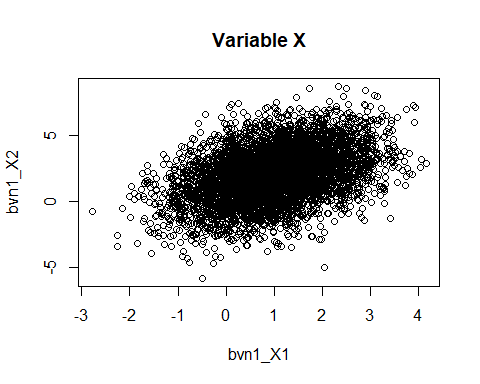
0504統算作業

108071601 賴冠維

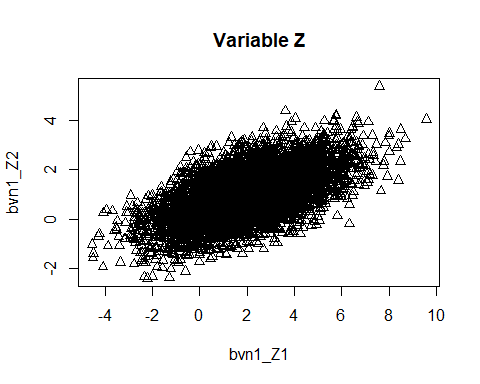
2021/5/3

## Problem 2

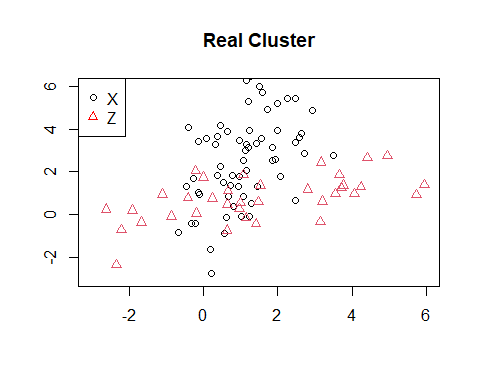
### 建立X,Z兩個二維常態分配

首先我們先建立X、Z變數，可見兩變數散佈圖如下： #### 變數X 

#### 變數Z



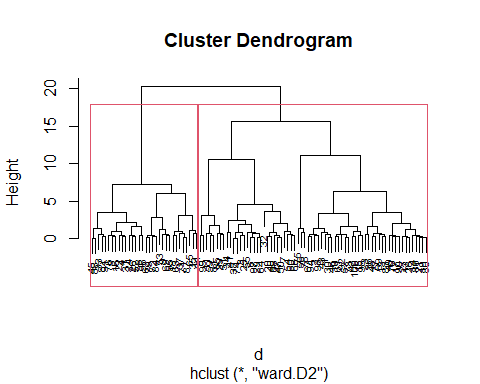
先模擬Uniform(0,1)樣本，在從該樣本中抽樣，若大於0.6則從Z當中抽一個樣本，反之從X中抽一個，直到達到100個樣本，以此形成兩者的混合樣本。

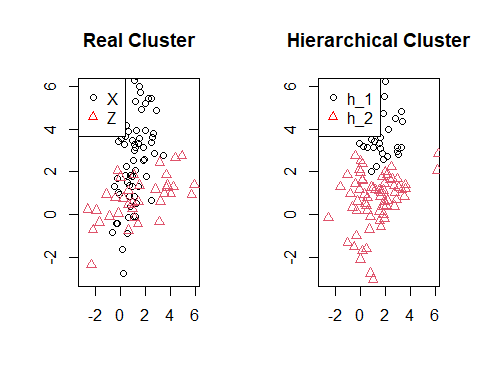
可以看到混合樣本散佈圖如下： 

#### (b)

對混合樣本採用,並選取作為合併標準，為最小組內變異法，其定義如下：

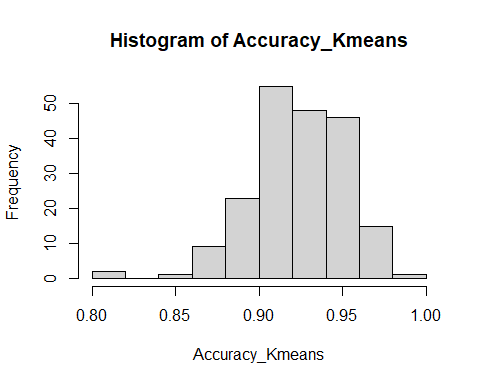
\* 為第i群集裡面第j個觀測值

以最小組內變異增加量為原則，逐步加入其他樣本直至合併到最後一群結束。  我們預先知道樣本來自兩母體，因此我們選取群數為2，結果如下： 

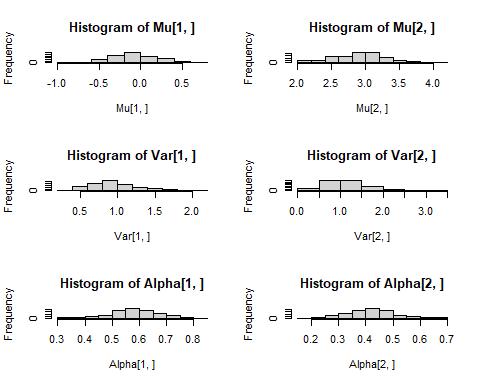
下圖左邊為真實兩變數在隨機樣本的分布，下圖右方為分群結果，可發現其分群的結果與真實的分群差異很大，可能來自於所選的合併標準是按照組內變異最小進行的，而真實樣本並不太符合兩組群組內變異較小的特性。 

## Problem 3

我們先重複的作法，先從,兩分配中，製造100筆隨機樣本，並重複200次，形成200組樣本數為100，由,組成的隨機樣本。  
再來對每組樣本進行的分群，設定

我們將200組Kmeans分群結果，於其正確分群計算，可得直方圖如下，可以發現大多數準確率都介在85%至95%之間，效果不錯。 

接著設定EM演算法的公式，我們將Kmeans分群結果，計算的平均數、期望值，作為起始值，在中計算以及兩群的機率，再來進入以MLE估計新的平均數、變異數，重複迭代此兩作法直到的改變量收斂至，結束演算法。

可看到200組收斂後所估計的所畫的直方圖，與我們真實分配 、進行比較，可發現非常接近真實分配的結果非常接近。 

最後看到EM演算法做出來的直方圖正確率一樣很高，大部分同樣落在85%至95%之間。 