您真的被檢驗出疾病了嗎? — 條件機率與貝氏定理

李幸娟(應數博 96)

在人們忙忙碌碌的二十一世紀,現代人的身體健康常常需要靠著”定期健康檢查 ” 來把關,但如何解讀和因應健檢報告,無論對醫生或病人而言,這都是一門學問。其中第一步是”**如何利用機率理論來計算某稀有疾病檢驗所帶來的偽陽性比率** ” 此偽陽性比率往往可用來說明某些疾病檢驗如呈現陽性反應,還需要回診複檢的理由。

為簡化問題(註:問題來源[1]),首先我們假設某一種稀有疾病,在一個人口群體中,每 1000 人會有 1 人受感染,且在醫學上已有方法可以檢驗此疾病,但不是百分之百有效(大部份的檢驗方法都有此特性),對於**已經染病**的人有 98 % 的檢驗結果呈現陽性反應,但對**未染病**的人也有大約 1.9 % 的檢驗結果為陽性,那麼如果某人剛作完此檢驗,結果為陽性反應,我們的問題是”**此人感染這稀有疾病的機會有多大呢 ?** ” 。

接著我們可以應用**條件機率**與**貝氏定理**(附錄 I)在這個非常正經的攸關人命的問題上,先解釋何謂條件機率,若已知 B 事件發生了,另一事件 A 又同時發生的機率,我們稱為” 在 B 條件下發生 A 的條件機率 ” 記為$P(A|B)$,可想而知根據機率的定義 $P(A|B)=P(A且B)/P(B)$

因此 A,B 兩事件同時發生的機率可用條件機率$P(A|B)$與 B 事件發生的機率$P(B)$相乘而得,也就是$P(A且B)=P(A|B)\times P(B)$,在我們的個案中

A 事件代表受檢驗者確實染病的事件

B 事件代表受檢驗者的檢驗結果為陽性反應的事件

則由上述資料,我們有

$P(A)=\frac{1}{1000}=0.001$ **(1000 人中有 1 人感染)**

$P(非A)=1-\frac{1}{1000}=0.999$ **(1000 人中有 999 人未感染)**

$P(B|A)=98\%=0.98$ **(已染病者 98%呈陽性反應)**

$P(B|非A)=1.9\%=0.019$ **(未染病者有 1.9%呈陽性反應)**

$P(B且A)=P(A且B)=P(B|A)\times P(A)$ **(已染病且呈陽性反應)**

$=(0.98)\times(0.001)=0.00098$

$P(B且非A)=P(非A且B)=P(B|非A)\times P(非A)$ **(未染病且呈陽性反應)**

$=(0.019)\times(0.999)=0.01898$

而我們要求的是”檢驗結果為陽性下真正染病的條件機率”

即$P(A|B)=P(A且B)/P(B)=P(A且B)/[P(B且A)+P(B且非A)]$

$$=\frac{0.00098}{0.00098+0.01898}\fallingdotseq 0.0491$$ (利用貝氏定理如附錄 I)

雖然檢驗的準確性很高,但在檢驗結果為陽性的人當中真正染病卻只有 4.91% (低於 5%),這就是” 偽陽性詭論 ” 。

然後我們假設此群體為 1000 人,可將上述結果簡化成下表:

檢驗結果	染病人數	未染病人數	合計
結果為陽性人數	1	19	20
結果為陰性人數	0	980	980
合計	1	999	1000

您看出結論了嗎?雖然此稀有疾病檢驗會產生偽陽性,但它也告訴我們一些重要的訊息,此疾病的感染率本來是 1/1000(平均 1000 人會有 1 人感染),但如有某人被檢驗出陽性反應,則此人感染此病的機率會增加到 1/20(因為 1000 人裡頭只有 20 人陽性反應,而這 20 人中會有 1 人感染此病),因此建議醫生可作更詳細或更精密的檢查,事實上這是一般醫師常作的決策。

看官們有興趣也可以算算看”已知自己染病但卻檢驗不出來的機會有多大呢? ”,這相當於此疾病檢驗所擔的風險大小(註:求$P(非B|A)$答案在附錄 II)最後從這個醫學的例子中,我們顯示在資訊不完整且有實際風險的情況下如何利用所學的這些抽象觀念,做出明智的決定—這也是統計學的最終目的。

參考資料:

[1] 看漫畫學統計(The Cartoon Guide to Statistics, Larry Gonick , Woollcott Smith) 鄭惟厚譯 天下文化出版

[2] Business Statistics Understanding Population and Processes , Mario F. Triola , Leroy A. Franklin, ADDISON-WESLEY PUBLISHING COMPANY, 1996

附錄:

[I] 貝氏定理:

$$P(A|B) = P(B|A) \times P(A) \ / \ [\ P(B|A) \times P(A) + P(B|非A) \times P(非A)\ ]$$

[II] $P(非B|A)=\dfrac{0.00002}{0.00002+0.00098}$ ≒ 0.02 即表示檢驗 100 個已患病的病人,會有兩個人驗不出已經患有此病,可見此疾病檢驗風險不大。