SindhUniv. Res. Jour. (Sci. Ser.) Vol. (49) (004) 849-852 (2017)

http://doi.org/10.26692/sujo/2017.12.0070

SINDH UNIVERSITY RESEARCH JOURNAL (SCIENCE SERIES)

Regression and correlation analysis on rotating biological contactor

M. R. DAUDPOTO[++], M. G. H. TALPUR1, A. KHOOHARO* K. ARAIN*

Department of Statistics, University of Sindh, Jamshoro 76062, Pakistan

Received 11th March 2017 and Revised 27th August 2017

Abstract: Regression widely used for prediction and forecasting a number of applications with function. Regression analysis estimates the uncertain expectation of according to the defined variables the independent variables. Regression is used to understand the relationships between the variables and arguments relating to the use survey. Overall, the regression variables defined by conditional expectation beam of argument-that the variable is normally the least resolved the disagreement; Focal points in the distribution parameters or arguments temporary variables. In Sindh University hostels the following experiment was conducted on Rotating Biological contactor through different parameters such as retention time and multiple numbers of disks. Experiment shows the removal on different retention time levels i-e; 2 hours, 2.5 hours and 3 hours with multiple number of disks of Biological Oxygen Demand (BOD). "r" indicates that there is a positive correlation in between the influent and effluent figures, where as the co-efficient of determination (r^2) indicates (when multiply by 100) the reliability of effluent on influent data. Co-efficient of determination (r^2) also indicates that the regression equations are how much reliable for the estimation.

Keywords: Regression, Correlation, Co-Efficient Of Determination, BOD

1. INTRODUCTION

In the populous states or countries, the water resources are under great stress specifically. (Gleick 2003) mentions that during last 20 years, shortages are revealed in per capita decline of available water in all regions of the world. The climatic change, ever increasing population expansion, swift urbanization and industrialization have significantly threatened the resources of fresh water which are depleting day by day as stated by Nghiem *et al* (2006) and USEPA (2004).

Regression analysis is a statistical procedure for estimating the relationship is surrounded by a variable which will include techniques for modeling and analysis of many variables when the center of attention lies in cooperation between variable and one or more independent variables, or predict (Jacob cohen et al 2003). More specifically, the regression analysis helps one understand that featured variable values according to the criteria or change variables when any one of a variety of variable while other free variables that are made permanent. In General, the regression analysis estimates the conditional expectation of according to the defined variables the independent variables-that is, the average of the independent variable when the variable is resolved less than usual; the focus is on the quintile or other places of temporary distribution parameter variables as independent variables (Andrew F. Hayes 2013). In all cases, the goal is to estimate the function of the independent variables, called the regression function in regression analysis; it is a curiosity that will describe the differences of the variables, based on the regression function around that can explain the probability distribution.

The symbol "X" is used to describe the independent variables and linear regression analysis about the relation between the variables and argument strength. When a single regression analysis of independent variable it is called a two way called (or simple). A linear regression analysis, when there are two or more independent variables involved in the analysis, it is called a multiple regression. The coefficient of correlation is generally used to analyze and determine the similarity interrelated degrees of variables **(**Yan 2003**).**

The numerical measure of strength of the linear relationship between two variables is known as the coefficient of correlation (Shahid 2016). Among two variables, the correlation is a measure of the closeness of linear relationship (Syed 2013).

Correlation of coefficient (r) =
$[N\Sigma XY - (\Sigma X)(\Sigma Y) / \text{Sqrt}([N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2])]$
(Where "r" is the sign or symbol of Pearson's Correlation Coefficient) linear correlation coefficient always lies between-1 and +1(Chaudhry and Kamal 2012)

2. MATERIAL AND METHODS

The following experiment was conducted in Sindh University hostels on Rotating Biological contactor

[++]Corresponding authors Email: doudpoto.rafique@usindh.edu.pk
* Department of Statistics, Sindh Agriculture University Tando Jam

through two factors, i.e. retention time and number of disks, affecting the performance of Rotating Biological Contactor. Significant results were obtained which reflect the performance efficiency of rotating biological contactor (RBC) simulator, during the course of work (A.A. Pathan 2015). In the present study, multiple procedures of retention time i.e. 2 hours, 2.5 hours and 3 hours and from 40 up to 52 numbers of disks were applied and results were noted. Forty numbers of disks showing 0% area increment, forty two disks showing 5% area increment forty four as 10%, forty six as 15%, forty eight as 20%, fifty as 25%, fifty two numbers of disks show 30% area increment respectively.

Statistical Package for Social Sciences software (SPSS: 21) was used for analysis. The data is gathered and interpreted the analysis table. In this study, the same data in Excel and other computer software and have been through their analysis.

The results of this study the formulas down analysis was performed. For linear equation

$Y = b_o + b_1x$

The number b_o is called the y-intercept and the number b_1 is called the slope of line. (Weiss NA 2007)
The values of b_o and b_1 (in straight line equation can be symbolize by a and b.

For multiple regressions with two predictor variables x_1 and x_2, the regression equation is in the form:

$y = b_o + b_1 x_1 + b_2 x_2$

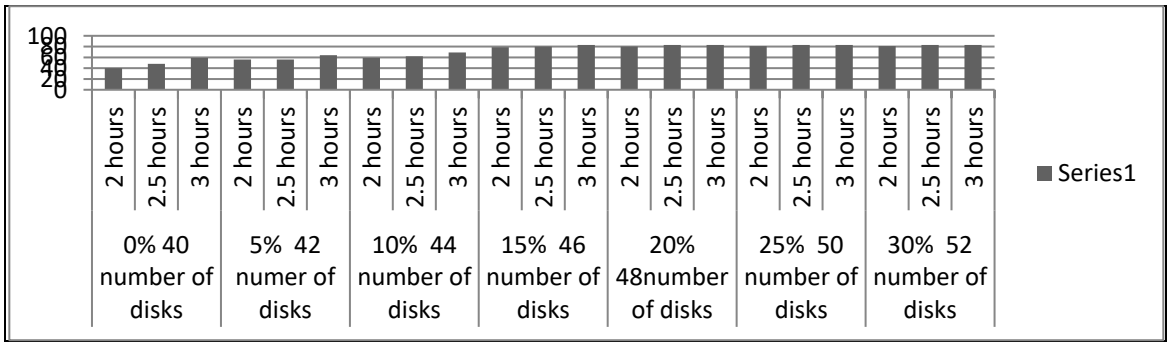
And in general; for multiple regression with k predictor variables x1, x2, ., xk the regression equation is in the form(Groeber et. Al 1993):

$y = b_o + b_1 x_1 + b_2 x_2 + \ldots\ldots\ldots\ldots + b_k + x_k$

Where b_o is y intercept and
b_1, b_2, and b_k are the parameters of regression of equation.

3. RESULTS AND DISCUSSION

Average removal of Biological Oxygen Demand on different time levels i-e; 2 hours, 2.5 hours and 3 hours with multiple numbers of disks are shown in the following table:



Following table shows the removal on different time levels i-e; 2 hours, 2.5 hours and 3 hours with multiple number of disks of biological oxygen demand. In the table the Co-efficient of correlation (r) indicates that there is a positive correlation in between the influent and effluent figures, where as the co-efficient of determination (r^2) indicates (when multiply by 100) the reliability of effluent on influent data. Co-efficient of determination (r^2) also indicates that the regression equations showing in the table are how much reliable for the estimation.

Disk area increment	Number of disks	Time	r	r^2	Minimum removal in %	Maximum removal in %	Mean removal in %	Range	Regression equation
0%	40	2h – 0.42 liter/min	0.987	0.974	37.1	42.9	40	5.8	0.536x+2.303
0%	40	2.5h – 0.331 liter/min	0.978	0.956	45.9	51.6	48	5.7	0.544x-0.825
0%	40	3h – 0.281 liter/min	0.885	0.783	50.0	63.0	59	13	0.351x+1.102
5%	42	2h – 0.42 liter/min	0.962	0.925	53.3	58.3	56	5	0.422x+0.430
5%	42	2.5h – 0.331 liter/min	0.872	0.760	47.8	62.5	56	14.7	0.459x-0.522
5%	42	3h – 0.281 liter/min	0.742	0.550	57.6	69.2	64	11.6	0.325x+1.102

10%	44	2h – 0.42 liter/min	0.884	0.781	55.9	62.5	59	6.6	0.501x-2.849
10%	44	2.5h – 0.331 liter/min	0.885	0.783	58.6	66.7	62	8.1	0.348x+0.954
10%	44	3h – 0.281 liter/min	0.922	0.850	65.8	70.7	69	4.9	0.320x-0.232
15%	46	2h – 0.42 liter/min	0.580	0.336	73.3	81.8	79	8.5	0.122x+3.248
15%	46	2.5h – 0.331 liter/min	0.580	0.341	73.1	85.7	80	12.6	0.214x-0.282
15%	46	3h – 0.281 liter/min	0.581	0.648	78.1	85.7	83	7.6	0.180x-0.034
20%	48	2h – 0.42 liter/4min	0.756	0.573	72.7	86.4	80	13.7	0.423x-4.863
20%	48	2.5h – 0.331 liter/min	0.805	0.648	76.3	87.9	83	11.6	0.392x-7.488
20%	48	3h – 0.281 liter/min	0.913	0.835	77.4	87.5	83	10.1	0.527x-9.583
25%	50	2h – 0.42 liter/min	0.178	0.031	76.7	84.5	81	7.8	0.156x+2.311
25%	50	2.5h – 0.331 liter/min	0.806	0.649	76.1	88.7	83	12.6	0.563x-23.07
25%	50	3h – 0.281 liter/min	0.836	0.698	78.2	87.5	83	9.3	0.725x-27.82
30%	52	2h – 0.42 liter/min	0.898	0.806	78.3	84.2	81	5.9	0.272x3.191
30%	52	2.5h – 0.331 liter/min	0.707	0.499	78.6	86.8	83	8.2	0.245x-2.624
30%	52	3h – 0.281 liter/min	0.960	0.921	81.0	88.5	83	7.5	0.242x-2.390

Regression Analysis

Table 1: coefficients as well as their consequent t and p values for Regression equation

Predictor	Coefficiet	S.D	T	P
Constant	-79.995	6.684	-11.97	0.000
Time	6.881	1.264	5.44	0.000
Disk	2.9008	0.1278	22.70	0.000

S = 7.151 R-Sq = 74.2% R-Sq(adj) = 74.0%

Table 2: Analysis of Variance (Table containing F and p-values) showing the effect of the removal of BOD

Source	DF	SS	MS	F	P
Regression	2	27857	13928	272.37	0.000
Error	189	9665	51		
Total	191	37522			

The amount of Biochemical Oxygen Demand and the other key information of grey water were composed from Sindh University Hostels by different procedures concerning the disks area, and the preservation time. Above results mentioned in table, indicates that there is a strong positive correlation in between the retention time and number of discs. The table mentioned above, reveals that estimate of retention time and number of discs have highly significant effect on the removal of BOD, as their corresponding p-values are less than even 0.01, while level of significance was 0.05. However ANOVA table was constructed to confirm the results obtained by applying Regression Analysis. Table 2 shows that retention time and number of discs have highly significant effect on removal of BOD through RBC, because the corresponding p-values are less than 0.01, while level of significance was 0.05.

The average removal percent is from 40 up to 83 percent. It also implies that as the number of discs and time expands there is a positive tendency with removal percent. Study also revealed that, as the time and number of discs increased the removal rate also increases but it happens at a particular level of retention time and number of discs, after than the removal rate remains parallel and afterwards tends towards decline. Average removal at Forty numbers of disks showing 0% area increment is from 40% up to 59% , at forty two disks showing 5% area increment is from 56% up to 64%, at forty four number of disks as 10% area increment is from 59% up to 69% , at forty six number of disks as 15% area increment is from 79% up to 83%, at forty eight number of disks as 20% area increment is from 80% up to 83%, at fifty number of disks as 25% area increment is from 81% up to 83% and fifty two numbers of disks show 30% area increment is from 81% up to 83% respectively.

As discussed, Statistics for natural sciences software (SPSS: 21) was used for analysis and found the results. The predictors are a constant with two variables time and disc. The dependent variable removal is explained through multiple regression frameworks by two explanatory variables time and number of disks. The constant representing b_0 (actually "y" intercept) comes out (-79.594) b_1 comes out (6.752) and b_2 comes out (2.899) respectively. The regression equation $y = b_0 + b_1 x_1 + b_2 x_2$ is in the form:

$$\hat{y} = -79.594+6.752(\text{Time}) +2.899(\text{Disk})$$

Now consider the interpretation of the explanatory variables of experiment, x_1 and x_2. The positive sign on coefficients of x_1 and x_2 indicates that removal of Biological Oxygen Demand increases when time and number of disks increase. The result are significantly different at $p<0.05$. Co-efficient of correlation ($r =0.861$) indicates that there is a strong positive correlation in between the removal with time and number of disks. Whereas the co-efficient of determination ($r^2 =0.742$) indicates (when multiply by 100) we get 74.2% the reliability of the above mentioned regression equation. In other words the regression equation is 74.2% reliable for the estimation.

REFERENCES:

Ashfaque A. P. (2015), Greywater Reclamation In Tropical Areas Through Rotating Biological Contactor (Muet) Jamshoro

Andrew F. H. (2013) Introduction to mediation, moderation and conditional process Analysis. A Regression based approach, Guilford press, New York.

Chaudhry, S. M., S. Kamal (2012). Introduction to statistical theory. Part-1. 8th ed. Ilmi kitab khana, Urdu bazaar Lahore, Pakistan.

Gleick, P. H. (2003), "Water Use," *Annual review of environment and resources*, Vol. 28, No.1, 275-314

Groeber, D. F. P, W. shanon, P.C. Fry, and K.D. Smith. (2013). Business Statistics, A decision making approach 8th ed. New Jersey.

Jacob C., P. Cohen, G. Stephen G. West, Leona S. Aiken (2003) Multiple Regression/Correlation Analysis for the Behavioral Sciences 3rd ed. Mahwah New Jersey.

Nghiem, D., N. Oschmann, and I. Schafer, (2006) "Fouling in Greywater Recycling by Direct Ultra filtration," *Desalination*, Vol. 187, 283-290

Syed K. A. (2013) Comprehensive Statistics concepts and methods, 17thed. Iqra publishers Urdu Bazar Karachi

Shahid J., (2016) Statistics Problems and practice, Ahmed academy Urdu Bazar Karachi

Weiss, N. A. (2017) Introductory statistics. Pearson Education Limited England ISBN 9780321989178.

Yan Li-kun (2003) Application of Correlation Coefficient and Biased Correlation Coefficient in Related Analysis, J. Yunnan Finance and Trade Institute