# Fuzzy Regression Analysis and Application of Interval Fuzzy Random Variables


## Abstract

The aim of this paper is to discuss the linear correspondance between two interval fuzzy random variables. We construct the regression equations of the upper and lower bounds of some interval fuzzy random variables, respectively, by the least squares. The upper and lower bounds of the estimated interval fuzzy random variables are derived by the regression equations of upper and lower bounds, respectively.

The collected upper and lower bounds are all crisp data, not fuzzy ones. In this paper, the interval fuzzy random variables discussed are constructed by crisp upper and lower bounds. In order to increase the reprsentative of the interval fuzzy random variables, we need to minimize the errors of the estimated upper and lower bounds. Applying the least squares along with the conventional regression analysis to construct regression lines of upper and lower bounds, respectively, should be the better way to minimize the errors of the estimated upper and lower bounds.

However, the errors of the upper and lower bounds estimated by the least squares are the least according to the arithmetic mean value. The more discrete the data we collected , the less representative of the arithmetic mean value is. That will also affect the accuracy of the estimated interval fuzzy random variables. This is what we are worried while we take the least squares as an tool to analyse the interval fuzzy random variables.

The coefficient of determination is a reference value which is mostly often used to distinguish the accuracy of the conventional regression model. In the view of the characteristics of fuzzy regression model, the conventional coefficient of determination cannot properly explain the fuzzy linear regression model. In this paper, we propose the fuzzy coverage rate to distinguish the accuracy of the fuzzy linear regression model between two interval fuzzy random variables. Finally, we give an example about the mean monthly

working-hour and the mean monthly salary of the manufacturing industry in Taiwan from 1991 to 2007, demonstrating the application of the fuzzy coverage rate in reality.

**Keywords:** interval fuzzy random variablesfuzzy, linear regression model, fuzzy coverage rate

# 摘要

本研究主要是探討兩個區間型模糊數之間的直線對應關係。主要的方式是以最小平方估計法(least squares estimation)分別求出區間型模糊數上、下界所對應的迴歸方程式，以該迴歸方程式所求得的上、下界，做爲所估計區間型模糊數的上、下界。

單就所蒐集到的上界或下界資料而言，它們是一組明確的資料，並不模糊。研究中所探討的區間型模糊數是由一組明確的上、下界值所構成的。考慮所估計的上、下界值需具有較小的誤差才能增加所構成區間型模糊數的代表性，使用最小平方估計法並以傳統的迴歸方式來求得上、下界迴歸直線，應該是減少估計的上、下界值誤差較佳的方式。

然而以最小平方估計法所估計的上、下界值是相對於資料算術平均數誤差最小。如果所蒐集到的數據愈分散，則算術平均數的代表性將愈低，連帶影響所估計區間型模糊數的準確性。這是使用最小平方估計法做爲研究工具的隱憂。

解釋係數是最常被用來判別迴歸模型優劣的參考數值。有鑑於區間型模糊數的模糊特性，傳統迴歸分析的解釋係數並不適用於模糊線性迴歸關係。本研究提出模糊覆蓋率的概念，來判別兩個區間型模糊數之間線型迴歸關係的優劣。最後以中華民國 80 年到 96 年間製造業平均月工時對應平均月薪資爲例，說明模糊覆蓋率在實務上的應用。

關鍵字：區間型模糊數、模糊線性迴歸分析、模糊覆蓋率