

HW2 <Data Assembly & Data Cleaning>

彭柏豪

(1)

資料種類：

四個樣區（栽植密度不同）中的Norway spruce的胸高直徑資料。因為不同的栽植密度，樣區中樹木的數量也有所不同。四個樣區的面積記錄在area.csv檔案中。四個樣區的胸高直徑資料則分別記載於四個.csv檔案中。

地點：

Austrian Research Center of Forests於Vienna Woods在栽種四個樣區。

資料年份：

於1892年開始計畫，於1923年於林分年齡35歲時開始第一次測量。而最後於1997年於收成前做最後一次測量，中間年份陸續有做測量，間隔1至3年不等。

詳細敘述：

沒有在同一年做過兩次以上的測量，因此時間尺度可以使用年作為標準。一開始測量總共有4507棵樹木，到了收成前剩下456顆樹木存活。

單位：

樣區面積單位為平方公尺（m²），而樹木胸高直徑則為毫米（mm）。

其他：

資料為Georg E. Kindermann, 2018出版<<109 years of forest growth measurements from individual Norway spruce trees>>中提供下載。

(2)

原本資料是作為栽植密度對於生長的影響，因此資料的整理上較為繁瑣要將以單株樹木最為標的物的資料轉為變成以樣區作為標的物，且樹木數量隨著時間逐漸減少，並未補上新的樹木，因次處理欄位會出現很多空白欄位。

(3)

將樹木資料轉變為以年為軸的逐年遞增陣列——排列。並且將所有樹木依照所屬樣區的不同，分別處理進入不同的.csv檔案。如果兩個測量資料中有missing data，會用時間去做加權平均，因為樹木的胸高直徑應隨著時間遞增，且做林分的Basal Area估計應該期待還存活的樹木都可以提供數據。