統計方法應用--- Fisher's Exact Test　應數博黃士峰

費雪（R. A. Fisher, 1890-1962, 英國統計學家）在他 1935 年發表的一篇文章中提到一個有趣的實驗。有一天喝下午茶時，費雪的一位女同事說到，下午茶的調製順序對其風味有很大的影響，把茶加進牛奶裡或者是把牛奶倒入茶中，兩者喝起來口感完全不同，她可以完全地分辨出來，而她個人的偏好是先放牛奶再加入茶。為了證實這位女同事的說法，費雪做了一個小實驗。首先他調製了 8 杯下午茶，並且告訴女同事其中 4 杯是先放牛奶再加入茶，另外 4 杯則是先倒茶再加入牛奶，接著費雪隨機的拿了一杯茶讓女同事試喝，並請她猜猜這杯茶中是先放牛奶還是先放茶，等女同事回答後，再隨機的拿第二杯請她試喝，直到全部試喝完畢。假設女同事的答案如下表：

	猜測先放的飲料		
實際先放的飲料	牛奶	茶	合計
牛奶	a = 3	b = 1	a+b = 4
茶	c = 1	d = 3	c+d = 4
合計	a+c = 4	b+d = 4	n = 8

由表中得知費雪的女同事猜對了 6 杯，但這是否只是碰巧她當天運氣好呢？能否以統計方法讓我們由實驗的數據判斷費雪的女同事能否真的分辨出兩種不同的下午茶調製方法？另一個值得注意的事情是，在這個實驗中樣本數非常的小，是否有適當的統計方法可以幫助我們做出統計推論呢？

對於以上的問題, 費雪在 1935 的文章中提出一個小樣本的檢定方法，稱為 Fisher's Exact Test。這個方法的前提是固定邊際分布，也就是 a+b 、c+d、a+c、與 b+d 的值不變，接著計算在此假設下，得到實際觀測值的機率：

$$\mathrm{P}(\mathrm{a}=3\,|\,\mathrm{a}+\mathrm{b}=\mathrm{c}+\mathrm{d}=\mathrm{a}+\mathrm{c}=\mathrm{b}+\mathrm{d}=4)=\binom{\mathrm{a}+\mathrm{b}}{\mathrm{a}}\binom{\mathrm{c}+\mathrm{d}}{\mathrm{c}}\Big/\binom{\mathrm{n}}{\mathrm{a}+\mathrm{c}}=0.229\text{，}$$

其中 $\mathrm{P}(A\,|\,B)$ 是在 B 事件發生的前提下 A 事件發生的條件機率，而 $\binom{\mathrm{a}}{\mathrm{b}}\equiv\dfrac{\mathrm{a}!}{\mathrm{b}!(\mathrm{a}-\mathrm{b})!}$。這個數值的意義是如果費雪的女同事只是隨便亂猜的話，最後會得到上述表格中結果的機率為 0.229。接著我們可更進一步的算出比表格中更極端情況（在此指費雪的女同事猜得更準時）的機率：$\mathrm{P}(\mathrm{a}=4\,|\,\mathrm{a}+\mathrm{b}=\mathrm{c}+\mathrm{d}=\mathrm{a}+\mathrm{c}=\mathrm{b}+\mathrm{d}=4)=0.014$，因此我們可以再計算出 $\mathrm{P}(\mathrm{a}\geq 3\,|\,\mathrm{a}+\mathrm{b}=\mathrm{c}+\mathrm{d}=\mathrm{a}+\mathrm{c}=\mathrm{b}+\mathrm{d}=4)=0.229+0.014=0.243$，此即為統計上單邊檢定的 p 值（p-value）。如果此 p 值小於型一誤的機率（the probability of type one error，一般常設定為 0.05 或 0.01），則我們可以說有充分的證據支持費雪的女同事所言不虛，反之則是證據仍不足以證實她的說法。費雪在 1935 年的文章中並沒有告訴我們當時這位女同事到底猜對了幾杯，但根據費雪女兒後來的說法是，這位女同事當時全答對了（Agresti 2002, p.92）。

上述的小故事看起來好像只是一則有趣的實驗，對統計的實際應用有什麼影響呢？實際上，Fisher’s Exact Test 不僅簡單實用，尤其適用於樣本數小的情況，而且這個方法並不需要假設資料母體的分布，換句話說，Fisher’s Exact Test 屬於一種無母數的檢定方法。以下再舉幾個實際應用的例子：

一、 現在大學生流行節食，一般大眾的猜測是：大學女生節食的比例比男生高。因此我們設定的虛無假設為 H_0：大學女生與男生節食的比例相同，對立假設為 H_a：大學女生節食的比例比男生高。倘若經過調查後我們得到以下的資料：

	男生	女生	合計
節食	a = 1	b = 9	a+b = 10
未節食	c = 11	d = 3	c+d = 14
合計	a+c = 12	b+d = 12	n = 24

則我們可根據上表中的資料算出單邊檢定的 p 值

$$P(a \le 1 \mid a+b=10, c+d=14, a+c=12, b+d=12)$$
$$=\left[\binom{10}{1}\binom{14}{11}+\binom{10}{0}\binom{14}{12}\right]\Big/\binom{24}{12}=0.0014.$$

若設定型一誤的機率為 0.01，則我們可根據此 p 值推論：上表中的資料提供了足夠的證據拒絕虛無假設，並支持對立假設，即大學女生節食的比例比男生高。

二、 某家醫院統計了罹患某種罕見疾病時，男性與女性的死亡與存活資料如下：

	男性	女性	合計
存活	a = 9	b = 4	a+b = 13
死亡	c = 1	d = 10	c+d = 11
合計	a+c = 10	b+d = 14	n = 24

根據上表，女性死亡率似乎遠比男性為高，因此設定虛無假設為 H_0：女性與男性患者死亡率相同，對立假設為 H_a：女性患者死亡率較男性高。則我們可根據上表中的資料算出單邊檢定的 p 值

$$P(a \ge 9 \mid a+b=13, c+d=11, a+c=10, b+d=14)$$
$$=\left[\binom{13}{9}\binom{11}{1}+\binom{13}{10}\binom{11}{0}\right]\Big/\binom{24}{10}=0.0042.$$

若設定型一誤的機率為 0.01，則我們可根據此 p 值推論：女性患者死亡率較男性高。

參考文獻

Agresti, A. (2002). *Categorical Data Analysis*. John Wiley, New Jersey.

Fisher, R. A. (1935). *The Design of Experiments*. Oliver ＆ Boyd, Edinburgh.