統計，讓數字說話！

Statistics

concepts and controversies

Chapter 8

推論:有信心的結論

2009/08/05

有些產品（比如藥物），必須有安全性和有效性的明顯證據；
法院在審訊共同起訴的歧視案件或其他法律程序中，
會問到統計顯著性；商人想知道新的廣告企畫是否明顯優於舊的，
而醫學研究者想知道新療法是否明顯有較好的效果。
在所有這些應用當中，統計顯著性很有價值，
因為統計顯著性會指出光靠機遇很難發生的效應。

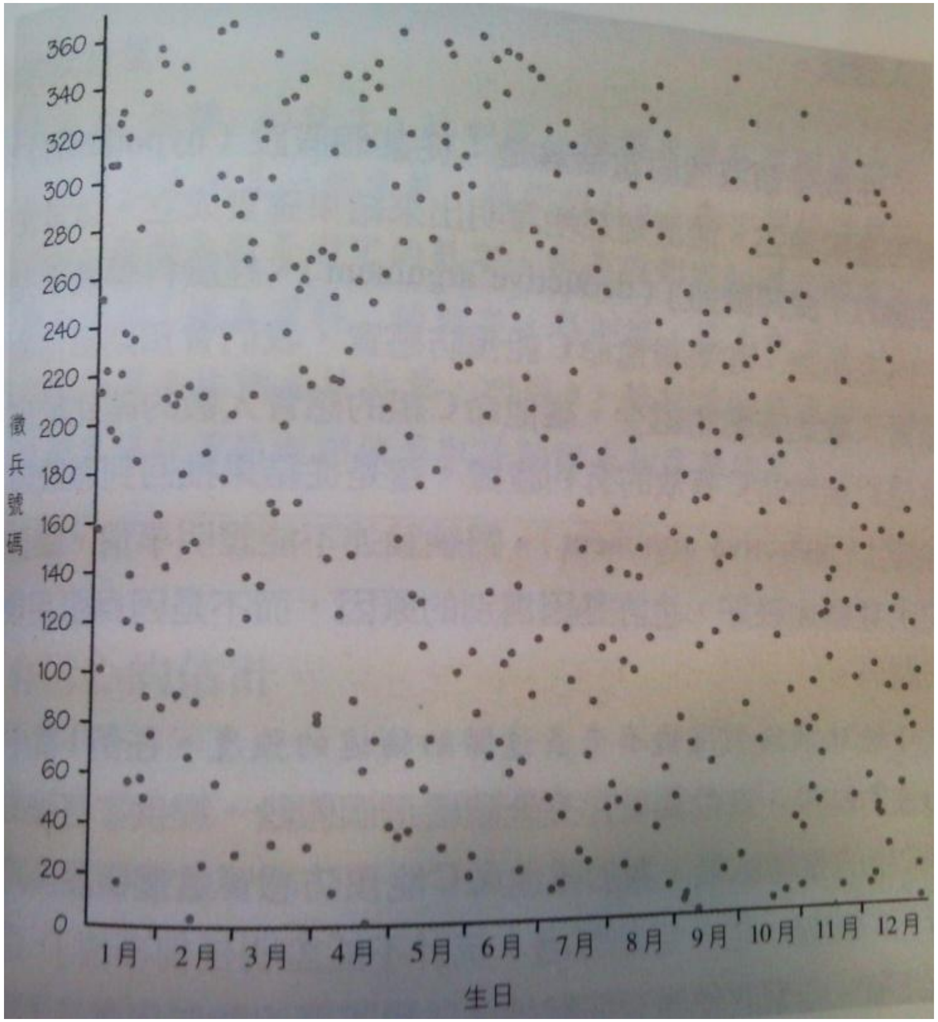
Ex. 徵兵抽籤公平嗎？ (1/3)

- **Ex.**

在越戰時，用**抽籤方法**決定男性徵兵**順序**。辦法是隨機抽取生日，來指定徵兵順序。隨機抽籤所產生的徵兵順序和生日的日期先後，應該沒有任何有系統的關聯。由徵兵的**散佈圖看不出什麼明顯的關聯**。然而，當我們計算生日和徵兵號碼之間的相關係數時，卻得到r=**-0.226**。即，生日比較靠**年尾的人**，似乎比較容易拿到**較小的徵兵號碼**。這個小小的相關係數是不是**抽籤不真正隨機的證據呢？**

Ex. 徵兵抽籤公平嗎？ (2/3)

- Ex. 散佈圖（生日vs. 徵兵號碼）



Ex. 徵兵抽籤公平嗎？ (3/3)

- **Sol：**

光是這樣，沒辦法判斷，原因是，只由於機遇，任何兩個變數就會有某些關聯。我們因此做了一些機率計算，發現，在真正隨機的抽籤中，要得到離開0這麼遠的相關係數之機率，小於0.0001。因為這麼強的相關係數幾乎永遠也不會在真正隨機的抽籤中發生，因此有強烈的證據指向抽籤不公平。

參數與統計量

- 參數（parameter）：

  描述母體的數字。ex. 母體中擁有某個我們感興的特質之比例是一個參數，稱為p。在一個統計推論問題中，母體參數是固定的數字，但是我們不知道它的值。

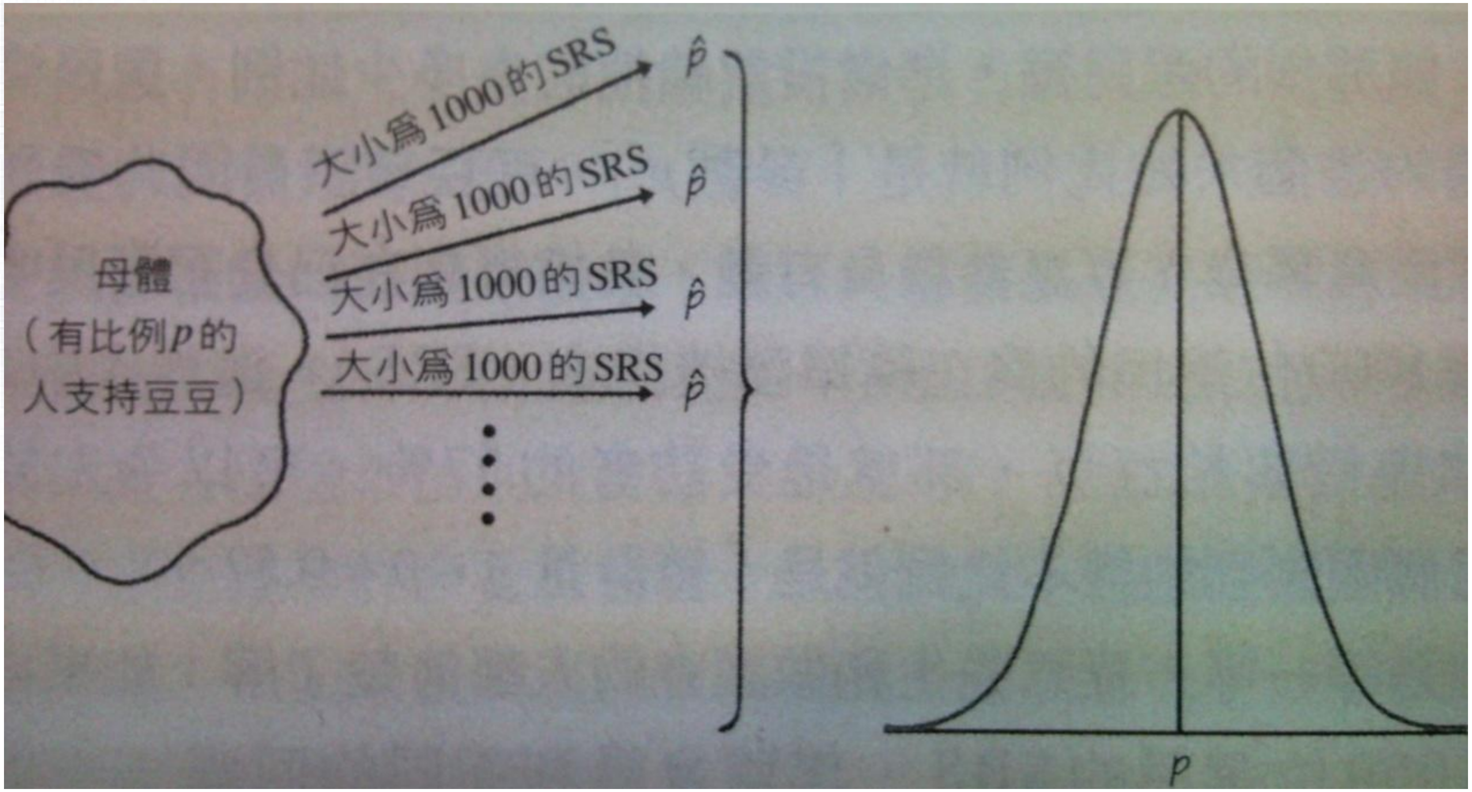
- 統計量（statistic）：

  描述樣本資料的數字。ex. 樣本中某個我們感趣的特質之比例是統計量，叫做$\hat{p}$ 。統計量的值隨樣本而變。我們用統計量的觀測值來取得關於未知參數的資訊。

抽樣分佈

- 抽樣分佈：

  從同一個母體抽許多樣本。對每一個樣本記錄統計量的值（樣本比例 $\hat{p}$）。這些值的分佈就是統計量的抽樣分佈。



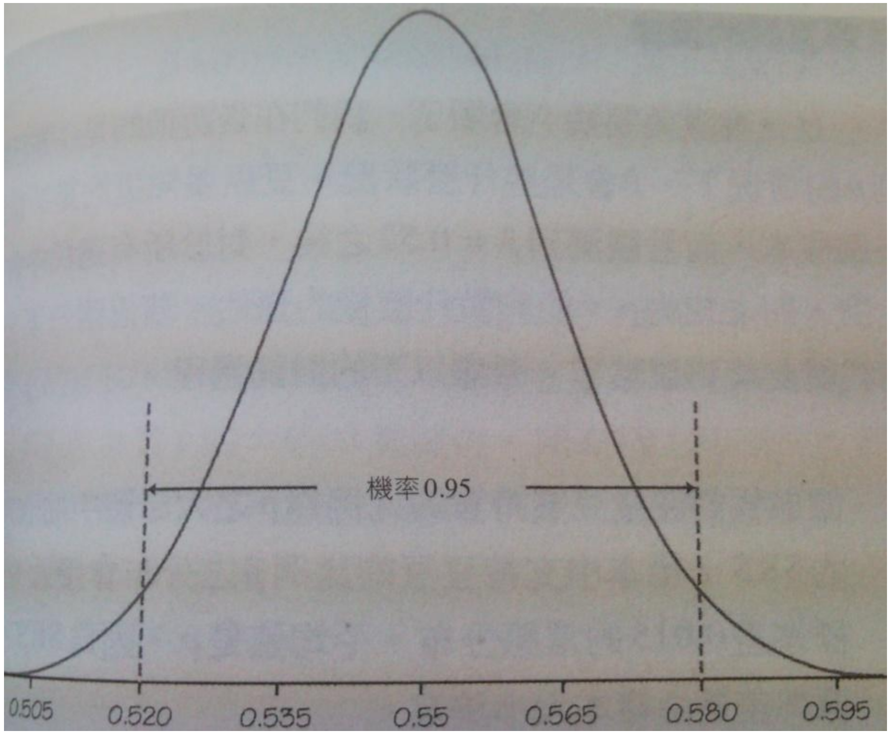
Ex. 抽樣分佈

- **Ex.**
  假設豆豆那州的幾百萬選民當中，事實上有**55%要投票給他**。許多大小各為**1000的獨立簡單隨機樣本（SRS）**所得到的支持豆豆之樣本比例的值，會有怎樣的型態（即樣分佈）呢？

Ex. 抽樣分佈 (Cont.)

- **Sol.**

  大小為1000的SRS中，支持參議員豆豆的比例$\hat{p}$的抽樣分佈。母體的真實情況是有**p=0.05**的比例支持豆豆。因此抽樣分佈的中心點是在**0.55**。$\hat{p}$落在**p**左右兩個標準差範圍內的機率是**0.95**。



Ex. 豆豆參議員的信賴區間

- **Ex.**

豆豆的1000個選民之SRS得到了$\hat{p}$=0.57。所有這樣子的樣本當中，有**95%得到區間 $\hat{p}\pm 0.03$會涵蓋到未知的母體比例*p***。所以，豆豆參議員有95%的信心，選民中支持他的真正比例，會落在

$$\hat{p}-0.03=0.57-0.03=0.54$$

和

$$\hat{p}+0.03=0.57+0.03=0.60$$

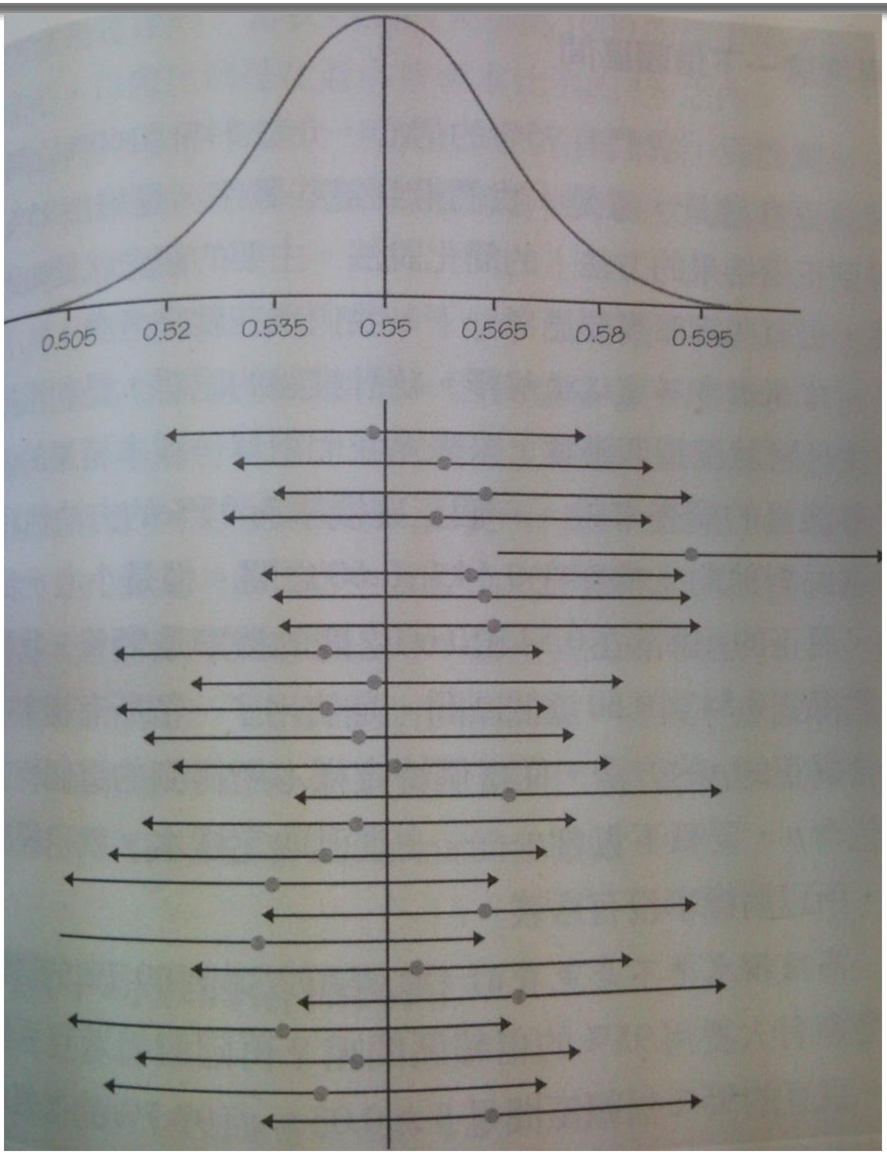
之間的區間內。即，議員有**95%**的信心，**母體當中有介於54%~60%之間的比例支持他**。

信賴區間

- 95%信賴區間是用某種方法從樣本資料得到的區間。經由該方法，所有樣本中的95%會產生包含真正母體參數的區間。
- 稱95%為信賴水準（confidence level）。這是該方法產生的區間會抓到真正參數的機率。
- 95%的信心是「我得到這個結果，是用95%的時侯會得到正確結果的方法」的精簡說法。

信賴區間 (Cont.)

- 重複抽樣所得到信賴區間的變化。這些區間從25個SRS所得到母體比例p的95%信賴區間。長期來說，這樣的區間當中有95%會包含*p*的真正值。



再考慮一下信賴區間

- 信賴水準不能這樣解釋：

  統計推論的語言，是利用長期下來會發生什麼況狀這個事實，來表達我們對單一樣本結果的信心。

- 高信賴水準不是免費的：

  如果要從同一個樣本要求較高的信賴水準，我們就必須願意接受較大的誤差界限（較寬的區間）。

- 較大的樣本會產生較窄的區間：

  樣本大小增加時，樣本統計量的精確度也會增加。抽樣分佈的標準差變小了。如果，信賴水準固定，則樣本愈大，信賴區間就愈短。

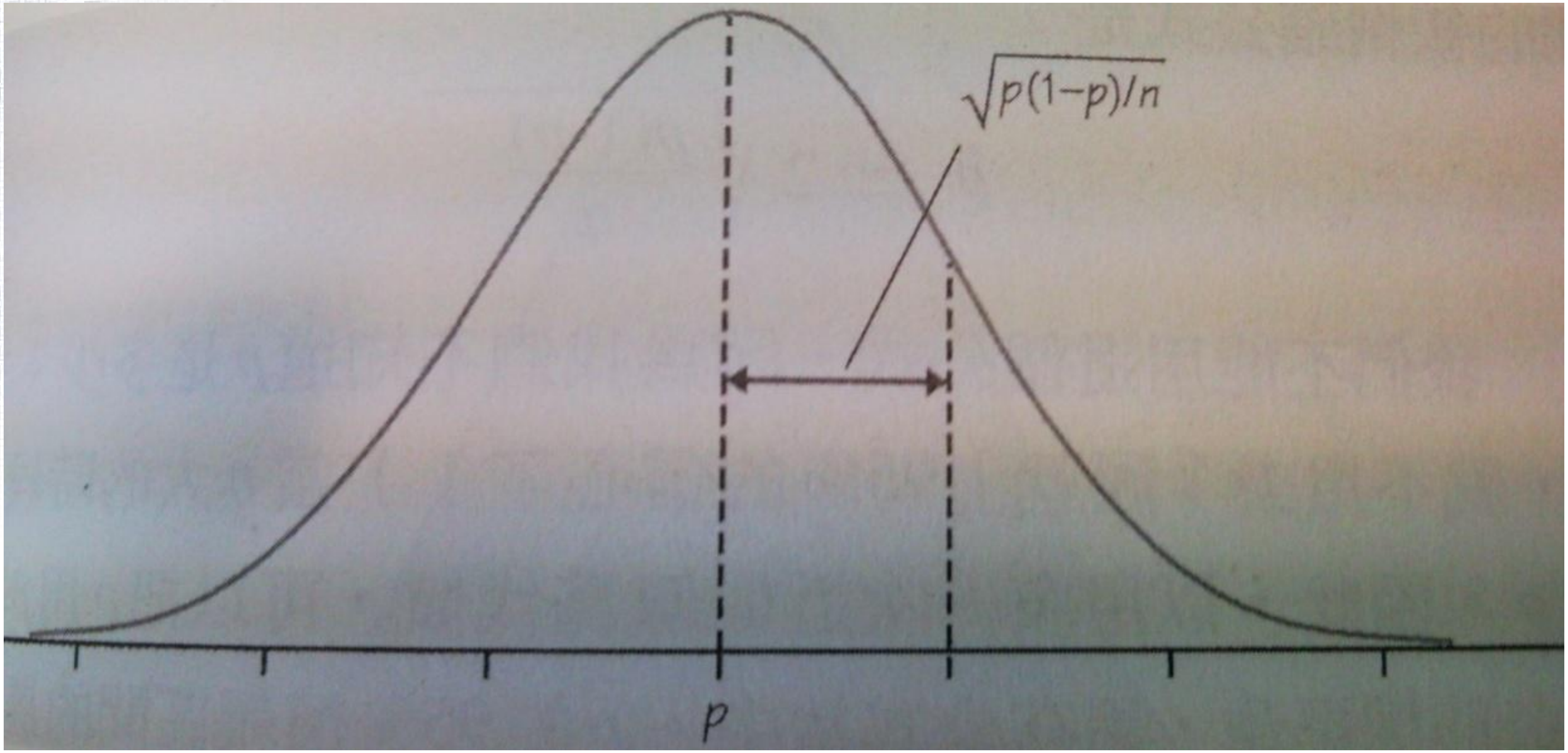
樣本比例的抽樣分佈

從大母體當中抽取大小為*n*的SRS，母體當中擁有我們感興趣的某種特質之比例是*p*。$\hat{p}$表示樣本當中擁有該特質的比例。則：

- 當樣本大小*n*夠大時，$\hat{p}$的分布為近似常態（approximately normal）
- 抽樣分布的平均數和*p*相等。

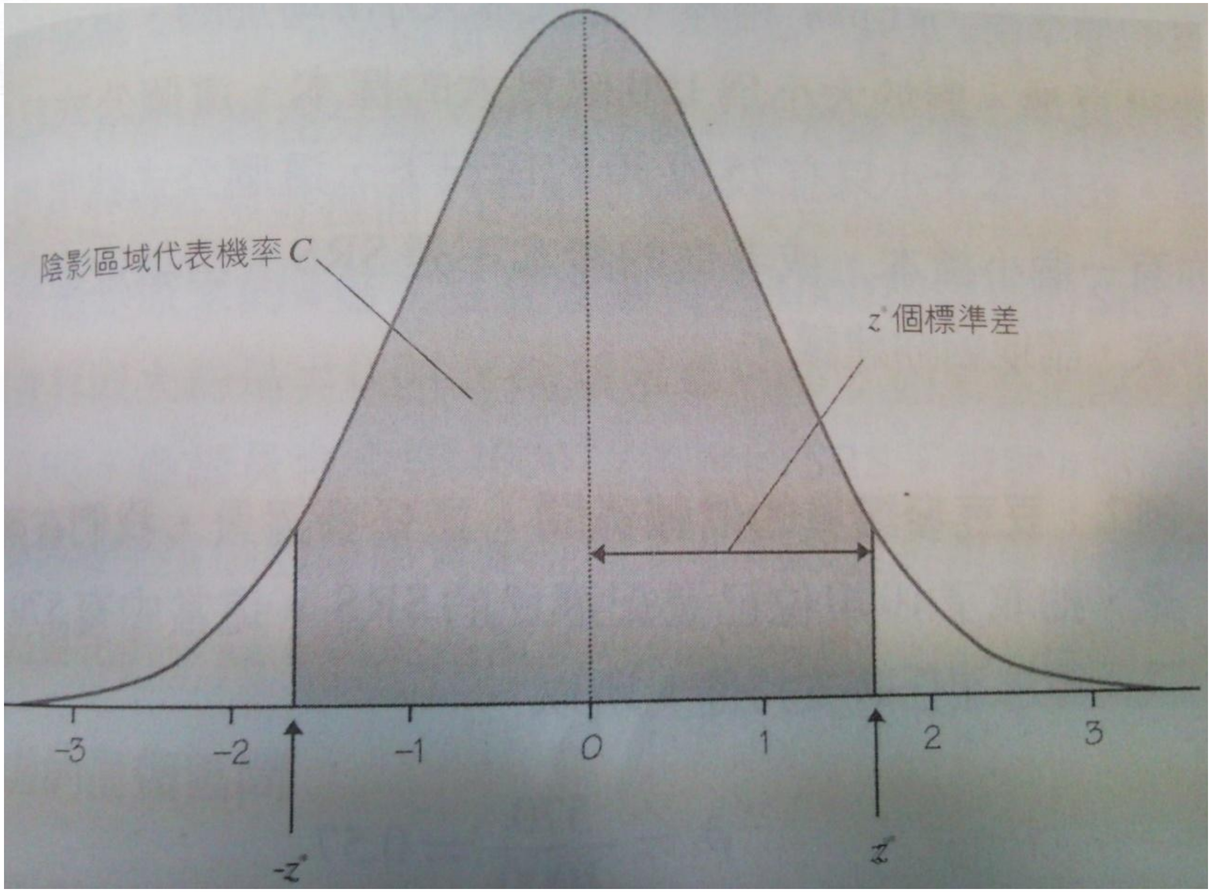
樣本比例的抽樣分佈 (Cont.)

- 樣本比例 $\hat{p}$的抽樣分布，樣本大時，$\hat{p}$的分布近似常態分佈。



其他信賴水準的信賴區間

- 常態分布的臨界值。在任何常態分布當中，平均數的 z^* 個標準差範圍內的機率是C。



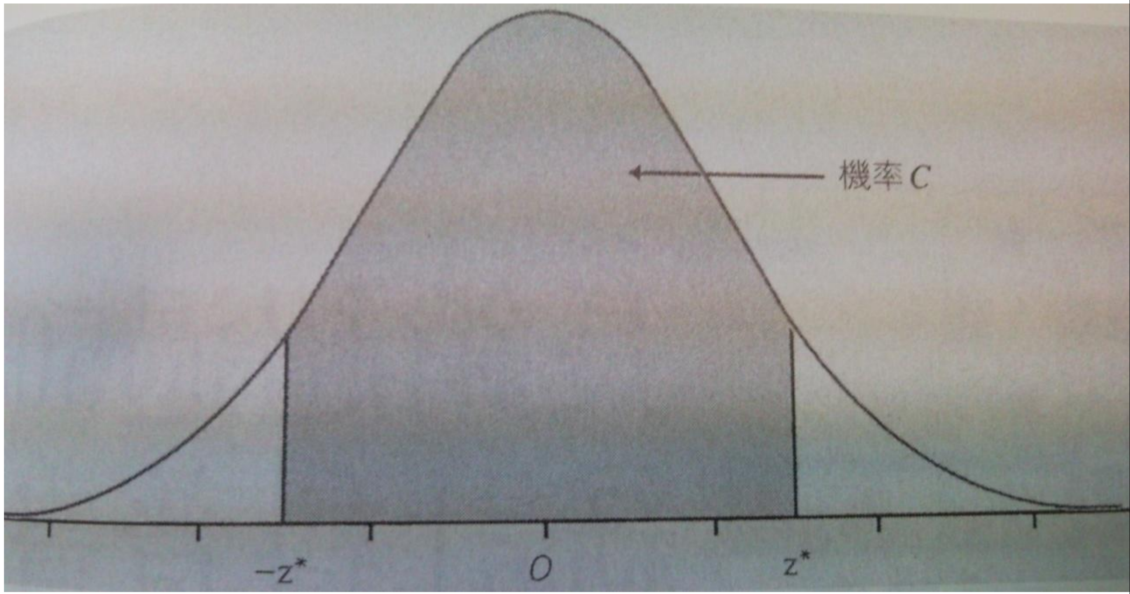
母體比例的信賴區間

- 母體中某個我們感興趣的特質之比例p，從母體抽取大小為n的SRS。當n大的時侯，p的近似水準C信賴區間為

$$\hat{p} \pm z^{*}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

此處 z^{*} 為機率C對應的臨界值，從表C所得。

常態分佈臨界值



表C　常態分布的臨界值

C	z^*	C	z^*
0.50	0.67	0.80	1.28
0.55	0.76	0.85	1.44
0.60	0.84	0.90	1.64
0.65	0.93	0.95	1.96
0.70	1.04	0.99	2.58
0.75	1.15	0.999	3.29

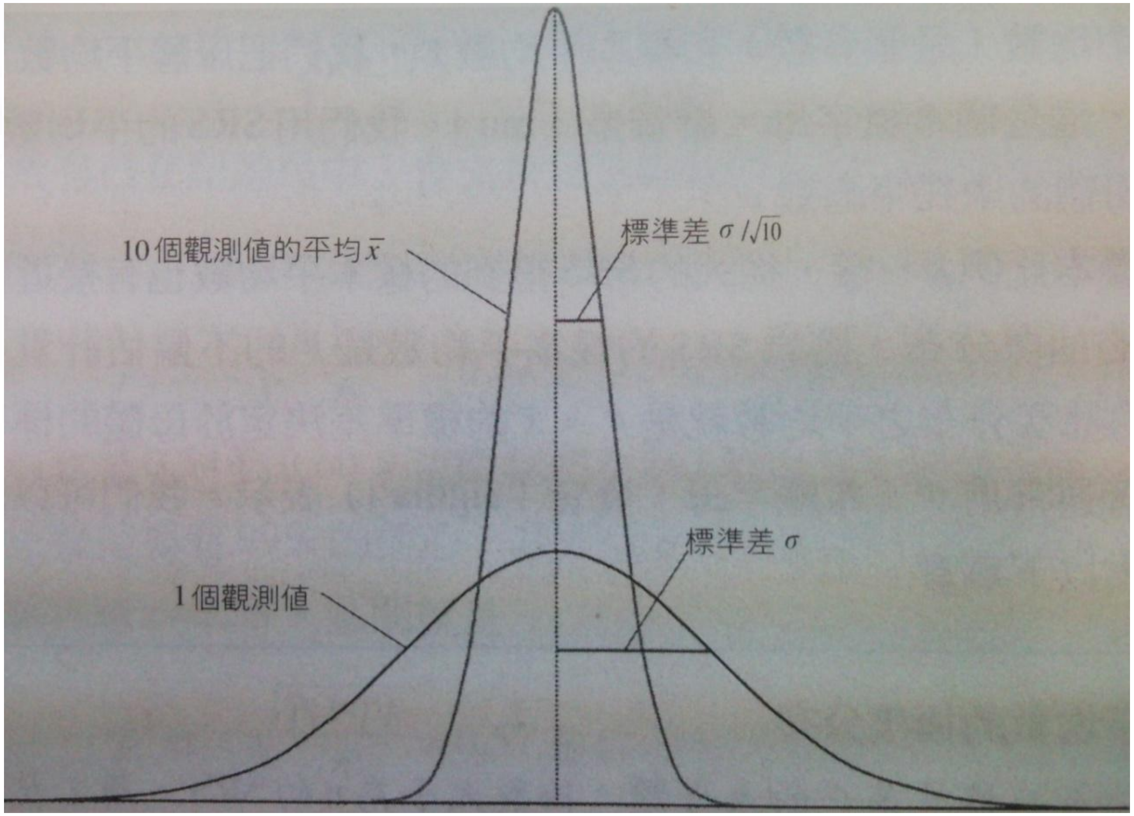
樣本平均數的抽樣分布

從平均數μ標準差σ的大母體，抽取大小為n的SRS。用 $\bar{x}$ 表示樣本平均數。則：

- 當樣本大小n較大時，$\bar{x}$ 的抽樣分布為近似常態。
- 抽樣分布的平均數等於μ。
- 抽樣分布的標準差是 $\sigma/\sqrt{n}$ 。

樣本平均數的抽樣分布 (Cont.)

10個觀測值的平均數之抽樣分布和1個觀測值之分布的比較。如果個別觀測值的標準差是σ，則10個觀測值的平均，其標準差就是 $\sigma/\sqrt{10}$ 。



母體平均數的信賴區間

- 從個體平均數為μ的母體中，抽取大小為n的SRS。當n大的時候，μ的近似水準C信賴區間為

$$\bar{x} \pm z^* \frac{s}{\sqrt{n}}$$

此處 z^* 為機率C對應的臨界值，從表C所得。

Ex. 牛奶裡的細菌數

- **Ex.** 為了做一個關於鮮奶被細菌污染狀況的研究，於美國東岸雜貨店裡購買了100件鮮奶當做樣本，並計算了每毫升的細菌數。資料如下：

5	8	6	7	8	3	2	4	7	8	6	4	4	8	8	8	6	10	6	5
6	6	6	6	4	3	7	7	5	7	4	5	6	7	4	4	4	3	5	7
7	5	8	3	9	7	3	4	6	6	8	7	4	8	5	7	9	4	4	7
8	8	7	5	4	10	7	6	6	7	8	6	6	6	0	4	5	10	4	5
7	9	8	9	5	6	3	6	3	7	1	6	9	6	8	5	2	8	5	3

Ex.牛奶裡的細菌數 (Cont.)

- **Sol. 該地區鮮奶的平均細菌數*μ*的90%信賴區間。**
- **樣本平均數：$\bar{x}$ =5.88；樣本標準差：*s*=2.02**
- **90%信賴區間是**

$$\bar{x} \pm z^* \frac{s}{\sqrt{n}} = 5.88 \pm (1.64)\frac{2.02}{\sqrt{100}} = 5.88 \pm 0.33$$

- **我們有90%的信心，整個鮮奶母體的平均細菌數，在每毫升5.55和6.21個之間。**

統計顯著性

- 統計推論利用樣本資料來對該筆資料代表的母體做結論。
- 推論的背後的依據是什麼？
  經過這麼多次試驗，會發生什麼狀況？
  Ans：根據抽樣分布，用機率表示。
- 信賴區間用來估計一個未知參數的值-抽許多樣本，得到大部分的區間會抓到真正的參數值。
- 統計檢定（statistical test）的目的：評估資料提供的對某一參數斷言的不利證據。
- 檢定：如果我們取許多樣本而且斷言正確，我們很少會得到這樣的結果。

Ex. 咖啡是現煮的嗎？

- Ex. 一位持懷疑態度的人斷言：喝咖啡的人裡，只有一半偏好現煮咖啡。

- 50位受測者（二樣本）： 0.72（36/50）比率偏好現煮
  0.56（28/50）比率偏好現煮

→可否斷言72%>56%是強烈證據？

→是不是可以當做母體中有大部分人喜歡現煮咖啡的信服證據呢？

原始假設

原始假設（**null hypothesis**）H_0：

- 在統計檢定中檢驗的敘述叫做原始假設。顯著性檢定是設計來評估：否定原始假設的證據有多強。通常原始假設是「沒有效應」或「沒有差別」的敘述。

- **Ex.** 所有喝咖啡的人裡面偏愛現煮咖啡的比例 p。

  原始假設為：　H_0：$p=0.5$

  對立假設為：　H_1：$p>0.5$

P值

P值（p-value）：

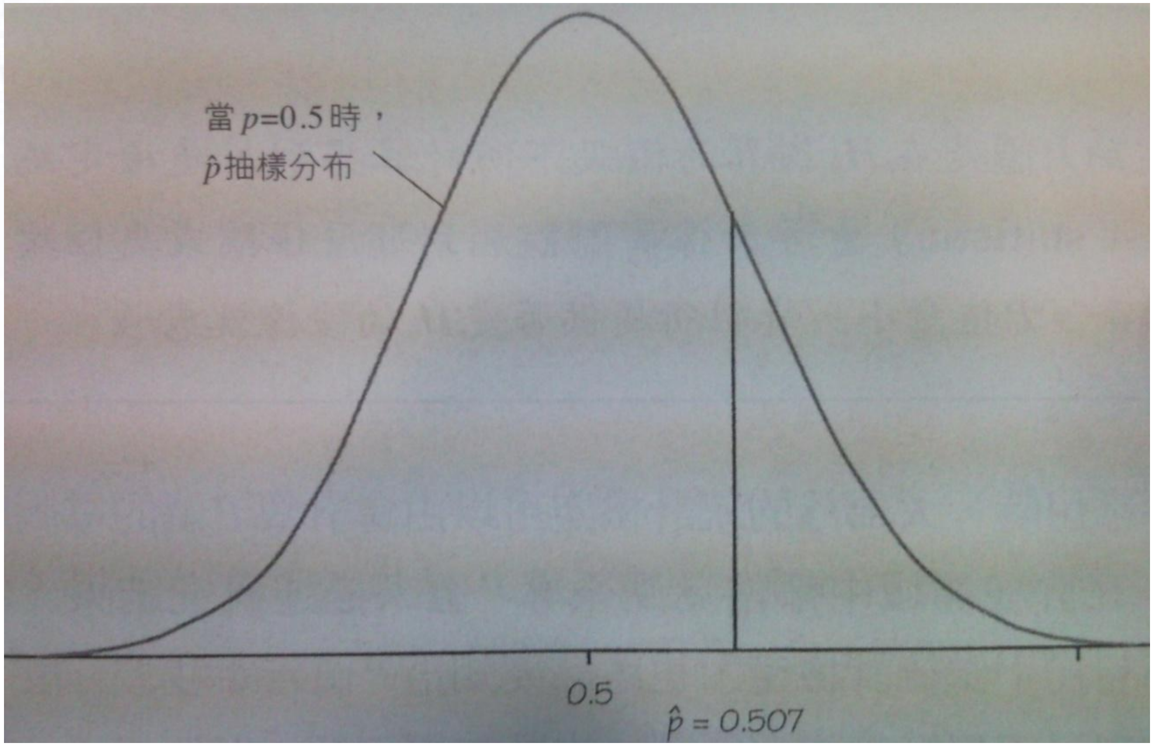
- 統計檢定的*p*值是在H_0為真的假設下所計算得到，是檢定統計量（test statistic）會等於像實際觀測到那麼極端或更極端的值之機率。*p*值愈小，資料所提供否定H_0的證據就愈強。

Ex. 布豐伯爵的銅板 (1/3)

- Ex. 法國自然主義者布豐伯爵擲了銅板4040次。得到2048正面。正面的**樣本比例**為2048/4040=0.507
- 原始假設銅板是平衡的： H_0： $p=0.5$
- 對立假設銅板是不平衡： H_1： $p \neq 0.5$

Ex. 布豐伯爵的銅板 (2/3)

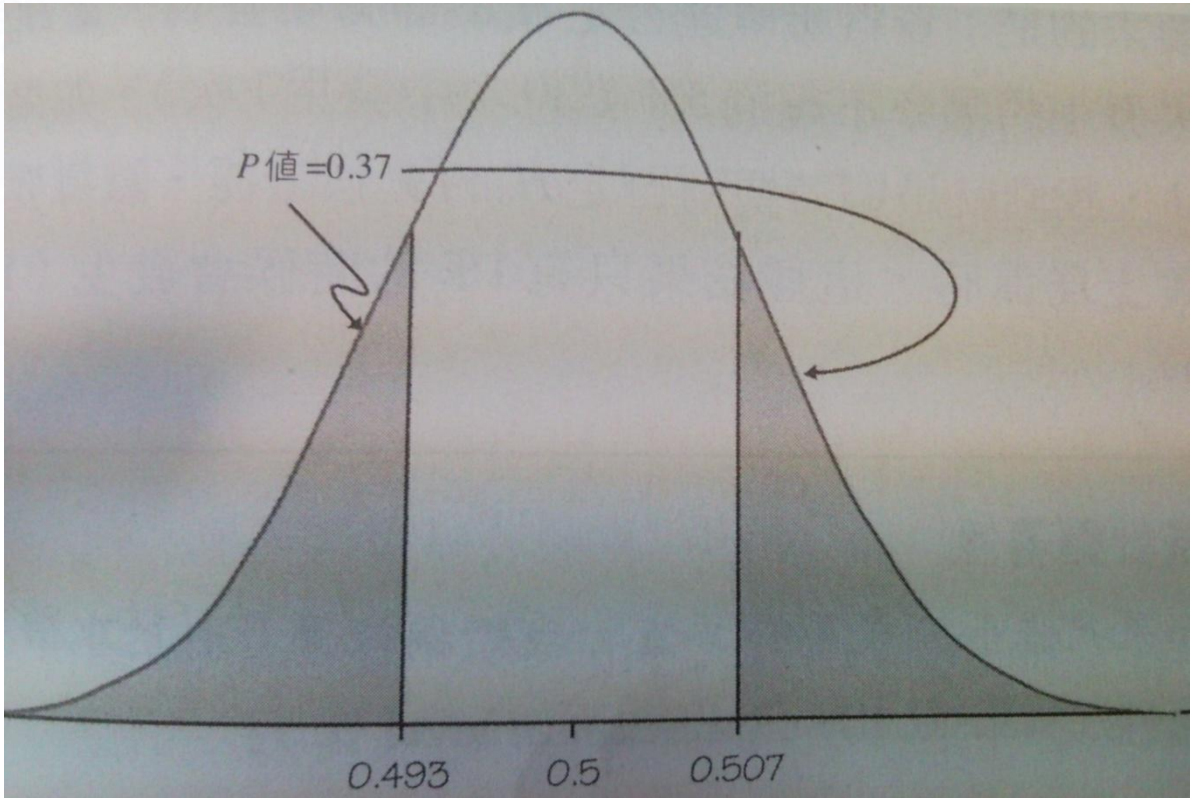
- 4040擲當中的正面比例之抽樣分布，接近平均數0.5；標準差0.008的常態分布。



- 樣本比例0.507，很容易在擲一個平衡銅板時因機遇而發生。

Ex. 布豐伯爵的銅板 (3/3)

- H_1：$p \neq 0.5$ 是雙邊對立假設（two sided alternative）。由對立假設是單邊還是雙邊決定了：樣本結果是往一個方向還是往兩個方向偏離，可以算做否定H_0而肯定H_1的證據。



統計顯著性

統計顯著性：

- 如果*P*值小於或等於α值，我們稱該筆資料於**水準α有統計顯著性**（statistically significant at level α ）。

- Ex. *P*值為0.03的結果於α=0.05水準有顯著性，但是在於α=0.01水準就沒有顯著性。

比例和平均數的顯著性檢定 (1/3)

- 執行這些檢定所遵循的步驟，通常包含兩個要素：

★度量我們在尋找的效應之統計量。
★可以算出對應的P值之該統計量的機率分配。

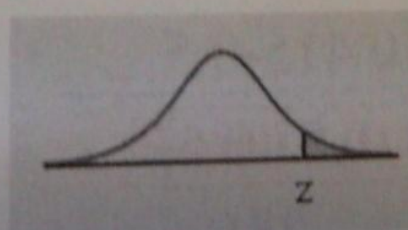
- 檢定現在通常由電腦程式執行。

比例和平均數的顯著性檢定 (2/3)

母體中某個我們感興趣的特質之比例為p，從母體抽取大小為n的SRS。若要對特定的值p_0檢定原始假設H_0：$p=p_0$，先計算標準計分

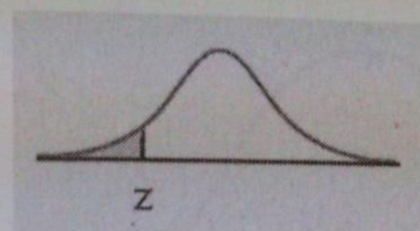
$$z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

H_0對應H_a：$p>p_0$檢定的近似P值，是在常態曲線之下、平均數之右超過z個標準差的面積。

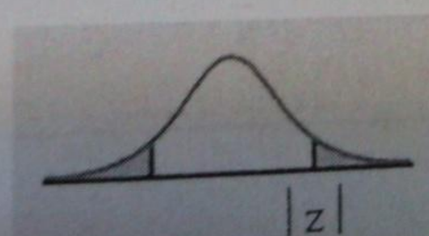


比例和平均數的顯著性檢定 (3/3)

H_0對應H_a：$p<p_0$檢定的近似P值，是在常態曲線之下、在平均數之左超過z個標準差的面積。



H_0對應H_a：$p\neq p_0$檢定的近似P值，是在常態曲線之下、離平均數超過z個標準差的面積，左右兩方向都算。



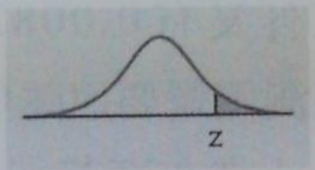
母體平均數的檢定

從母體中取大小為n的SRS，母體平均數μ未知，標準差σ已知，若要對特定值μ_0檢定假定H_0：$\mu=\mu_0$，先計算標準計分。

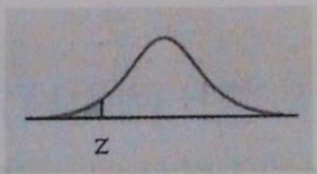
$$z=\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$$

母體平均數的檢定 (Cont.)

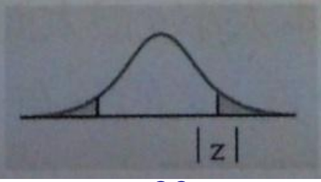
則 H_0 對應 H_a：$\mu > \mu_0$ 檢定的近似 P 值，是在常態曲線之下、平均數之右超過 z 個標準差的面積。



H_0 對應 H_a：$\mu < \mu_0$ 檢定的近似 P 值，是在常態曲線之下、平均數之左超過 z 個標準差的面積。



H_0 對應 H_a：$\mu \neq \mu_0$ 檢定的近似 P 值，是在常態曲線之下、往任一方向離開平均數超過 z 個標準差的面積。



真正了解統計顯著性的意義

統計顯著性及實際的顯著性

- 如果我們可以取得很大的樣本，即使只是具原始假設只有微小的偏離，也會得到顯著性。

選擇顯著水準

- 對於樣本所提供的否定原始假設之證據強度，給一個清楚的敘述，這就是顯著性檢定的目的，P值就是這個作用。

Thanks for your attention