Article

Regression modeling of different proteins using linear and multiple analysis

Shruti Jain
Department of Electronics and Communication Engineering, Jaypee University of Information Technology, Solan-173234, India
E-mail: jain.shruti15@gmail.com

Received 1 September 2017; *Accepted 25 September 2017*; *Published 1 December 2017*
(cc) BY

Abstract
There are different types of regression analysis. Out of which simple regression and multiple regressions was considered in this paper. For calculation purpose we have used PLS analysis which calculates squared *r* values. This paper considers eleven different proteins and one output. We have validated our results by calculating adjusted regression coefficient, predicted regression coefficient regression coefficient cross validation, rm^2 and *F*-test values. Later multiple regressions were used as we have different independent variable (proteins). For that analysis we have calculated the coefficient, standard error, standard coefficient, tolerance, *t* value and *p* value, variation explanation of predictors and estimators which gives percentage and cumulative percentage. Correlation matrixes were also shown at the end for eleven proteins and one output.

Keywords linear regression analysis; multiple regression analysis; marker proteins; PLS.

Network Biology
ISSN 2220-8879
URL: http://www.iaees.org/publications/journals/nb/online-version.asp
RSS: http://www.iaees.org/publications/journals/nb/rss.xml
E-mail: networkbiology@iaees.org
Editor-in-Chief: WenJun Zhang
Publisher: International Academy of Ecology and Environmental Sciences

1 Introduction

Regression analysis (RA) is a statistical method for investigating relationship between one variable and other variables (Farahani, 2010; Ringle, 2010). A statistical model is a simple description of state. There are three types of RA: linear regression (LR), multiple linear regression (MLR) and non linear regression (NLR). If we have to model the linear relationship between dependent and independent variables than LR was used but if there are more than one independent variable and one dependent variable than MLR is used. The MLR considers co linearity, variance inflation, graphical display of regression diagnosis, and detection of regression outlier and influential observation. In NLR the variables (dependent and independent) are not linear. NLR can be written as

$$y = \frac{\alpha}{1+e^{\beta t}} + \varepsilon \tag{1}$$

where *y* is the growth of a particular organism as a function of time *t*, α and β are model parameters, and ε is

the random error. NLR model is more complicated than LR model in terms of estimation of model parameters, model selection, model diagnosis, variable selection, outlier detection, or influential observation identification. In this paper we have calculated regression coefficient or regression coefficient cross validation (r^2 or q^2_{cv}), adjusted regression coefficient (r^2_{adj}), predicted regression coefficient (r^2_{pre}), regression coefficient without intercept (r^2_0) for ten different concentration (Jain, 2012a; Suzzane, 2005; Weiss, 2001) of three input proteins TNF (Thoma, 1990; Jain, 2009a, 2009b, 2010a, 2011a), EGF (Janes, 2005; Normano, 2006; Jain, 2014, 2015a, 2016a, 2017) and Insulin (Jain, 2010b, 2010c, 2011b, 2012b, 2012c; Morris, 2003). For validation of our results we have calculated rm^2 and *F*-test values. Different plots were plotted which are showing r^2 values. Later in paper we have shown the results of multiple regression. We have different marker proteins: AkT (Coffer, 1998; Hemmings, 1997; Bruent, 1999; Jain, 2010d, 2012d, 2015b, 2017b), MK2 (Jain, 2011c, 2016b, 2016c), JNK (Jain, 2010e, 2015b, 2015c), FKHR (Jain, 2015c, 2011d), MEK, ERK, IRS, IKK, pAkT, ptAkT and EGFR for the HT carcinoma cells which helps in cell survival/ apoptosis. If these proteins are present in the pathway than it leads to cell survival otherwise cell death (Jain, 2009c).

2 Material and Methods

There are different types of regression analysis. In this paper we are working on Linear Regression/ Simple regression (LR) and Multiple regression analysis (MR). For calculation of LR there are two techniques: Ordinary least square method (OLS) and Partial least square method (PLS). PLS is a technique which is helpful in predictive models when the factors are many and highly collinear. PLS approach is beneficial for relating dependent variables to many independent variables.

LR as the name suggests the shape of regression line is linear whose intercept is *a* and slope of line is *b*. For LR the dependent variable (*Y*) is continuous while independent variable (*X*) can be continuous or discrete. We can consider the error term '*e* or ε' . LR is expressed as:

$$\underbrace{Y}_{actual(observed)} = \underbrace{aX+b}_{explained(predicted)} + \underbrace{\varepsilon}_{error} \quad (2)$$

Equation 1 is also known as linear population regression model, or linear population regression. For LR error ε is normally distributed with $E(\varepsilon) = 0$ and a constant variance $\mathrm{Var}(\varepsilon) = \sigma^2$. LR can be also be represented as :

$$\underbrace{Y}_{observed} = \underbrace{\hat{Y}}_{predicted/estimator} + \underbrace{\hat{\varepsilon}}_{error/residual} \quad (3)$$

Predicted value is also known as conditional mean. For predicted values equation 2 can be written as:

$$\hat{Y} = \hat{a}X + \hat{b} \quad (4)$$

Equation 4 is known as sample regression function (SRF) where intercept is represented by equation 5.

$$\hat{b} = \bar{Y} - \hat{a}\bar{X} \quad (5)$$

$\bar{X}$ and $\bar{Y}$ are the sample means of *X* and *Y*. Slope is represented as:

$$\hat{a}=\frac{SS_{XY}}{SS_{XX}}=\frac{\sum\left(X-\bar{X}\right)\left(Y-\bar{Y}\right)}{\sum\left(X-\bar{X}\right)^2}=\frac{\sum x\,y}{\sum x^2}=\frac{Cov\left(X,Y\right)}{Var\left(X\right)} \tag{6}$$

where

$$Cov(X,Y)=Var(X,Y)=\frac{\sum\left(X-\bar{X}\right)\left(Y-\bar{Y}\right)}{n-1}$$

and

$$Var(X)=\frac{\sum\left(X-\bar{X}\right)^2}{n-1}, \quad Var(Y)=\frac{\sum\left(Y-\bar{Y}\right)^2}{n-1}$$

equation 6 can be written as

$$\hat{a}=\frac{\sum xY}{\sum X^2-n\bar{X}^2}=\frac{\sum X\,y}{\sum X^2-n\bar{X}^2} \tag{7}$$

We can define $x=\left(X-\bar{X}\right)$ and $y=\left(Y-\bar{Y}\right)$ where lower case letters denote deviations from mean values.

OLS method and PLS is used for calculation of LR. OLS minimizes the SS of the vertical deviations from each data point to the line while in PLS, initially the values are squared, then added up so as there is no cancellation of positive and negative terms. Finally, the minimum (least) square value is considered.
We have three different types of sum of squares (SS):

a. regression sum of squares (SSreg) / explained SS which is a measure of explained variation,

$$SS_{reg}=\sum(\hat{y}-\bar{y})^2 \tag{8}$$

b. residual sum of squares or error sum of squares (SS_{err}) / unexplained SS which is a measure of unexplained variation and

$$SS_{err}=SS_{residual}=\sum(y-\hat{y})^2=\sum\hat{\varepsilon}^2 \tag{9}$$

c. total sum of squares (SS_{total}) which is a measure of total variation.

$$SS_{total}=SS_{reg}+SS_{err} \tag{10}$$

$$SS_{total}=\sum(y-\bar{y})^2 \tag{11}$$

The ratio of SS_{reg} to SS_{total} is known as coefficient of determination (r^2) is expressed as:

$$r^2=\frac{SS_{reg}}{SS_{total}}=\frac{SS_{reg}}{SS_{reg}+SS_{err}}=1-\frac{SS_{err}}{SS_{total}} \tag{12}$$

For deviation form, the SRF can be written as:

$$= \frac{\sum \hat{y}^2}{\sum y^2} \tag{13}$$

$$= \frac{\hat{a}^2 \sum x^2}{\sum y^2} \tag{14}$$

$$= \hat{a}^2 \left(\frac{\sum x^2}{\sum y^2} \right) \tag{15}$$

If numerator and denominator are divided by sample size '*n*' than r^2 is expressed as:

$$r^2 = \hat{a}^2 \left(\frac{S_x^2}{S_y^2} \right) = a^2 \left(\frac{Var\ X}{Var\ Y} \right) \tag{16}$$

where S_x^2 and S_y^2 is a sample variance of *X* and *Y* respectively.

Replace the value of $\hat{a}$ in equation 15 by equation 6; we get

$$r^2 = \frac{\left(\sum x\, y \right)^2}{\sum x^2 \sum y^2} \tag{17}$$

Or

$$r = \frac{\left(\sum x\, y \right)}{\sqrt{\sum x^2 \sum y^2}} = \frac{Cov\left(X, Y \right)}{\sqrt{Var\,(X).Var\left(Y \right)}} \tag{18}$$

In above equation; *r* is known as sample/linear correlation coefficient. Equation 12 can be written as

$SS_{reg} = r^2\, SS_{total}$ where

$$SS_{total} = \sum y^2 \tag{19}$$

Equation 10 can be written as

$$SS_{reg} = r^2 \sum y^2 \tag{20}$$

$$SS_{err} = SS_{total} - SS_{reg}$$

Placing values from eq 19 and eq 20 to equation 10 we get:

$$SS_{err} = \sum y^2 \left(1 - r^2 \right) \tag{21}$$

Finally, placing all the values i.e. from eq 19 , eq 20 and eq 21 in equation 120 we get

$$\sum y^2 = r^2 \sum y^2 + \sum y^2 \left(1 - r^2 \right) \tag{22}$$

The r^2 lies in the range of 0 and 1, greater the value of r^2 more accurate the model. If the value of r^2 is 1 means

a perfect fit on the other hand if r^2 value is zero it means that there is no relationship between regress and regressor. r^2 is a measure of goodness of fit which tells how close the estimate values are to their actual values. r^2 can also be calculated as the squared coefficient of correlation between actual *Y*, estimated *Y* i.e. $\hat{Y}$ and is expressed as

$$r^2 = \frac{\sum\left(Y-\overline{Y}\right)^2\left(\hat{Y}-\overline{Y}\right)^2}{\sum\left(Y-\overline{Y}\right)^2\sum\left(\hat{Y}-\overline{Y}\right)^2} \tag{23}$$

$$= \frac{\left(\sum y\,\hat{y}\right)^2}{\sum y^2\sum \hat{y}^2} \tag{24}$$

Multiple Regression (MR): If the *x* parameters are more than one i.e. x_1, x_2…. than the regression analysis is known as multiple regression (MR) but if the *x* parameter is one than it is LR. MR equation can be represented as:

$$y = a + b_1x_1 + b_2x_2 + b_3x_3 + \ldots b_nx_n + e \tag{25}$$

For MR it is usually assumed that the error term ε follows the normal distribution with E(ε) = 0 and a constant variance Var(ε) = σ^2 . Forward selection (FS), backward elimination (BE) and step wise approximation (SWA) is used for analysis of MR.

3 Validation Tests

In this paper we are considering LR and MR methods for which we have discussed how to calculate r^2 values. To validate the results we have different approaches. In this paper we are using r^2_{adj}, r^2_{pre}, q^2_{cv}, rm^2 and *F*-test values. The predictive capability of the equation is determined using the leave-one-out cross validation method. q^2_{cv} was calculated by the following equation:

$$q^2_{cv} = 1 - \frac{PRESS}{TOTAL} \approx r^2 \tag{26}$$

For a perfect model value of q^2_{cv} should be close to one and its value is approximately equal to r^2.

The evaluation of the predictive ability of the model for the external test set compounds was done by determining the value of rm^2 which was determine by :

$$rm^2 = r^2\left(1-\left|\sqrt{r^2-r_0^2}\right|\right) \tag{27}$$

r^2_0 is the squared correlation coefficient for regression without using intercept and the regression equation becomes $y = ax$. For a perfect model value of rm^2 should be close to one.

4 Discussion

In this paper we have considered eleven different proteins MK2, JNK, FKHR, MEK, ERK, IRS, AkT, IKK,

pAkT, ptAkT and EGFR for the HT carcinoma cells which occurs due to the combination of TNF, EGF and Insulin. These proteins yield four different output: phosphatidylserine exposure (PE), membrane permeability (MP), nuclear fragmentation (NF) and caspase substrate cleavage (CCK). We have first taken average of all outputs and then normalized the output to maximum that's why in results we are showing only one output. Table 1 shows the minimum, maximum, median, mean, standard deviation, variance, and coefficient of variance for eleven different proteins and output.

Table 1 Different values for eleven different proteins and output.

	MK2	**JNK**	**FKHR**	**MEK**	**ERK**	**IRS**	**AKT**	**IKK**	**pAKT**	**pTAKT**	**EGFR**	**OUTPUT**
N of Cases	300	300	300	300	300	300	300	300	300	300	300	300
Minimum	221.466	221.526	191.019	110.233	140.711	191.123	164.909	59.976	110.359	90.321	49.437	0.407
Maximum	262.76	255.651	245.391	189.092	189.29	244.545	188.985	90.748	201.321	133.322	78.153	0.593
Median	242.156	235.749	210.593	136.447	172.221	216.314	177.376	76.804	157.058	108.792	68.604	0.48
Arithmetic Mean	242.798	236.224	214.208	144.297	167.113	216.768	177.868	75.085	154.684	110.499	64.849	0.483
Standard Deviation	10.541	8.302	14.94	26.971	14.425	14.304	6.149	9.22	30.922	13.876	9.505	0.051
Variance	111.115	68.923	223.217	727.422	208.083	204.606	37.808	85.006	956.141	192.552	90.348	0.003
Coefficient of Variation	0.043	0.035	0.07	0.187	0.086	0.066	0.035	0.123	0.2	0.126	0.147	0.106

In this section we have shown all the results r^2, r^2_{adj}, q^2_{cv}, rm^2 and F-test values for our 10 data sets of TNF/ EGF and Insulin in ng/ml.

1. **For r^2**: We have observed values of our 10 data set. First we have predicted the values using STAT SOFT software. We have put all results in excel and get the r^2 value from their and then calculate the same r^2 value using formula as shown in Eq 1. We have also calculated the r^2 value using MINITAB software. All the results of r^2 are shown in Table 2 which proves that all the values are same.

Table 2 All possible values of r^2 using excel, formula and software.

S. No	**Possible Values**	**r^2 from excel**	**r^2 from formula using equation 1**	**r^2 from software**
1	0-0-0	0.984	0.984	0.985
2	5-0-0	0.991	0.991	0.991
3	100-0-0	0.966	0.966	0.966
4	0-100-0	0.765	0.8265	0.765
5	5-1-0	0.953	0.9589	0.953
6	100-100-0	0.981	0.9817	0.981
7	0-0-500	0.991	0.992	0.991
8	0.2-0-1	0.985	0.985	0.985
9	5-0-5	0.9916	0.992	0.992
10	100-0-500	0.9105	0.9116	0.911

2. **For r^2_{pred}, r^2_{adj}**: We have calculated the r^2_{pred}, r^2_{adj} using MINITAB software shown in Table 3 which is coming out to be OK. Table 4 shows the cumulative r^2 values for *X* and *Y*. this table also represents the Eigen values and Q^2 cumulative values.

Table 3 Values of r^2, r^2_{pred}, r^2_{adj}.

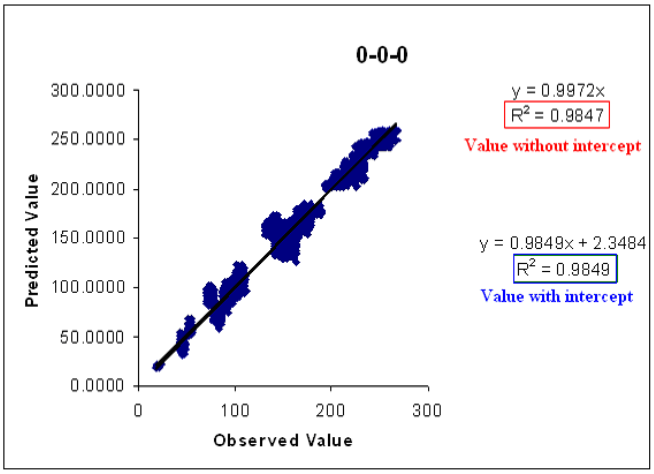
S. No	Possible Values	r^2 (%)	r^2_{pred} (%)	r^2_{adj} (%)
1	0-0-0	98.5	98.5	98.49
2	5-0-0	99.1	99.13	99.1
3	100-0-0	96.6	96.59	96.6
4	0-100-0	76.5	76.44	76.5
5	5-1-0	95.3	95.32	95.3
6	100-100-0	98.1	98.10	98.1
7	0-0-500	99.1	99.10	99.1
8	0.2-0-1	98.5	98.49	98.5
9	5-0-5	99.2	99.16	99.2
10	100-0-500	91.1	91.05	91.1

Table 4 Eigen values and cumulative Q2 values of ten different concentrations.

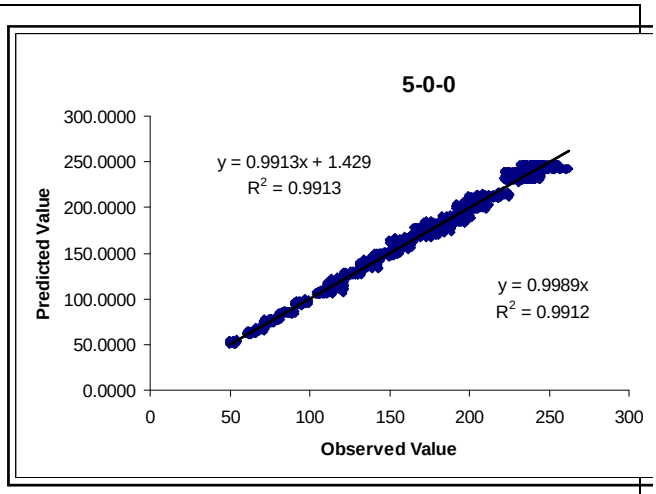
S. No	Possible Values	r^2 X (Cumul.)	Eigen values	r^2 Y (Cumul.)	Q^2 (Cumul.)	Iterations
1	0-0-0	0.661527	6.778547	0.642263	0.530556	3
2	5-0-0	0.706644	7.770976	0.664049	0.662104	4
3	100-0-0	0.716885	7.654734	0.70051	0.648019	6
4	0-100-0	0.643378	6.55393	0.589539	0.48238	5
5	05-01-2000	0.449319	4.906164	0.515036	0.460442	49
6	100-100-0	0.704479	7.743815	0.663912	0.661397	4
7	0-0-500	0.794276	8.711558	0.517066	0.512291	4
8	0.2-0-1	0.76338	8.297655	0.704509	0.685306	3
9	5-0-5	0.806147	8.854371	0.482898	0.480075	4
10	100-0-500	0.658964	7.232619	0.6151	0.609608	9

3. **For q^2_{cv}**: As we know for a perfect model value of q^2_{cv} should be close to one and its value should be equal to r^2. We have already calculated the value of r^2. q^2_{cv} is calculated from MINITAB software both the values are equal. So it proves that the value of q^2_{cv} is close to r^2.

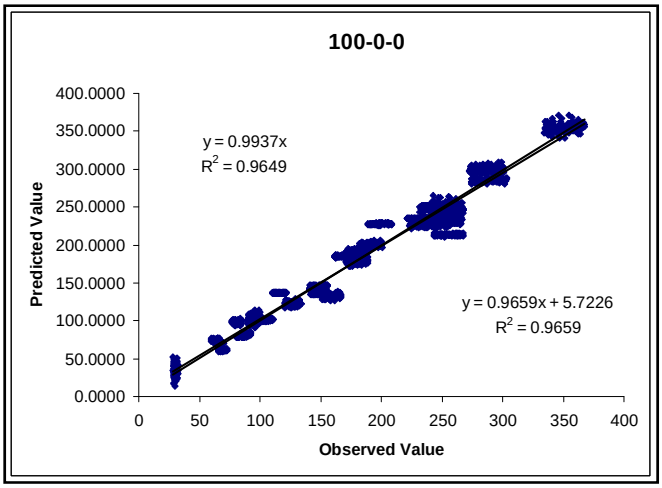
4. **For rm^2**: Fig. 1 shows the r^2 and r^2_0 value for ten different concentrations. For r^2 and r^2_0 value we have plotted the observed value and predicted value in Excel and get the equations from the same. Similarly we have plotted the data in MINITAB software and get the same equation which verifies the result as shown in Table 5. Putting the values of r^2 and r^2_0 in Eq. 27 we get the rm^2 value which is coming close to one as shown in Table 6 which verifies our result.


0-0-0
y = 0.9972x
R² = 0.9847
Value without intercept
y = 0.9849x + 2.3484
R² = 0.9849
Value with intercept
Predicted Value
Observed Value
300.0000
250.0000
200.0000
150.0000
100.0000
50.0000
0.0000
0
100
200
300

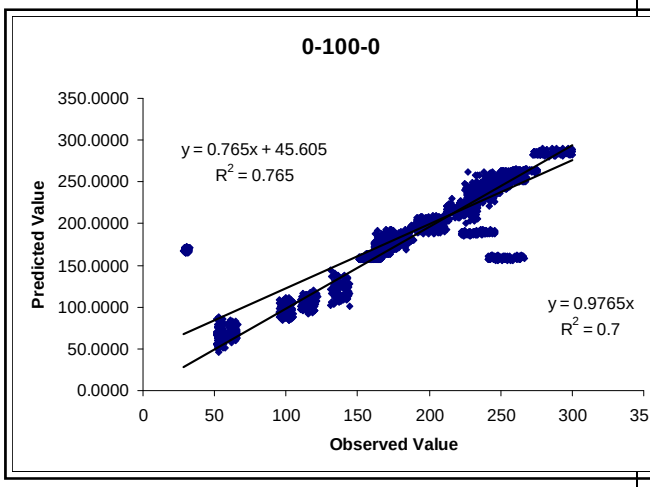
(*a*)


5-0-0
y = 0.9913x + 1.429
R² = 0.9913
y = 0.9989x
R² = 0.9912
Predicted Value
Observed Value
300.0000
250.0000
200.0000
150.0000
100.0000
50.0000
0.0000
0
50
100
150
200
250
300

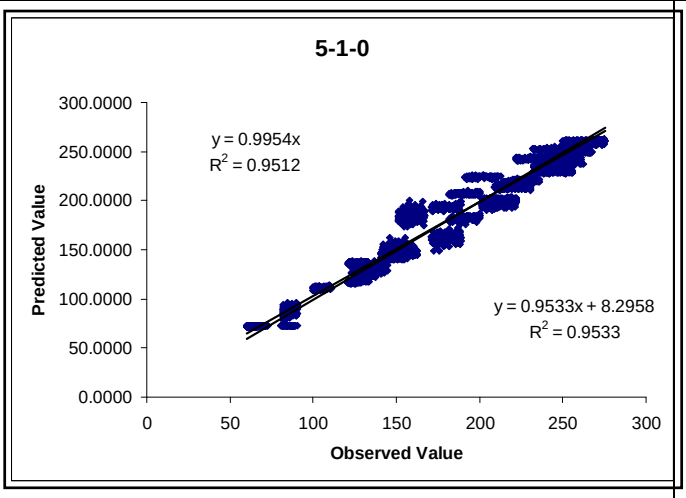
(*b*)


100-0-0
y = 0.9937x
R² = 0.9649
y = 0.9659x + 5.7226
R² = 0.9659
Predicted Value
Observed Value
400.0000
350.0000
300.0000
250.0000
200.0000
150.0000
100.0000
50.0000
0.0000
0
50
100
150
200
250
300
350
400

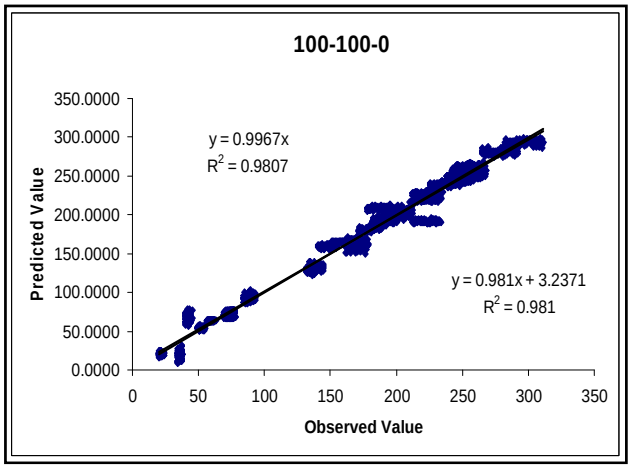
(*c*)


0-100-0
y = 0.765x + 45.605
R² = 0.765
y = 0.9765x
R² = 0.7
Predicted Value
Observed Value
350.0000
300.0000
250.0000
200.0000
150.0000
100.0000
50.0000
0.0000
0
50
100
150
200
250
300

(*d*)


5-1-0
y = 0.9954x
R² = 0.9512
y = 0.9533x + 8.2958
R² = 0.9533
Predicted Value
Observed Value
300.0000
250.0000
200.0000
150.0000
100.0000
50.0000
0.0000
0
50
100
150
200
250
300

(e)


100-100-0
y = 0.9967x
R² = 0.9807
y = 0.981x + 3.2371
R² = 0.981
Predicted Value
Observed Value
350.0000
300.0000
250.0000
200.0000
150.0000
100.0000
50.0000
0.0000
0
50
100
150
200
250
300
350

(f)

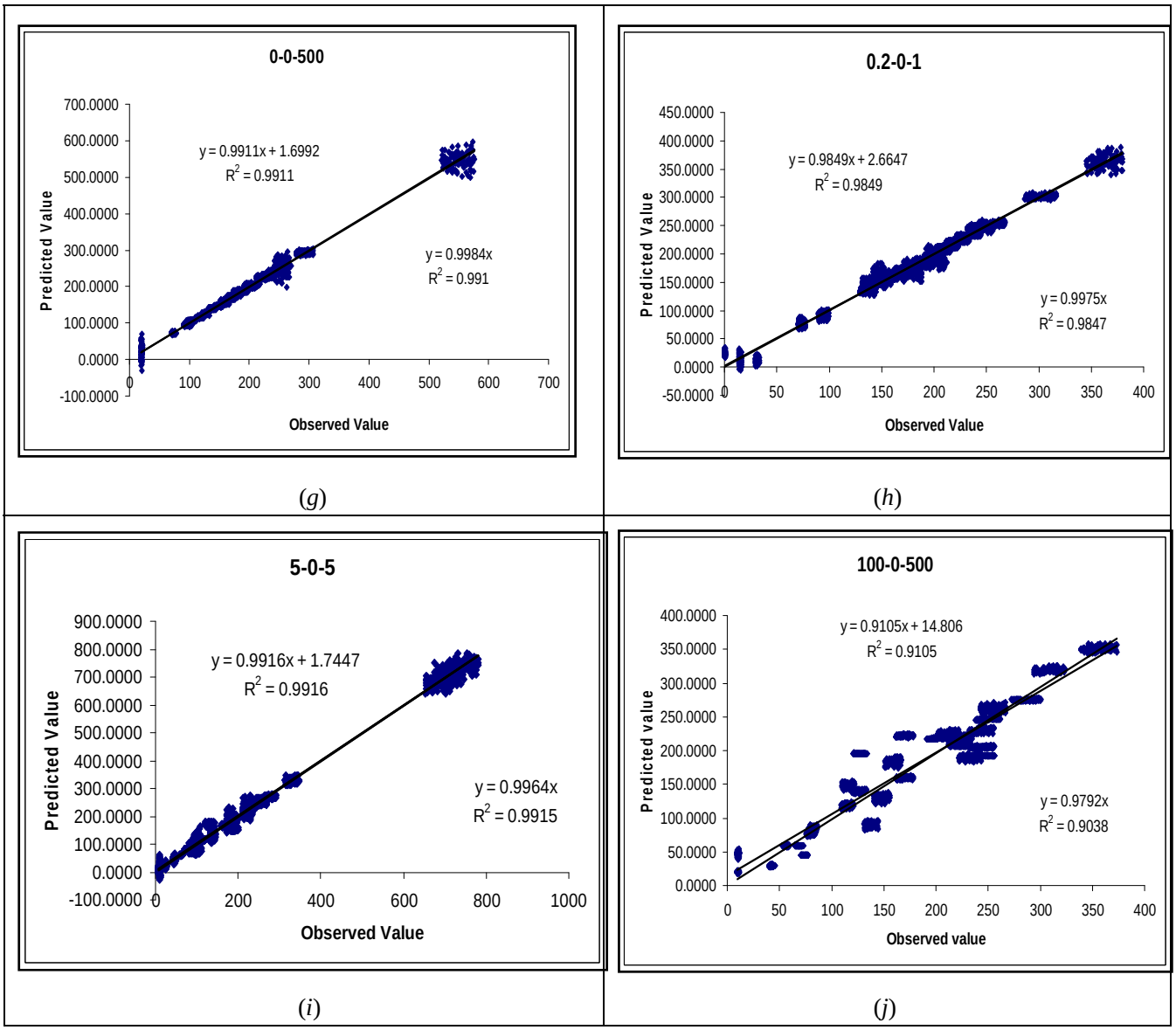


Fig. 1 Values for r^2 and r^2_0 for 10 data sets.

Table 5 Equations of r^2 and r^2_0.

S. No	Possible Values	r^2 with intercept From excel	r^2 with intercept From MINITAB	r^2_0 without intercept
1	0-0-0	0.9849*x* + 2.3484	0.985*x* + 2.35	0.9972*x*
2	5-0-0	0.9913*x* +1.429	0.991*x* +1.43	0.9989*x*
3	100-0-0	0.9659*x* + 5.7226	0.966*x* + 5.72	0.9937*x*
4	0-100-0	0.765*x* + 45.605	0.765*x* + 45.6	0.9765*x*
5	5-1-0	0.9533*x* + 8.2958	0.953*x* + 8.30	0.9954*x*
6	100-100-0	0.981*x* + 3.2371	0.981*x* + 3.24	0.9967*x*
7	0-0-500	0.9911*x* + 1.6992	0.991*x* + 1.7	0.9984*x*
8	0.2-0-1	0.9849*x* + 2.6647	0.985*x* + 2.66	0.9975*x*
9	5-0-5	0.9916*x* + 1.7447	0.992*x* + 1.74	0.9964*x*
10	100-0-500	0.9105*x* + 14.806	0.911*x* + 14.8	0.9792*x*

Table 6 Values of rm^2.

S. No	Possible Values	r^2	r^2_0	rm^2 (close to 1)	$(r^2 - r^2_0) / r^2$ (less than 0.1)
1	0-0-0	0.984	0.9847	0.970971	0.0002
2	5-0-0	0.991	0.9912	0.981387	0.0001
3	100-0-0	0.966	0.9649	0.935356	0.00104
4	0-100-0	0.765	0.7	0.569963	0.08497
5	5-1-0	0.953	0.9512	0.909614	0.0022
6	100-100-0	0.981	0.9807	0.964009	0.00031
7	0-0-500	0.991	0.991	0.981189	0.0001
8	0.2-0-1	0.985	0.9847	0.970971	0.0002
9	5-0-5	0.9916	0.9915	0.981684	0.0001
10	100-0-500	0.9105	0.9038	0.835972	0.00736

5. **For *F*-value**: With the help of observed and predicted values we have calculated *F*-value shown in Table 7 which is coming out to be very large.

Table 7 Values of *F*-value.

S. No	Possible Values	PRESS	*F*-value
1	0-0-0	273446	215195.4
2	5-0-0	106089	375401.5
3	100-0-0	698648	93465.53
4	0-100-0	2484216	10735.86
5	5-1-0	505386	97249.82
6	100-100-0	382434	170650
7	0-0-500	235785	365494
8	0.2-0-1	296709	215009.8
9	5-0-5	896652	388299.1
10	100-0-500	2231417	33567.88

As the independent variables are many so MR can be applied. Table 8 shows the coefficient, standard coefficient, *t* value and *p* values of the independent variables. Table 9 shows the variation explanation for predictors and responses of eleven proteins and output. Table 10 shows the correlation matrixes for eleven different proteins and one output.

Table 8 Different parameters for MR.

Predictor	Coeff	SE coeff	*t*	*p*
Constant	0.2225	0.187	1.19	0.235
MK2	0.001157	0.000186	6.21	0
JNK	-0.00046	0.000222	-2.07	0.039

FKHR	-0.0007	0.000232	-3.03	0.003
MEK	-0.00045	0.000296	-1.5	0.133
ERK	0.001518	0.000306	4.97	0
IRS	0.000117	0.000249	0.47	0.639
AkT	-0.00124	0.00027	-4.6	0
IKK	-0.00033	0.000624	-0.53	0.597
pAkT	-0.00122	0.000266	-4.58	0
ptAkT	0.001206	0.000423	2.85	0.005
EGFR	0.005013	0.000674	7.44	0

Table 9 Variation explained for predictors and responses of eleven proteins.

Factors	**Variation Explained for**		**Variation Explained for**	
	Predictor(s)		**Response(s)**	
	Percentage	**Cum. Percentage**	**Percentage**	**Cum. Percentage**
MK2	68.938	68.938	45.166	45.166
JNK	14.369	83.307	22.41	67.576
FKHR	4.333	87.64	3.151	70.727
MEK	2.077	89.717	3.485	74.212
ERK	4.783	94.5	0.648	74.86
IRS	2.176	96.675	0.639	75.499
AkT	0.797	97.473	0.221	75.72
IKK	0.971	98.443	0.042	75.763
pAkT	0.738	99.181	0.017	75.779
ptAkT	0.287	99.468	0.002	75.781
Output	0.532	100	0	75.781

Table 10 Correlation matrixes.

	MK2	JNK	FKHR	MEK	ERK	IRS	AKT	IKK	PAKT	PTAKT	EGFR	OUTPUT
MK2	1											
JNK	-0.026	1										
FKHR	0.226	0.511	1									
MEK	0.263	0.538	0.895	1								
ERK	-0.223	-0.533	-0.85	-0.934	1							
IRS	0.366	0.442	0.8	0.876	-0.831	1						
AKT	-0.313	0.163	0.108	0.105	-0.135	0.029	1					
IKK	-0.295	-0.515	-0.875	-0.955	0.909	-0.888	-0.08	1				
PAKT	-0.377	-0.517	-0.882	-0.966	0.917	-0.899	-0.043	0.954	1			
PTAKT	0.364	0.507	0.865	0.947	-0.903	0.884	0.035	-0.938	-0.962	1		
EGFR	-0.218	-0.564	-0.889	-0.963	0.925	-0.865	-0.162	0.937	0.945	-0.929	1	
OUTPUT	0.306	-0.497	-0.596	-0.6	0.637	-0.444	-0.442	0.546	0.497	-0.487	0.666	1

5 Conclusion

We have made a best fit linear model using partial least square for ten cytokine combinations of TNF, EGF and Insulin. In this we have find all the results like regression coefficient, adjusted regression coefficient, regression coefficient cross validation, rm^2 and F-test values for our 10 data sets which comes out to be correct. Later multiple regressions were applied as we have eleven different input independent variables (proteins). We have calculated coefficient, standard error, standard coefficient, tolerance, t value and p value, variation explanation of predictors and estimators which gives percentage and cumulative percentage of all eleven proteins and one output. Later, Correlation matrixes were also for eleven proteins and one output.

References

Brunet A, Bonni A, Zigmond MJ, Lin MZ, Juo P, Hu LS. 1999. Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor. Cell, 96: 857-868

Coffer PJ, Jin J, Woodgett JR. 1998. Protein kinase B (c-Akt): a multifunctional mediator of phosphatidylinositol 3-kinase activation. Biochemical Journal, 335: 1-13

Farahani H.A., Rahiminezhad A, Samec L, Immannezhad K. 2010. A comparison of Partial Least Squares (PLS) and Ordinary Least Squares (OLS) regressions in predicting of couples mental health based on their communicational patterns. Procedia Social and Behavioral Sciences, 5: 1459-1463

Hemmings BA. 1997. Akt signaling: linking membrane events to life and death decisions. Science, 275: 628-630

Jain S. 2012a. Communication of Signals and Responses Leading To Cell Survival/Cell Death Using Engineered Regulatory Networks. PhD Thesis. Jaypee University of Information Technology, Solan, Himachal Pradesh, India

Jain S. 2014. Implementation of fuzzy system using operational transconductance amplifier for ERK pathway of EGF/ Insulin leading to cell survival/ death. Journal of Pharmaceutical and Biomedical Sciences, 4(8): 701-707

Jain S. 2015d. Mathematical Analysis and Probability Density Function of FKHR pathway for Cell Survival /Death. Control System and Power Electronics – CSPE 2015. 84-93, Banglore, India

Jain S. 2016a. Compedium model using frequency / cumulative distribution function for receptors of survival proteins: Epidermal growth factor and insulin, Network Biology, 6(4): 101-110

Jain S. 2016b, Regression analysis on different mitogenic pathways. Network Biology, 6(2): 40-46

Jain S. 2016c. Mathematical analysis using frequency and cumulative distribution functions for mitogenic pathway. Research Journal of Pharmaceutical, Biological and Chemical Sciences, 7(3): 262-272

Jain S. 2017a. Implementation of marker proteins using standardised effect. Journal of Global Pharma Technology, 9(5): 22-27

Jain S. 2017b. Parametric and non parametric distribution analysis of AkT for cell survival/death. International Journal of Artificial Intelligence and Soft Computing, 6(1): 43-55

Jain S, Bhooshan SV, Naik PK. 2010a. A system model for cell death/ survival using SPICE and ladder logic. Digest Journal of Nanomaterials and Biostructures, 5(1): 57-66

Jain S, Bhooshan SV, Naik PK. 2010e. Model of mitogen activated protein kinases for cell survival/death and its equivalent bio-circuit. Current Research Journal of Biological Sciences, 2(1): 59-71

Jain S, Bhooshan SV, Naik PK. 2011a. Mathematical modeling deciphering balance between cell survival and cell death using Tumor Necrosis Factor α. Research Journal of Pharmaceutical, Biological and Chemical Sciences, 2(3): 574-583

Jain S, Bhooshan SV, Naik PK. 2011b. Mathematical modeling deciphering balance between cell survival and cell death using insulin. Network Biology, 1(1): 46-58

Jain S, Bhooshan SV, Naik PK. 2011c. Model of Protein Kinase B for Cell Survival/Death and its Equivalent Bio Circuit. In: Proceedings of the 2nd International Conference on Methods and Models in Science and Technology (ICM2ST-11). 69-73, Institution of Engineers, Technocrats and Academician Network (IETAN), Jaipur, Rajasthan, India

Jain S, Bhooshan SV, Naik PK. 2012d. Compendium model of AkT for cell survival/death and its equivalent BioCircuit. International Journal of Soft Computing and Engineering (IJSCE), 2(3): 91-97

Jain S, Chauhan DS. 2015a. Mathematical analysis of receptors for survival proteins. International Journal of Pharma and Bio Sciences, 6(3): 164-176

Jain S, Chauhan DS. 2015b. Linear and Non Linear Modeling of Protein Kinase B/ AkT. International Conference on Information and Communication Technology for Sustainable Development (ICT4SD - 2015). 81-88, Ahmedabad, India

Jain S, Chauhan DS. 2015c. Implementation of fuzzy system using different voltages of OTA for JNK pathway leading to cell survival/ death. Network Biology, 5(2): 62-70

Jain S, Naik PK, Sharma R 2009a. A computational modeling of cell survival/death using VHDL and MATLAB simulator. Digest Journal of Nanomaterials and Biostructures (DJNB), 4(4): 863-879

Jain S, Naik PK, Sharma R. 2009b. A Computational modeling of apoptosis signaling using VHDL and MATLAB simulator. International Journal of Information and Communication Technology, 2(1-2): 7-12

Jain S, Naik PK, Chauhan DS, Sharma R. 2009c. Computational Modeling of Cell Survival/ Death Using MATLAB Simulator. 6th International Conference on Information Technology: New Generations ITNG. 64-68, Las Vegas, Nevada, USA

Jain S, Naik PK. 2010b. Computational modeling of cell survival using VHDL, BVICAM's. International Journal of Information Technology (BIJIT), 2(1): 47-51

Jain S, Naik PK, Bhooshan SV. 2010c. Computational modeling of cell survival/death using BiCMOS. International Journal of Computer Theory and Engineering, 2(4): 478-481

Jain S, Naik PK, Bhooshan SV, 2010d. Petri net implementation of cell signaling for cell death. International Journal of Pharma and Bio Sciences, 1(2): 1-18

Jain S, Naik PK, Bhooshan SV. 2011d. Nonlinear Modeling of cell survival/ death using artificial neural network. In: The Proceedings of International Conference on Computational Intelligence and Communication Networks. 565-568, Gwalior, India

Jain S, Naik PK. 2012b. System modeling of cell survival and cell death: A deterministic model using Fuzzy System. International Journal of Pharma and BioSciences, 3(4): 358-373

Jain S, Naik PK. 2012c. Communication of signals and responses leading to cell death using Engineered Regulatory Networks. Research Journal of Pharmaceutical, Biological and Chemical Sciences, 3(3): 492-508

Janes KA, John AG, Suzanne G, Peter SK, Douglas LA, Michael YB. 2005 A systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis. Science, 310: 1646-1653

Morris F. White 2003 Insulin signaling in health and disease science, 302: 1710-1711

Normanno N, De Luca A, Bianco C, Strizzi L, Mancino M, Maiello MR, et al. 2006. Epidermal growth factor receptor (EGFR) signaling in cancer. Gene, 366: 2-16

Ringle C, Wende S, Will A. 2010. Finite mixture partial least squares analysis: Methodology and numerical examples. In: Handbook Partial Least Squares (Vinzi VE, Chin W, Henseler J, Wang H, eds). Springer, Heidelberg, Germany

Suzanne G, Janes KA, John AG, Emily PA, Douglas LA, Peter SK. 2005. A compendium of signals and responses trigerred by prodeath and prosurvival cytokines. Manuscript M500158-MCP200.

Thoma B, Grell M, Pfizenmaier K, Scheurich P. 1990. Identification of a 60-kD tumor necrosis factor (TNF) receptor as the major signal transducing component in TNF responses. The Journal of Experimental Medicine, 172: 1019-1023

Weiss R., 2001 Cellular Computation and Communications Using Engineered Genetic Regulatory Networks. PhD Thesis. MIT, USA