

Proceedings of the 1st Annual World Conference  
of the Society for Industrial and Systems Engineering,  
Washington, D.C, USA  
September 16-18, 2012

## **Multivariate Analysis of Breast Cancer Prediction Parameters**

**Benjamin Schleich, Hema Sudarsanam, and Chanchal Saha**

Department of Systems Science and Industrial Engineering  
State University of New York at Binghamton  
Binghamton, New York 13902

Corresponding author's Email: [bschlei1@binghamton.edu](mailto:bschlei1@binghamton.edu)

**Author Note:** The authors are from the Department of Systems Science and Industrial Engineering (SSIE) and working in the Watson Institute for Systems Excellence (WISE) and Center for Autonomous Solar Power (CASP). The authors would like to convey their most profound thanks to SSIE, WISE and CASP for providing information, resources, and incessant support throughout the entire study period. Benjamin Schleich's (corresponding author) tel.: +1-719-214-4924.

**Abstract:** This study explores breast cancer data with multivariate techniques. Breast cancer is the second leading disease that causes death, 100 deaths per day, of women living in the USA. Therefore, this study aims in identifying the independent decision parameters given in the data-set which are accounted for correct prediction of breast cancer. Logistic Regression Analysis (LRA) is performed to assess the correct diagnosis probability. A Principal Component Analysis (PCA) is also performed to determine the number of decision parameters can be reduced. The experimental results indicate that the LRA has a very high correct validation while the PCA suggested that the principal components can be significantly reduced. In future, Factor Analysis (FA) can be conducted to find the latent relations among the variables, which might give a better understanding of the correlation among the variables.

**Keywords:** Breast cancer, Logistic regression analysis, Principal component analysis