* there indexes in dataset are:

```
>>

  Columns 1 through 11

    18    27    38    69    74    86    99   107   140   171   174

  Columns 12 through 17

   213   214   216   217   230   236
```

For more information's about the correspondence of this points in the dataset see Table 2 see for instance [10] [11].

3.3. Multi Linear Regression Model of Verbs

For multi linear regression model of verbs:

$$Y \approx \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3$$

where α_1, α_2 and α_3 are the parameters of the model.

For estimate the parameters α_1, α_2 and α_3, and test hypothesis of multi linear regression model we used the following matlab code:

```
tbl = table(y,x1,x2,'VariableNames',{'falon','tfalon','yfalon'});
fitlm(tbl,'falon~tfalon+yfalon')

The results of test hypotheses by fitlm.m function are:
>>

Linear regression model:
    falon ~ 1 + tfalon + yfalon

Estimated Coefficients:
                   Estimate       SE        tStat       pValue
                   ________    ________    ______    __________

    (Intercept)     1.0335      0.34425    3.0021     0.0028563
    tfalon         0.97168     0.045881    21.178    1.6537e-66
    yfalon          1.0425     0.068244    15.275    1.6984e-41
Number of observations: 387, Error degrees of freedom: 384
Root Mean Squared Error: 6.18
R-squared: 0.836,   Adjusted R-Squared 0.835
F-statistic vs. constant model: 981, p-value = 1.2e-151
```

The estimation of parameters $(\alpha_1, \alpha_2, \alpha_3) \approx (1,1,1)$.

For plotting a dataset in 3D we use the following Matlab commends:

[X,Y] = meshgrid(x1,x2);
Z1 = meshgrid(x3);
Z = X+Y+Z1 ;
contour(X,Y,Z,10,'ShowText','on')

Number of observations: 387, *Error degrees of freedom*: 384
Root Mean Squared Error: 6.18
R-squared: 0.836, *Adjusted R-Squared* 0.835
F-statistic vs. constant model: 981, *p-value* = 1.2e−151

Also in the **Figure 2**, we show the name of Allah in Arabic الله.

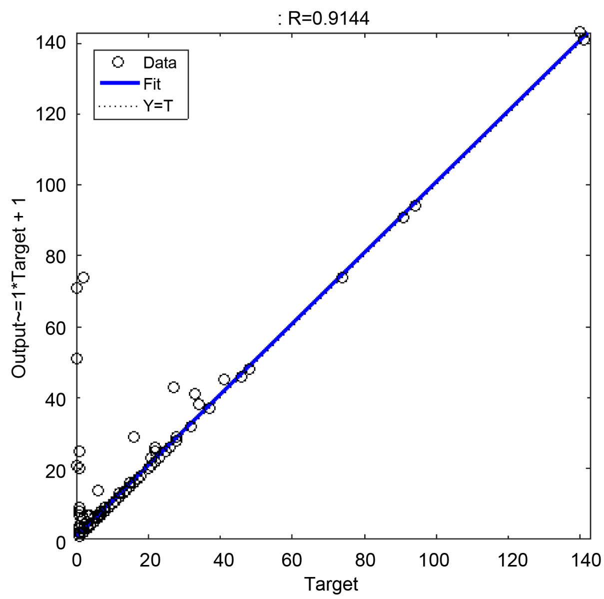

Figure 1. Plotting dataset in 2D.

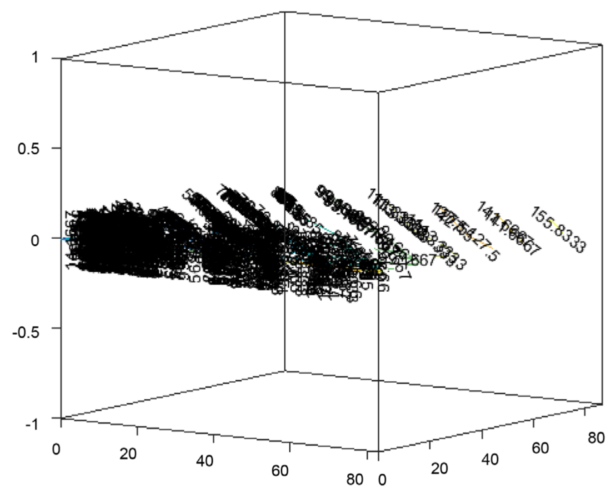

Figure 2. Plotting dataset in 3D.

Table 2. The verbs in line fit.

Index	Words in Arabic	Words in English	Translate in English	Frequency (X_1, X_2, X_3)
18	شكرون	Shakurun	(Be) grateful	(9, 19, 1)
27	ملكون	Malikun	(Are the) owners	(10 , 2, 1)
38	خافون	Kafun	They are afraid	(11, 4, 1)
69	شهدون	Shahidun	(Were) witnesses	(4, 4, 1)
74	جعلون	Jaelun	They made	(5, 2, 1)
86	دخلون	Dakhilun	Enter (it)	(6, 0, 1)
99	كتبون	Katabun	(Are) recorders	(5, 0, 1)
107	جهلون	Jahilun	Ignorance	(1, 4, 1)
140	حذرون	Hadharun	Forewarned	(2, 1, 1)
171	سلمون	Salamun	(Were) sound	(1, 1, 1)
174	غلبون	Ghalibun	Victorious	(1, 1, 1)
213	كيدون	Kidun	Scheme against me	(1, 0, 1)
214	ركعون	Rakaeun	(are) those who bow down	(1, 0, 1)
216	طوفون	Tufun	(As) moving about	(1, 0, 1)
217	خصمون	Khsimun	Argumentative	(1, 0, 1)
230	قومون	Kuamun	(Are) protectors	(1, 0, 1)
236	فكهون	Fakahun	(In) amusement	(0, 1, 1)

4. Conclusions and Future Work

The present dataset in this paper fined in Quran gives a good simple and multi linear regression model between the frequency of verbs with a form (—un, ون---) made by the frequency of plural present verbs (t—un, ت---ون) or (y—un, ي---ون) and there inverses.

The results show that the parameters of the model are ones vector (1,1) and mean of dataset is (6, 7). It corresponds to the verb point enter (yadkolun دخلون، يدخلون dakilun), also other 17 points exist in the line and in the dataset of 387 verbs and their derivate verbs in Quran. The name of Allah (ﷲ) showed when we use tree variables and plot it in 3D with option "Show Text".

For future work, the estimation parameter of this dataset will be done by using ℓ_1 norm and sub-gradient method, and comparing this model in other drawing of the Quran.

Acknowledgements

The author thanks Allah for this miracle dataset in Quran. Also we are indebted to the anonymous reviewers and editors of AJCM for many suggestions and stimulating comments to improve the original manuscript.

References

[1] Al-Faqih, K.M. (2017) A Mathematical Phenomenon in the Quran of Earth-Shattering

Proportions: A Quranic Theory Based on Gematria Determining Quran Primary Statistics (Words, Verses, Chapters) and Revealing Its Fascinating Connection with the Golden Ratio. *Journal of Arts and Humanities*, **6**, 52-73. https://doi.org/10.18533/journal.v6i6.1192

[2] Al-Kaheel, A. (2017) http://www.kaheel7.com/ar/index.php/1/1690-2014-07-03-19-11-02

[3] El Mouatasim, A. and Al-Hossain, A. (2009) Reduced Gradient Method for Minimax Estimation of a Bounded Poisson Mean. *Journal of Statistics*: *Advances in Theory and Applications*, **2**, 183-197.

[4] El Mouatasim, A. and Wakrim, M. (2015) Control Subgradient Algorithm for Image ℓ_1 Regularization. *Journal of Signal*, *Image and Video Processing* (*SViP*), **9**, 275-283. https://doi.org/10.1007/s11760-015-0815-z

[5] Smadi, A.A. and Abu-Afouna, N.H. (2012) On Least Squares Estimation in a Simple Linear Regression Model with Periodically Correlated Errors: A Cautionary Note. *Austrian Journal of Statistics*, **41**, 211-226. https://doi.org/10.17713/ajs.v41i3.175

[6] Uyanik, G.K. (2013) A Study on Multiple Linear Regression Analysis. *Procedia-Social and Behavioral Sciences*, **106**, 234-240. https://doi.org/10.1016/j.sbspro.2013.12.027

[7] Rawlings, J.O., Pantula, S.G. and Dickey, D.A. (1998) Applied Regression Analysis: A Research Tool. 2nd Edition, Springer, Berlin. https://doi.org/10.1007/b98890

[8] Al-Kaheel, A. (2006) Ishraqat al-Raqam Sab`a fial-Qur'an al-Karim. Dubai International Holy Quran Award, Dubai.

[9] Al-Hakkak, G. (2016) Arabic Verbs. Arabic Online for English Speakers.

[10] https://www.almaany.com/ar/dict/ar-en/

[11] https://translate.google.com/