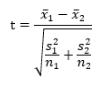
圖1藍色的為不同年齡層的男性，因腦血管疾病死亡的曲線圖，紅色的為不同年齡層的女性，因腦血管疾病死亡的曲線圖，由圖中可以看出兩者有類似的分布情形。

t value 計算公式: f-value 計算公式為: s1^2 / s2^2 where s is variance

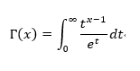
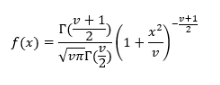
 平均值的變異量 / 樣本標準化

「男生去減掉女生」或「女生去減掉男生」都可以，差別是一個t值會是正的一個是負的，所以當需無假設是=的狀況，我們會用雙尾檢定，這時候單一邊算出來的p-value必須\*2(正負兩邊同時要考慮)。

但當需無假設是「平均數(男生死亡人數)≥平均數(女生死亡人數)」，這時候就有分右尾檢定還是左尾檢定，這時後算出來的p-value就不必在\*2了。

一般統計檢定都只看有沒有差異，所以幾乎都是用雙尾檢定。

t 分佈公式:



* 統計假設檢定:

虛無假設(H0): 男生死亡人數=女生死亡人數

對立假設(H1): 男生死亡人數≠女生死亡人數

* 計算出來的結果

女性樣本總人數:**118168** (已減去1)

男性樣本總人數:**167010** (已減去1)

女性平均死亡人數是 **6951** 標準差為**7568.59**

男性平均死亡人數是 **9824** 標準差為**9487.49**

t value is **89.79701158186265**

f value is **1.5713521171912956**

f value越大，如果維度很大，可以看哪一個維度的東西比較可靠。

圖二為最後的結果，藍色的曲線幾乎和常態分佈一模一樣。

紅色的為t value和 - t value，此次雙尾檢定，可以看出t值和分佈曲線非常遙遠，所以假定男生死亡人數=女生死亡人數是要拒絕的，也就是男生死亡人數和女生死亡人數有統計上顯著的差異。

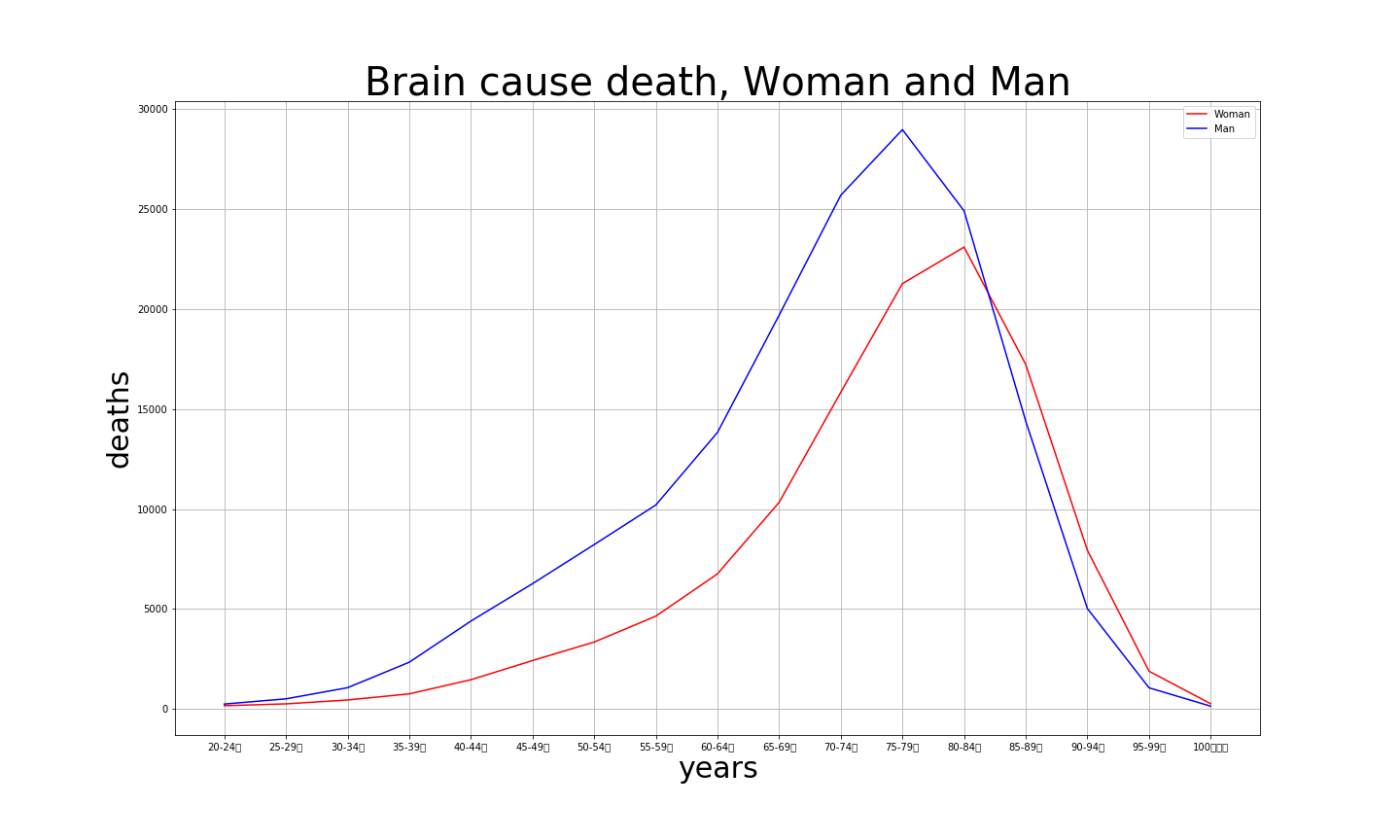


圖1 腦血管疾病，不同年齡層男女的死亡曲線圖

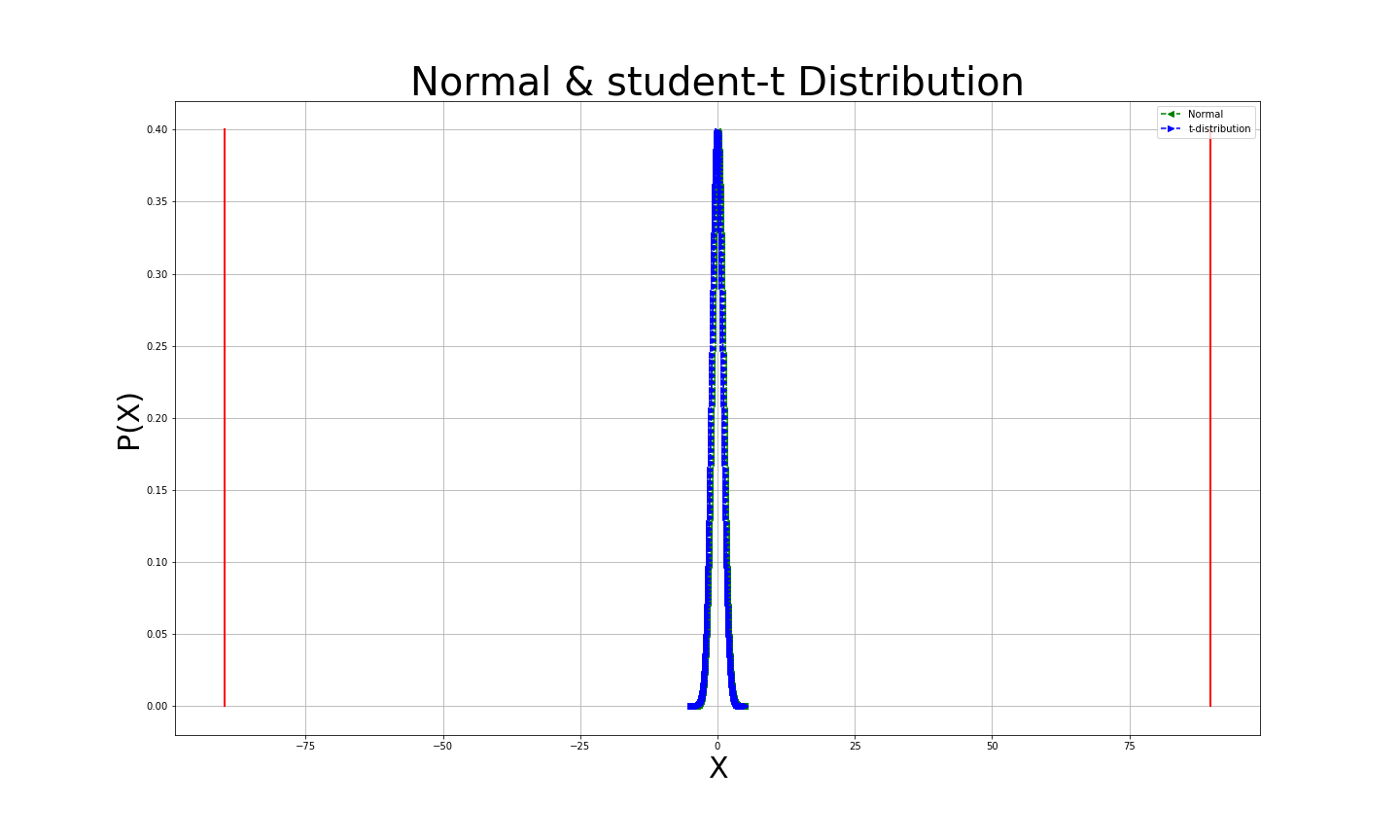


圖2 t-distribute的分布曲線與t value的圖形

PS:資料來源:政府開放資料平台

參考資料:

f-test:

<https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/hypothesis-testing/f-test/>

t-test wiki:

<https://en.wikipedia.org/wiki/Student%27s_t-test>

Basic Statistics in Python: Probability:

<https://www.dataquest.io/blog/basic-statistics-in-python-probability/>