Indira Gandhi
National Open University
School of Sciences

MSTE-002
Industrial Statistics-II

Block

3

REGRESSION MODELLING

Curriculum and Course Design Committee

Prof. K. R. Srivathsan
Former Pro-Vice Chancellor
IGNOU, New Delhi

Prof. Rahul Roy
Theoretical Statistics and Mathematics Unit
Indian Statistical Institute, New Delhi

Prof. Parvin Sinclair
Former Pro-Vice Chancellor
IGNOU, New Delhi

Prof. Diwakar Shukla
Department of Mathematics and Statistics
Dr. Hari Singh Gaur University, Sagar

Prof. Geeta Kaicker
Former Director, School of Sciences
IGNOU, New Delhi

Prof. Rakesh Srivastava
Department of Statistics
M.S. University of Baroda, Vadodara

Prof. Jagdish Prasad
Department of Statistics
University of Rajasthan, Jaipur

Prof. G. N. Singh
Department of Applied Mathematics
I.S.M., Dhanbad

Prof. R. M. Pandey
Department of Bio-Statistics
AIIMS, New Delhi

Prof. Gulshan Lal Taneja
Department of Mathematics
M.D. University, Rohtak

Dr. Manish Trivedi
School of Sciences, IGNOU

Prof. Poornima Mital
School of Sciences, IGNOU

Dr. Nitin Gupta
School of Sciences, IGNOU

Prof. Sujatha Varma
School of Sciences, IGNOU

Dr. Neha Garg
School of Sciences, IGNOU

Dr. S. Venkataraman
School of Sciences, IGNOU

Mr. Rajesh Kaliraman
School of Sciences, IGNOU

Dr. Deepika
School of Sciences, IGNOU

Block Preparation Team

Prof. Rakesh Srivastava (Editor)
Department of Statistics
M.S. University of Baroda,
Vadodara (GUJ)

Prof. G. K. Shukla (Units 10 to 12)
(Retd.), Department of Applied Mathematics,
Indian Institute of Technology,
Kanpur (UP)

Prof. Vijayshri (Language Editor)
School of Sciences, IGNOU

Dr. Manish Trivedi (Units 9 to 12)
School of Sciences, IGNOU

Programme and Course Coordinator: Dr. Manish Trivedi

Block Production

Mr. Sunil Kumar, AR (P)
School of Sciences, IGNOU

September, 2014

ISBN-978-81-266-

Further information on the Indira Gandhi National Open University may be obtained from the University's Office at Maidan Garhi, New Delhi-110068 or visit University's website http://www.ignou.ac.in

Printed and published on behalf of the Indira Gandhi National Open University, New Delhi by the Director, School of Sciences.

Laser Typeset by: Tessa Media & Computers, C-206, A.F.E.-II, Okhla, New Delhi

Printed at:

BLOCK 3 REGRESSION MODELLING

In Blocks 1 and 2 of this course, you have learnt some basic methods of solving linear programming problems, transportation problems and assignment problems. In these blocks, you have also learnt various techniques to optimise the processing/service time or total cost while dealing with the queueing systems, sequential problems or inventory problems, respectively.

In this block, we discuss the concepts of regression modelling, which is one of the most widely used techniques for analysing multivariable data. Regression analysis is the process for investigating and analysing the average relationship between the dependent or response variable and a set of independent or regressor variables. The purpose of regression analysis is to establish a relationship between a dependent variable and one or more than one regressor variables and to predict the optimum value of dependent variable using this relationship. Regression analysis helps us model practical problems dealing with real world data. Therefore, a thorough knowledge of regression modelling is essential for a statistician. We present the concepts of regression modelling in 4 units of this block.

In Unit 9 entitled **Simple Linear Regression**, we discuss how to establish the linear relationship between a dependent variable Y and an independent variable X using an equation $Y = a + bX$. The coefficients a and b are unknown and for given data on Y and X, we obtain the least squares estimates of these coefficients. Residual analysis is also performed for detecting the deficiencies in this regression model.

In Unit 10 entitled **Statistical Inference in Simple Linear Regression**, we discuss the inferential aspect of simple linear regression. It is of interest to examine whether a significant linear relationship exists between Y and X. The coefficient of determination, R^2, which gives a measure of goodness of fit of the model, is explained in this unit. We also discuss the confidence interval for the predicted value of E(Y) corresponding to a given value of X.

Unit 11 entitled **Multiple Linear Regression** deals with multiple regression modelling. We examine the adequacy and overall fit of the model with the help of the coefficient of determination R^2. We also discuss a method for calculating R^2 and adjusted R^2.

In Unit 12 entitled **Selection of Variables and Testing Model Assumptions**, we discuss how to build a regression model for satisfactory explanation of variability in the dependent (response) variable Y. The forward selection, backward elimination and stepwise selection procedures for selection of variables in multiple regression modelling are also described in this unit. The other important aspects covered in this unit are verification of assumptions such as normality and the homogeneity of variances.

Notations and Symbols

Y_i	:	i^{th} observation of the dependent variable Y
X_i	:	i^{th} observation of the independent variable X
$\hat{Y}_i$	:	i^{th} observation of the fitted value
r_i	:	i^{th} residual value, i.e., $Y_i - \hat{Y}_i$
d_i	:	i^{th} standardised residual value
e_i	:	i^{th} error term
p_i	:	Cumulative probability of i^{th} ranked residual value
P_i	:	Percentile cumulative probability of i^{th} ranked residual value
E	:	Sum of squares of the error terms
$Var(\overline{Y})$	:	Variance of $\overline{Y}$
SS_{Res}	:	Sum of squares of residuals
$\hat{a}$ and $\hat{b}$	:	Unbiased estimates of regression parameters a and b
R^2	:	Coefficient of determination
F_X	:	Lowest partial F-value
SS_{Reg}	:	Sum of squares due to regressor variables
SS_T	:	Total sum of squares
B_j	:	Amount of change in Y for a unit change in X_j
B_0	:	Intercept and coefficient of dummy variable X_0
I_n	:	Identity matrix of order n
$X'X$	:	A Non-singular Matrix
$V(\hat{B})$	:	Variance-Covariance matrix of $\hat{B}$
Secs.	:	Sections/Sub-sections
Fig.	:	Figure