第一章 緒論

迴歸分析目的主要就是要去探索反應變數(Y)與解釋變數($X_1,\ldots,X_p$)之間的關係，並希望利用這個關係經由觀察($X_1,\ldots,X_p$)去對反應變數(Y)做預測。在這探索過程中，最重要的就是如何決定哪一個變數可納入於模型。當然如果我們分析目的只考慮在預測優劣，我們可以採納所有解釋變數。但我們知道引入變數過多，將使得預測偏誤(bias)較小，而預測變異(variance)較大；相反的引入變數太少，將使得預測偏誤(bias)較大，而預測變異(variance)較小，這之間存在著權衡關係。傳統的變數選取方法有子集挑選(subset selection)、逐步迴歸(stepwise regression)；其中子集挑選即在所有可能迴歸分析(a11 possible regression)中，利用某些準則去選定其中一子集使得準則達到最佳，當解釋變數的量太大，將因計算量龐大而受限制；逐步迴歸則利用序列假設檢定去決定合適的子集，雖然計算量較小，就某些準則而言，它並不保證可以達到最佳。在另一方面，壓縮(shrink)係數方法則是放棄傳統最小平方估計和子集挑選，利用所有解釋變數配適模型，而是利用控制偏誤的(biased)估計大小來做變數選取。其中如脊迴歸(ridge regression) (Hoerl and Kennard，1970)、Garrote (Breiman，1995)、LASSO (Tibshirani，1996)等，已經成為各領域在處理多維度資料的重要工具。

James 與 Stein 在 1961 年首先提出 James-Stein 估計量(JS)，此估計具有 minimax 性質但此估計並無門檻值。Baranchik 在 1964 年為了解決 JS 估計量其壓縮比例變成負數時的問題，提出 James-Stein positive part 估計量(JS^+)，但是 JS^+ 只能在完全模型(full model)與原始模型(origin model)兩者去做挑選。Zhou 與 Hwang 在 2005 年，為了改善 JS^+其缺點，建立了具有 Minimax 性質同時加上門檻值的估計量，即 James-Stein with Threshoding positive part 估計量($JSWT^+$)。由於 $JSWT^+$估計量具有門檻值，使得此估計量可以在完全模型與其線性子集下做變數選取。Zhou 與 Hwang 將 $JSWT^+$估計量應用於小波分析對函數做估計，展示

JSWT$^+$估計量與一些可同時做估計與模型挑選的估計量比較結果顯示 JSWT$^+$表現最好，即在均方誤差(MSE)比較下為最小。

Sclove 在 1968 年提出將 JS$^+$應用於線性迴歸模型的 Sclove 估計量，但是並沒有建立合適的門檻值，所以此估計量依舊限制於只能在完全模型(full model)與原始模型(origin model)兩者去做挑選。而既然在 JSWT$^+$估計應用於小波分析時對函數做估計得到不錯的估計效果，我們想進一步了解如果將 JSWT$^+$估計量應用於線性迴歸分析時，藉由 JSWT$^+$估計具有門檻值的性質去做變數選取的效果如何？

本文目的將利用 JSWT$^+$估計量具有門檻值的性質，建立 JSWT$^+$估計量應用於線性迴歸分析變數挑選的流程，並以模擬的方式與可同時做係數壓縮及變數選取的 LASSO 方法去比較；如係數路徑及變數挑選差異等，最後將 JSWT$^+$變數選取的流程應用於攝護腺癌資料(Tibshirani，1996)作變數挑選。