樣本比例數的抽樣分配

如果母群體中的個體是按某種特性比如性別、教育程度、下不下雨、支持何候選人……等歸入各種類別中，在研究由該母群體中抽取的樣本時，我們計算的是歸入某一類別的數目，這種資料就為計數或點計資料。

在劃分類別時，如果類別只有二種，就符合二項分配。例如打過疫苗和沒打過疫苗、男性或女性(不是男性)、看過或沒看過某廣告……等等，本質上祗是屬於「是」與「否」二類。我們在調查研究時，只要記下樣本是否具有此特性即可，為了避免絕對次數受到總樣本數的影響，因此都是以百分比(比例數)作為衡量的基準，這個百分比首起來帶有小數，但在本質上，它還是屬於間斷資料。

假設某候選人在所有選民(母群體)中的支持率為 π (通常以希臘字母表示)，可是這個 π 要在投票那天才會顯現出來，在選前，我們只能從所有的母群體中隨機抽出 n 個樣本，調查這 n 個人中有多少人支持該候選人，這就是樣本比例數，常以 p 來表示。然後利用統計方法，由樣本統計是 p 對母群體參數 π 來作估計或檢定。

由一個比率為 π 的母群體中，每次抽取 n 個樣本，其中含有某項屬性之個數為 x 個，x/n 即為樣本的比率 p 。如果以同樣的方式做了 k 次以後，就可以得到 k 個 p，這 k 個 p 可以組成一個 p 的抽樣分配，我們稱為樣本比例數的抽樣分配。此分配是一個二項分配，其

平均數　$\mu_P = \sum p/k = \pi$

標準差　$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$

當樣本數 n 相當大時，此二項分配可以用常態分配來求近似解，也就是說，樣本比例數的分配會接近常態分配，至於樣本要多大才夠大，依經驗法則，當 nπ 與 n(1-π) 均等於或大於 5 時，我們即可用常態曲線來求解。但在實務上， π 是未知的，所以我們可以用樣本比例 p 來代替 π ，同時標準差 σ_p 可以用 $\sqrt{\frac{p(1-p)}{n}}$ 來計算。

依照抽樣分配原理，當間斷數據要以連續數據來求近似解時，正確的方式

應該要作「連續性校正」(continuity correction) ，但由於計算比例數的樣本一般都比較大，所以有無校正影響並不大，實務上均予忽略。

有關單一母群體比例數的估計及檢定，其原理與單一母群體平均數的方法一樣，只是用到的是比例數的抽樣分配，使用的公式如下：

1. 以常態分配來求近似解

2. $\mu_P = \pi$

   $\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$ 或 $\sigma_p = \sqrt{\frac{p(1-p)}{n}}$

3. π 的 $(1-\alpha)\%$ 信賴區間

$$p \pm Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

4. 檢定

$$Z_0 = (p - \mu_P) / \sigma_p = (p - \pi) / \sqrt{\frac{\pi(1-\pi)}{n}}$$

將 Z_0 與 $Z_{\frac{\alpha}{2}}$ 或 Z_α 比較，若落於拒絕區，則棄卻 H_1。