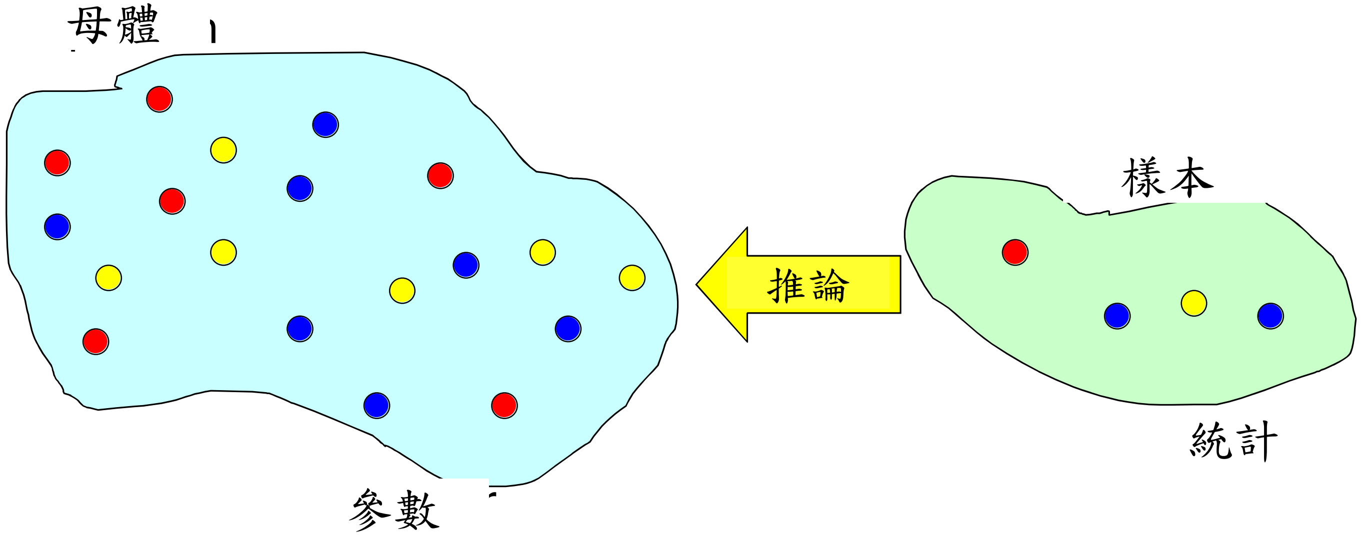
第 12 章

單一母體的推論

對單一母體的推論



我們將發展估計與檢定三個母體參數的方法：

母體平均數 μ

母體比例 p

Preliminaries

- Recall estimation and hypothesis testing in Chapters 10 and 11 for a population mean with the known population variance.
- For the interval data, the parameters of interest are the population mean and variance.
- In this chapter, the inference on the population parameters is the main focus.
  - Section 12.1: population mean with the unknown population variance (interval data);
  - Section 12.2: population variance (interval data);
  - Section 12.3: population proportion (categorical data)
  - Section 12.4: the application of marketing survey-市場區隔

12.1 在母體標準差未知下進行單一母體平均數的推論

- 以前，我們著重於當母體標準差已知或給定時，估計與檢定母體平均數：

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

但是我們如何知道實際的母體變異數?

我們以學生 t 統計量取代，給定：

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

The above t-statistic has a Student t distribution (or t distribution) with degrees of freedom n-1.

在母體標準差未知下進行單一母體平均數的推論

- 當**母體標準差未知**且**母體是常態**時，我們使用其點估計 ***s***

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \quad \Longrightarrow \quad t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

t- 統計量取代了 *z*- 統計量，其自由度 ν 為 n–1 。

當 σ 未知時，μ 的檢定統計量

- 當**母體標準差未知**且**母體是常態**時，有關假設檢定的檢定統計量是：

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

它是自由度為 $v = n - 1$ 的學生t 分配，μ 的信賴區間估計量：

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

範例 12.1 報紙回收計劃

- 在未來可能很多國家必須更努力於拯救環境。
- 人們的行動包括減少使用能源和加強回收。
- 目前 (2007年)，大部分由回收材料所製造的產品比起那些由地球原始材料所製造的產品明顯地比較昂貴。
- 報紙例外。回收報紙可能是有利可圖的。
- 從各家庭回收報紙是一項主要的費用。最近，數家公司加入蒐集家庭過期報紙並再利用它們的行業。

範例 12.1

- 一位財務分析師為這類的公司計算，如果平均每星期從每戶家庭蒐集的報紙超過2.0磅，則公司將可以賺取利潤。
- 一項研究決定回收場的可行性，從一個大型的社區中抽取一個148戶家庭的隨機樣本，每星期每一戶家庭回收丟棄的報紙重量被記錄。Xm12-01*
- 這些資料是否提供充分的證據讓這位分析師下結論回收計畫是有利可圖的？
- Raw data, see p423. Could you find any useful information from raw data? Hard to see……..

辦識方法

範例 12.1

我們的目的是**描述 (describe)** 每一戶家庭丟棄報紙總量的母體。資料的尺度是區間的，表示母體平均數 μ 是需要被檢定的參數。

我們想知道是否有足夠的證據決定平均數大於2.0磅，因此：

$$H_1: \quad \mu > 2.0$$

因此，虛無假設中的數值是

$$H_0: \mu = 2$$

辨識方法

範例 12.1

檢定統計量是：

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \qquad \nu = n - 1$$

因為對立假設是：

$$H_1: \quad \mu > 2.0$$

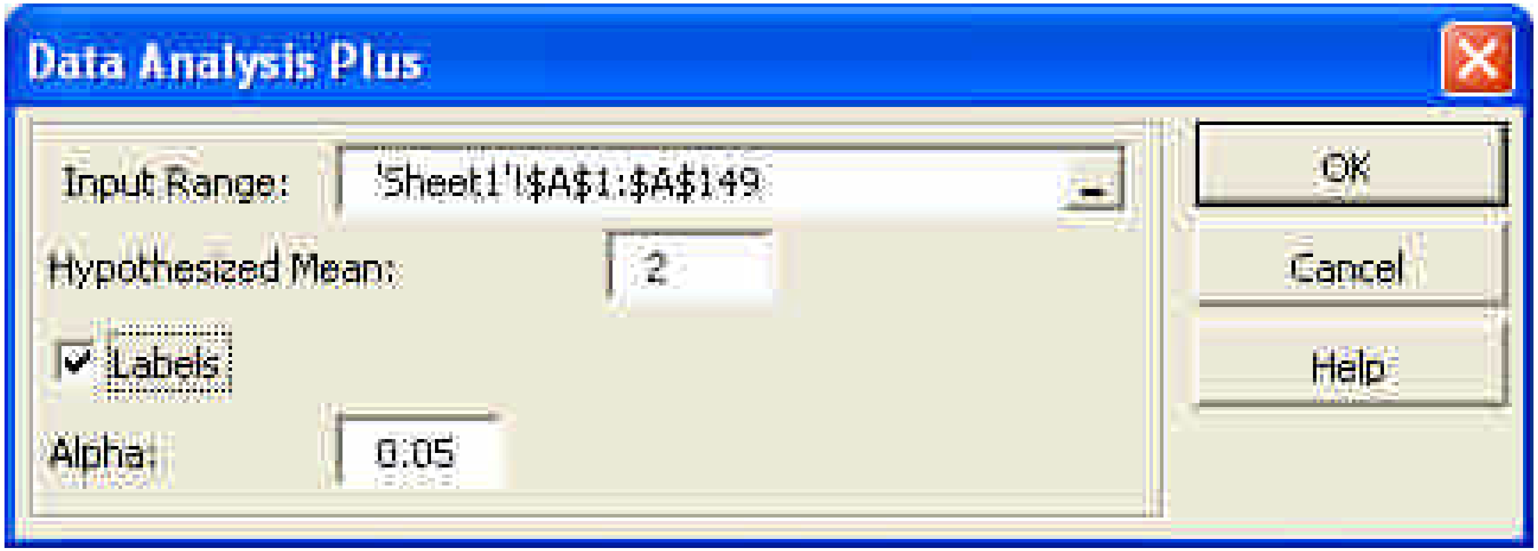
拒絕域則為：

$$t > t_{\alpha,\nu} = t_{.01,148} \approx t_{.01,150} = 2.351$$

範例 12.1

計算

點選 **Add-Ins**、**Data Analysis Plus**，與 **t-Test: Mean**



範例 12.1

計算

	A	B	C	D
1	**t-Test: Mean**			
2				
3				Newspaper
4	Mean			2.18
5	Standard Deviation			0.98
6	Hypothesized Mean			2
7	df			147
8	t Stat			2.24
9	P(T<=t) one-tail			0.0134
10	t Critical one-tail			2.3520
11	P(T<=t) two-tail			0.0268
12	t Critical two-tail			2.6097

詮釋

範例 12.1

檢定統計量的值是 $t = 2.24$ 而且它的 p- 值是.0134。

沒有充分的證據去推論被丟棄報紙的平均重量是大於 2.0的。

注意證據是有一些的；p- 值是 .0314 。然而，因為我們想要犯型 I 錯誤的機率小一點，我們堅持要求1%的顯著水準。

因此，我們不能下結論說回收場將會有利可圖。

範例 12.2 審核報稅的稅收

- 在 2004年 (最新年份的報告)，在美國提出的報稅共有130,134,000件。
- 美國國稅局 (Internal Revenue Service, IRS) 審查其中的0.77%或1,008,000件以確定他們是否正確報稅。
- 為了判斷審計人員的執行績效如何，從這些報稅中選出一個隨機樣本，並且報告其附加的稅收。

Xm12-02

- 以95%的信心估計從1,008,000件審查的報稅中平均徵收的附加所得稅。
- Raw data, see p425.

範例 12.2 審核報稅的稅收

辨識方法

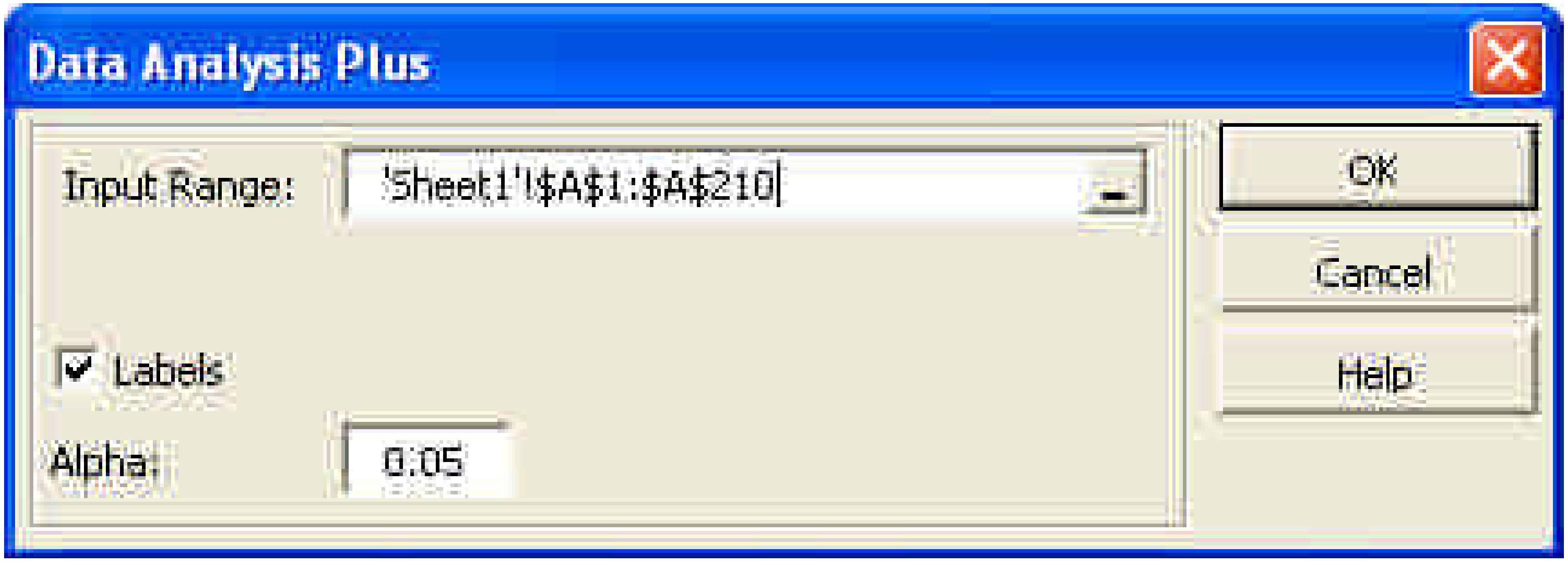
- 問題的目的是描述附加所得稅的母體。
- 資料的尺度是區間的。
- 此問題要求我們估計的參數是附加所得稅的母體。
- 信賴區間的估計量是

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

計算

範例 12.2 審核報稅的稅收

點選 **Add-Ins**、**Data Analysis Plus**，與 **t-Estimate: Mean**



範例 12.2 審核報稅的稅收

計算

	A	B	C	D
1	**t-Estimate: Mean**			
2				
3				Taxes
4	Mean			6001
5	Standard Deviation			2864
6	LCL			5611
7	UCL			6392

詮釋

範例 12.2 審核報稅的稅收

我們估計所徵收的平均附加所得稅是介於$5,611與$6,392之間。

我們可以使用這項估計值去決定是否IRS審核了應該被審核的個人。

檢查必要的條件

- 學生 *t* 分配是**穩健的**，指的是假如母體不是常態，*t*-檢定的信賴區間估計值仍然是合適的，只要母體「不是**極端地**非常態」。

- 為了檢查這個必要的條件，我們**繪製直方圖**來決定它們是否嚴重地偏離鐘形分配。假如直方圖極端地(假定在一個指數分配的情況下)，其可能被視為「極端地非常態」，因此 *t*- 統計量在這範例中不是合理的。

圖12.2　範例12.1的直方圖

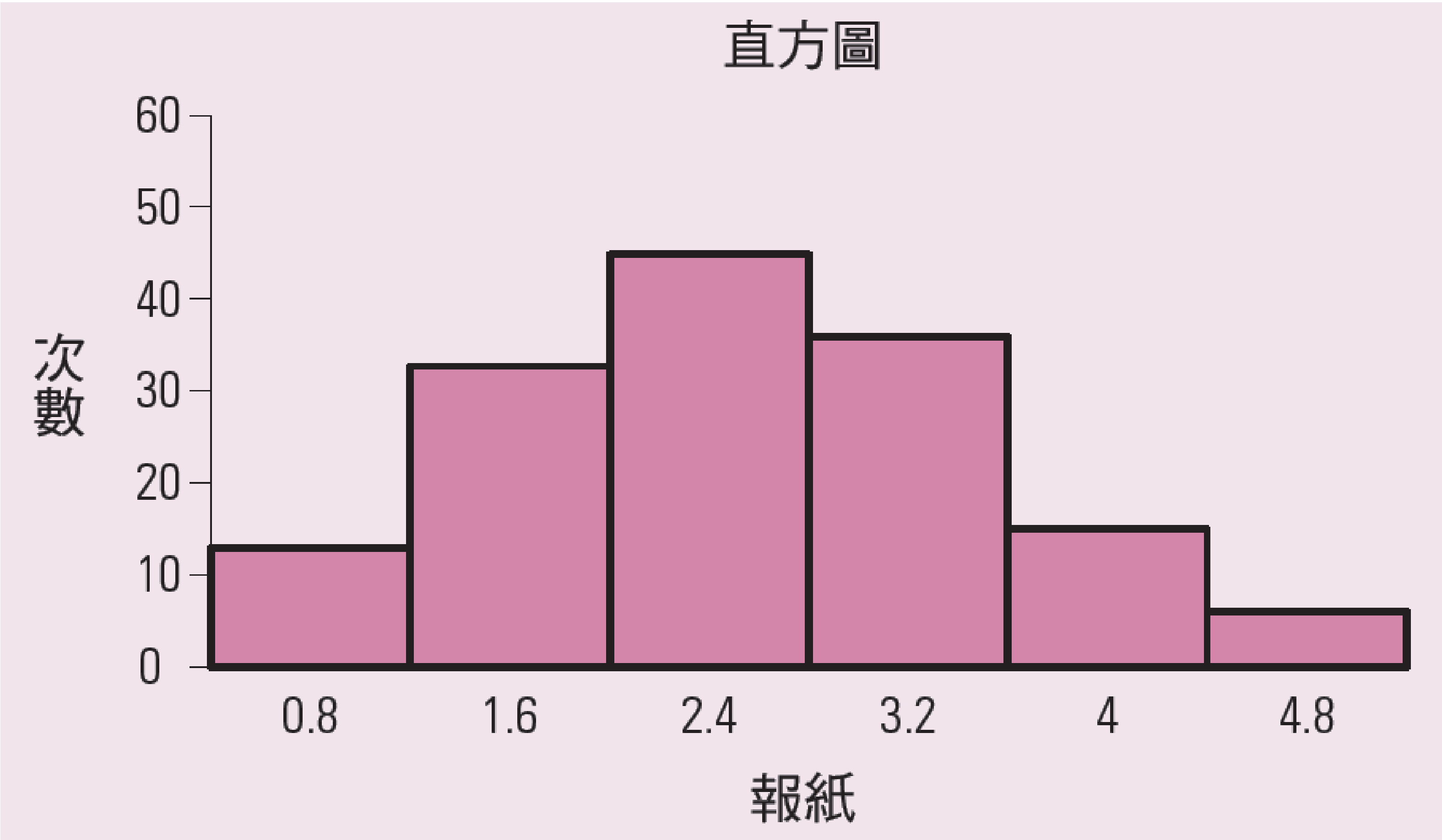
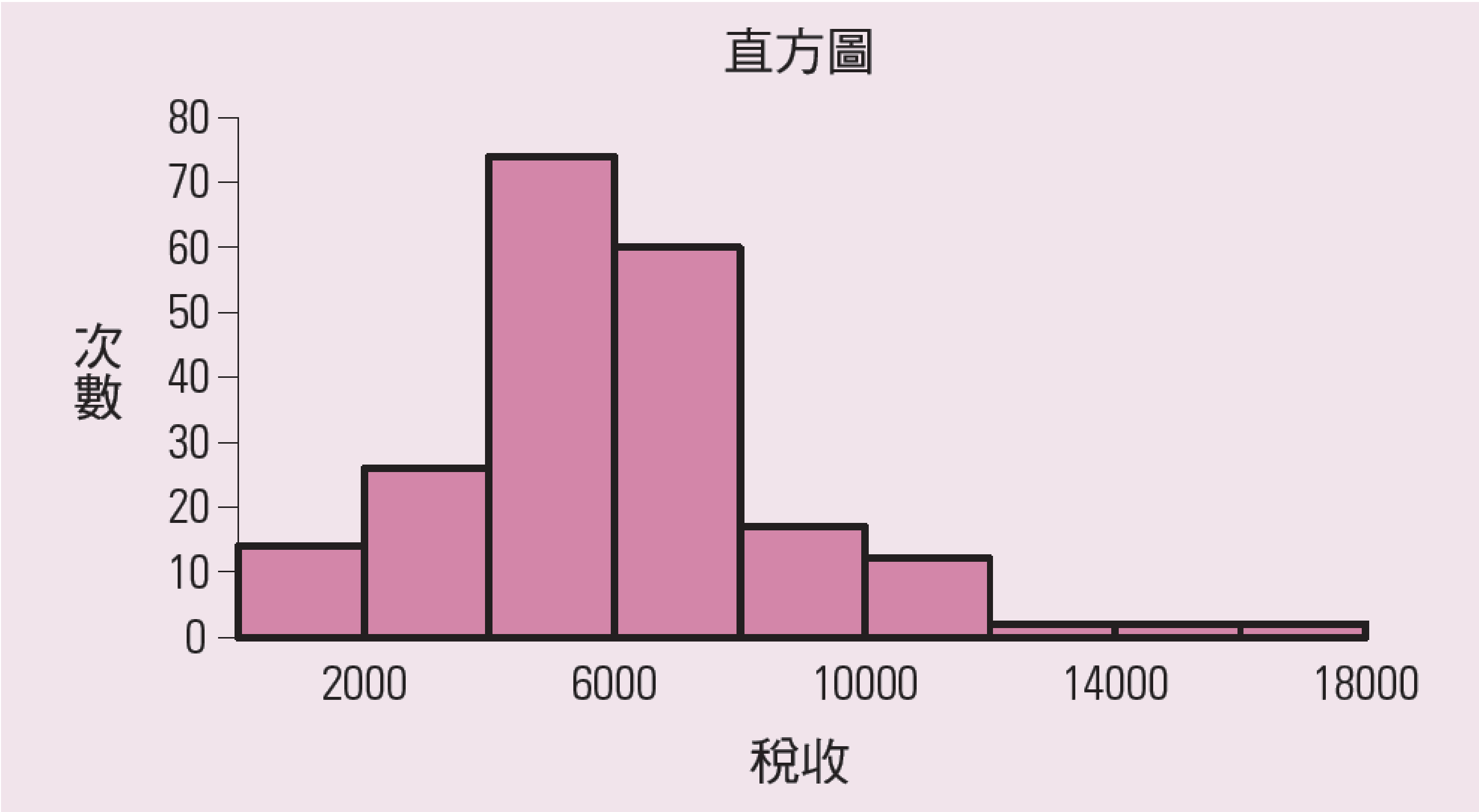


圖12.3　範例12.2的直方圖



估計有限母體的總和

- 到目前為止，所介紹的推論方法都是在母體無限大的假設下推導出的。然而，實務上大多數的母體都是有限的。

- 當母體是小的，我們必須使用在第9章所介紹的有限母體校正因子來調整檢定統計量和區間估計量。
- 然而，當母體相對於樣本大小是夠大的，我們可以忽略校正因子。大母體的定義是母體至少是樣本大小的20倍。

估計有限母體的總和

- 有限母體容許我們使用一個平均數的信賴區間估計量去產生母體總和的信賴區間估計量。
- 要估計總和，我們將平均數信賴區間估計值的上下限乘以母體的大小。
- 總和的信賴區間估計量是

$$N\left(\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}\right)$$

估計有限母體的總和

例如，假設我們想要估計從1,008,000份審核的報稅中，所能徵收的附加所得稅總額。總額的 95% 信賴區間估計值是

$$N\left(\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}\right) = 1{,}008{,}000\ (6{,}001 \pm 391)$$

它是

LCL= 5,654,880,000　　和　　UCL= 6,443,136,000

發展對統計觀念的了解

- 與 z-統計量相同，t-統計量以相當於幾個標準誤的方式，衡量樣本平均數與假設值之間的差異。

  然而，當母體標準差是未知時，我們以 $s/\sqrt{n}$ 估計標準誤。

發展對統計觀念的了解

當我們在 8.4 節中介紹學生 t 分配，我們指出它比標準常態分配來的分散。

這個狀況是合乎邏輯的。

在z-統計量的唯一隨機變數是樣本平均數，其隨著樣本的不同而改變。

發展對統計觀念的了解

t- 統計量則有兩個變數，樣本平均數 $\bar{x}$ 和樣本標準差 *s*，兩者皆會隨著不同的樣本而改變。

因為比較大的不確定性，*t*- 統計量將會呈現較大的變異性。

辨識因素

- 辨識 *t*- 檢定與 μ 的統計量的因素：

1. 問題目的：描述一個母體。
2. 資料類型：區間。
3. 敘述性測量的類型：中央位置。

12.2 單一母體變異數的推論 (normal population)

假如我們對推論母體的變異性感興趣，我們必須調查的參數是**母體的變異數** σ^2。

樣本變異數 s^2 是 σ^2 的一個不偏的、一致的估計量。此外， σ^2 的檢定統計量

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

被稱為卡方統計量 (**chi-squared statistic**, χ^2-statistic)，它會服從自由度為 $\nu = n - 1$ 的卡方分配。Normal population is required for卡方統計量。

檢定與估計一個母體變異數

用於 σ^2 假設檢定的檢定統計量是：

$$\chi^2 = \frac{(n - 1)s^2}{\sigma^2}$$

它會服從自由度為 $\upsilon = n - 1$ 的卡方分配。

檢定與估計一個母體變異數

我們可以做下列機率敘述：

$$P\left(\chi^2_{1-\alpha/2} < \chi^2 < \chi^2_{\alpha/2}\right) = 1 - \alpha$$

代入

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

可以推導出一個母體變異數的信賴區間估計量 σ^2：

$$\text{(LCL)} = \frac{(n-1)s^2}{\chi^2_{\alpha/2}}$$

信賴下限

$$\text{(UCL)} = \frac{(n-1)s^2}{\chi^2_{\alpha/2}}$$

信賴上限

範例 12.3 容器填充機的一致性

- 容器填充機使用於填充各種的液體，包括牛奶、飲料與油漆等。
- 理想的情況是，各個瓶中的液體量應該只有些微的不同，因為較大的變異會導致有些容器裝填不足 (欺騙顧客) 而有些填過量(導致成本上的浪費)的情況。
- 一家開發新型機器公司的總裁誇耀他們的機器能夠非常一致地填充 1 公升(1,000 立方公分)的容器，使得裝填的變異數小於 1 立方公分。

範例 12.3 容器填充機的一致性

- 為了檢定此宣稱的真實性，選取 25 個填滿 1 公升的容器的隨機樣本，並且記錄結果。Xm12-03
- 這些資料是否容許該總裁在 5% 的顯著水準下做此宣稱。
- Raw data (填充量), see p436.

辨識方法

範例 12.3 容器填充機的一致性

問題的目的是描述這部機器裝填 1 公升容器的母體。

資料是區間尺度，而我們對裝填的變異性感興趣。

感興趣的參數是母體變異數 σ^2。

辨識方法

範例 12.3 容器填充機的一致性

因為我們想決定是否存在充分的證據支持這項宣稱，對立假設為

$$H_1: \quad \sigma^2 < 1$$

虛無假設為

$$H_0: \quad \sigma^2 = 1$$

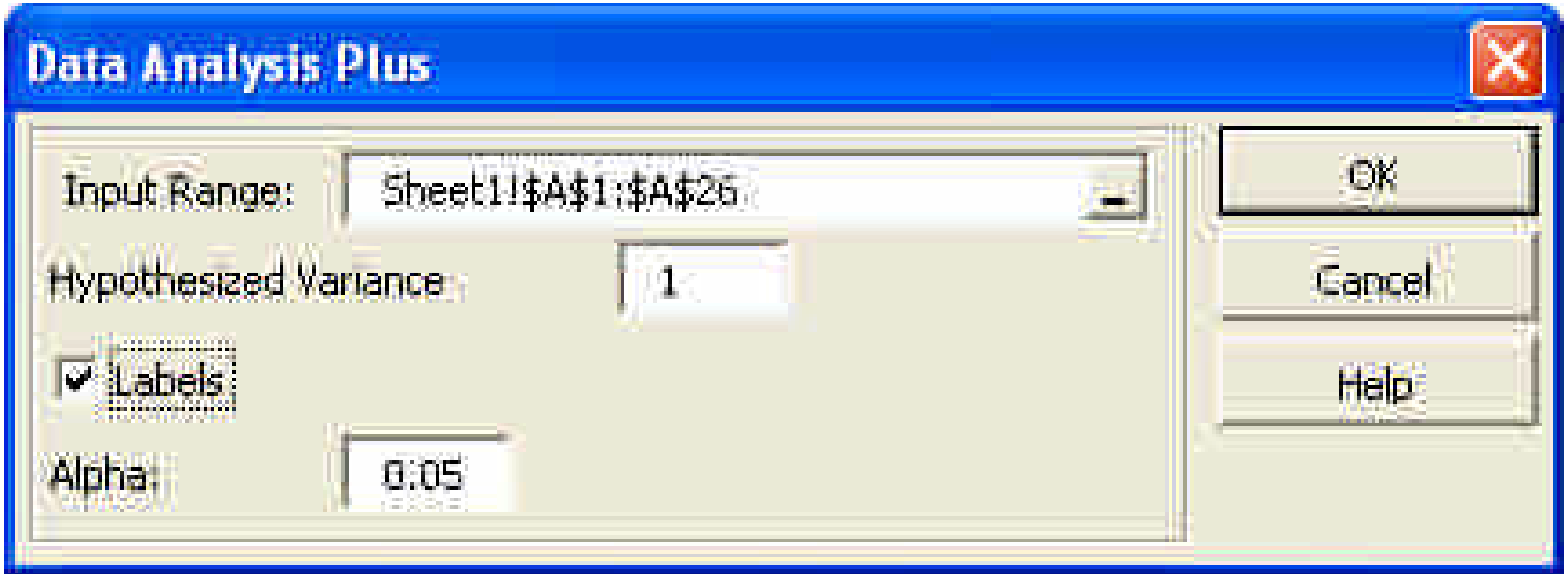
以及我們將使用的檢定統計量是

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

計算

範例 12.3 容器填充機的一致性

點選 **Add-Ins、Data Analysis Plus，與Chi-squared test: Variance**



範例 12.3 容器填充機的一致性

計算

	A	B	C	D
1	**Chi Squared Test: Variance**			
2				
3				*Fills*
4	Sample Variance			0.6333
5	Hypothesized Variance			1
6	df			24
7	chi-squared Stat			15.20
8	P (CHI<=chi) one-tail			0.0852
9	chi-squared Critical one tail	Left-tail		13.85
10		Right-tail		36.42
11	P (CHI<=chi) two-tail			0.1705
12	chi-squared Critical two tail	Left-tail		12.40
13		Right-tail		39.36

詮釋

範例 12.3 容器填充機的一致性

沒有足夠的證據去推論這項宣稱為真。

如同我們先前討論過的，此結果並未說明變異數是等於 1；它只敘述我們無法證明變異數是小於 1。

範例 12.4

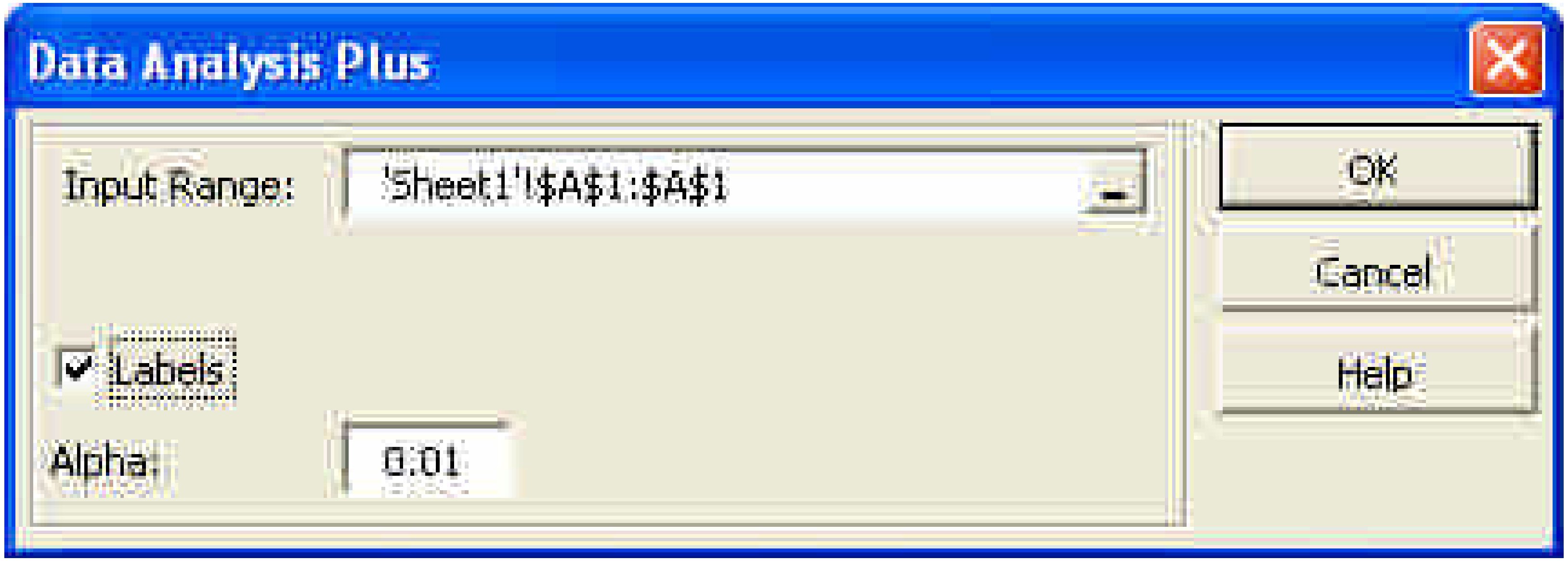
以 99% 的信心估計範例12.3中裝填量的變異數。

Xm12-03

計算

範例12.4

點選Add-Ins、Data Analysis Plus，與Chi-squared Estimate: Variance



範例12.4

計算

	A	B
1	**Chi Squared Estimate: Variance**	
2		
3		*Fills*
4	Sample Variance	0.6333
5	df	24
6	LCL	0.3336
7	UCL	1.5375

詮釋

範例 12.4

在範例12.3中，我們了解沒有充分的證據去推論母體變異數是小於 1。在此我們看到裝填量變異數被估計落在 .3336 與 1.537 之間。

此區間的一部分是在 1 之上，這告訴我們該變異數可能大於 1，確證我們在範例12.3中所達成的結論。

詮釋

範例 12.4

我們可能可以使用這個估計值來預測填充過量與填充不足瓶子的百分比。這可能容許我們在具競爭性的機種中做選擇。

檢查必要的條件

- Normal population is required for 卡方統計量。假如母體不是常態，卡方檢定的信賴區間估計值仍然是合適的，只要母體「不是**極端地**非常態」。
- 為了檢查這個必要的條件，我們**繪製直方圖**來決定它們是否嚴重地偏離鐘形分配。假如直方圖極端地(假定在一個指數分配的情況下)，其可能被視為「極端地非常態」，因此卡方統計量在這範例中不是合理的。

辨識因素

辨識卡方檢定與 σ^2 之估計量的因素：

1. 問題目的：描述一個母體。
2. 資料類型：區間。
3. 敘述性測量的類型：變異性。

12.3 單一母體比例的推論 (categorical data)

當資料是名目時，我們計數每一個數值發生的次數。因此，在描述一個名目資料的母體時，感興趣的參數是母體比例 p。

這個參數被用來計算二項實驗的機率。

回想這個統計量的用法：樣本比例 $\hat{p} = \frac{x}{n}$

where x 是樣本大小 n 中成功的次數。

單一母體比例的推論

np 與 ***n(1–p)*** 大於 5 的條件下，$\hat{p}$ 的抽樣分配近似於常態

平均數：$\mu = p$

標準誤：$\sigma = \sqrt{\dfrac{p(1-p)}{n}}$

因此，$\hat{p}$ 的抽樣分配為：　$z = \dfrac{\hat{p} - p}{\sqrt{p(1-p)/n}}$

單一母體比例的推論

p 的檢定統計量為：

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

p 的信賴區間估計量。結果是

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

(當 ***np*** 與 ***n(1–p)***大於5時)

範例 12.5

在選舉日投票結束後，電視網競爭著搶先預測哪一位候選人將勝選。

預測透過某些選區的投票或是投票處出口民意調查。

出口處民意調查基於投票者的一個隨機樣本，當他們離開投票處時，詢問他們的票投給了誰。

範例 12.5

美國總統選舉中，在一個州當中得到最多選票的候選人獲得該州代表投票選總統的投票權。

實際上，這意味共和黨或是民主黨的候選人其中之一會贏。

假設一個州當中，出口處民意調查的結果被記錄為 1 =民主黨編碼與2 =共和黨編碼 。 Xm12-05*

範例 12.5

投票結束於晚間 8:00。

由這些資料，電視網是否能夠下結論說共和黨的候選人將會勝選？

電視網是否應該在晚間 8:01 宣佈共和黨的候選人將會勝選？

辨識方法

範例 12.5

問題的目的是描述一個州的投票母體．資料是名目的，因為這些數值是「民主黨」與「共和黨」。因此要檢定的參數是整個州之內投票給共和黨候選人的比例。因為我們想決定是否電視網可以在晚間8:01 宣佈共和黨的候選人是勝選者，對立假設為

$$H_1: \quad p > .5$$

虛無假設為

$$H_0: \quad p = .5$$

範例 12.5

辨識方法

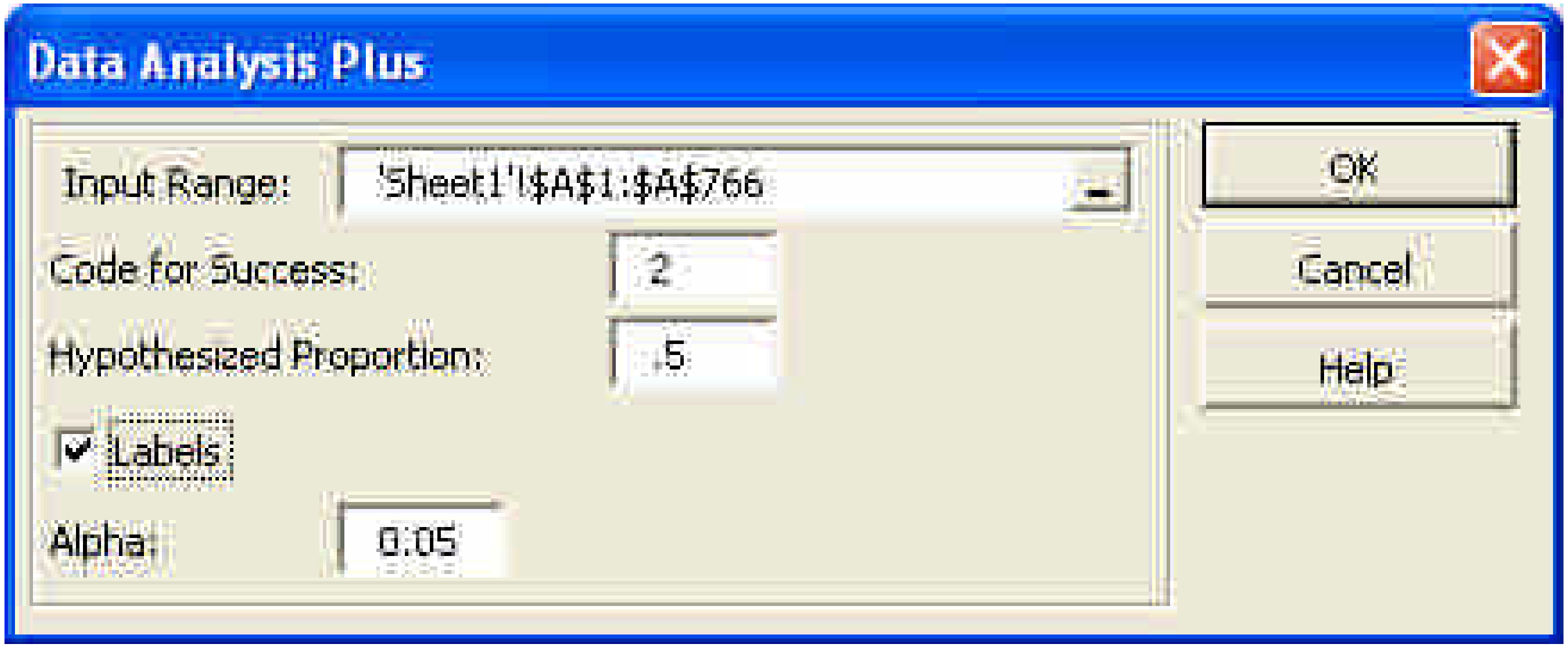
檢定統計量為

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

範例 12.5

計算

點選 **Add-Ins、Data Analysis Plus，與z-test: Proportion**



範例 12.5

計算

	A	B	C	D
1	**z-Test: Proportion**			
2				
3				*Votes*
4	Sample Proportion			0.532
5	Observations			765
6	Hypothesized Proportion			0.5
7	z Stat			1.7716
8	P(Z<=z) one-tail			0.0382
9	z Critical one-tail			1.6449
10	P(Z<=z) two-tail			0.0764
11	z Critical two-tail			1.9600

範例 12.5

詮釋

使用5%的顯著水準，我們拒絕虛無假設並且下結論說有充分的證據去推論共和黨贏得選舉。

然而，這是否為正確的選擇？

詮釋

範例 12.5

在此考慮的關鍵問題是型 I 與型 II 錯誤的成本。

如果我們下結論說共和黨將會贏，事實上共和黨輸了，則發生型 I 錯誤。

詮釋

範例 12.5

這類錯誤的意思是電視網在晚間 8:01 宣佈共和黨候選人已經贏得選舉，稍晚必須承認這是一項錯誤。

如果只有一家特定的電視網犯此一錯誤，將造成觀眾對他們的誠信產生懷疑並可能影響觀眾的人數。

詮釋

範例 12.5

這確實發生在 2000 年 11 月美國總統選舉當晚的情況。

在晚間 8:00 投票結束後不久，所有的電視網皆宣佈民主黨候選人 Albert Gore 將會在佛羅里達州獲勝。

數小時之後，各家電視網承認錯誤並宣佈共和黨候選人 George W. Bush已經獲勝。

詮釋

範例 12.5

幾小時之後他們再次承認錯誤，並宣佈此一選舉的結果太接近而無法分出勝負。

幸運地，所有的電視網犯了相同的錯誤。

但是，如果有一家電視網沒有犯此一錯誤，它將保有更好的記錄，在未來可能被使用於新聞節目的廣告中，並可能吸引更多的觀眾。

在此考慮的關鍵問題是型 I 與型 II 錯誤的成本，其使用1%的顯著水準較好。

在一個大的母體中估計總成功次數

就像我們先前看到的，如果母體是大的而且是有限的，我們可以估計母體中的總成功次數，藉由母體大小(N)的產生以及信賴區間估計量的母體大小為：

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

尼爾森排行(用來測量電視觀眾)使用這個方法。從小樣本觀眾(5000個電視觀眾)來推斷電視家庭的總數(1.1億)。

尼爾森排行

統計方法在協助廣告商決定多少電視觀眾觀看他們所贊助的節目扮演重要的角色。

雖然有數家公司以抽樣電視觀眾來決定他們觀看的電視節目，但是最著名的是A. C. Nielsen 公司。

尼爾森排行是根據一個大小約5,000 的隨機樣本，它是從美國境內至少擁有一部電視的1.1 億個家庭中抽出(在2007 年)。

尼爾森排行

一個測量器被安裝在每一個選取家庭的電視上，以記錄開電視的時間與收視的頻道。

每天晚上資料被傳送到尼爾森公司的電腦以計算排行，因此贊助商可以決定觀眾的人數以及任何商業廣告的潛在價值。

尼爾森排行

- 在2007 年4 月1 日星期日晚間 9:00 到 9:30 時段，使用下列編碼來記錄結果：

電視網	節目	編碼
ABC	*Desperate Housewives*	1
CBS	*The Amazing Race 11*	2
NBC	*Deal or No Deal*	3
Fox	*Family Guy*	4
沒開電視或看其他頻道		5

資料來源：*Televisionweek*, April 7, 2007.

辨識方法

尼爾森排行

此問題的目的是要描述全國觀眾所收看的電視節目的母體。

資料是名目的。

問題目的與資料型態的組合使得被估計的參數是觀看“ *Deal or No Deal* ”全部母體的比例。

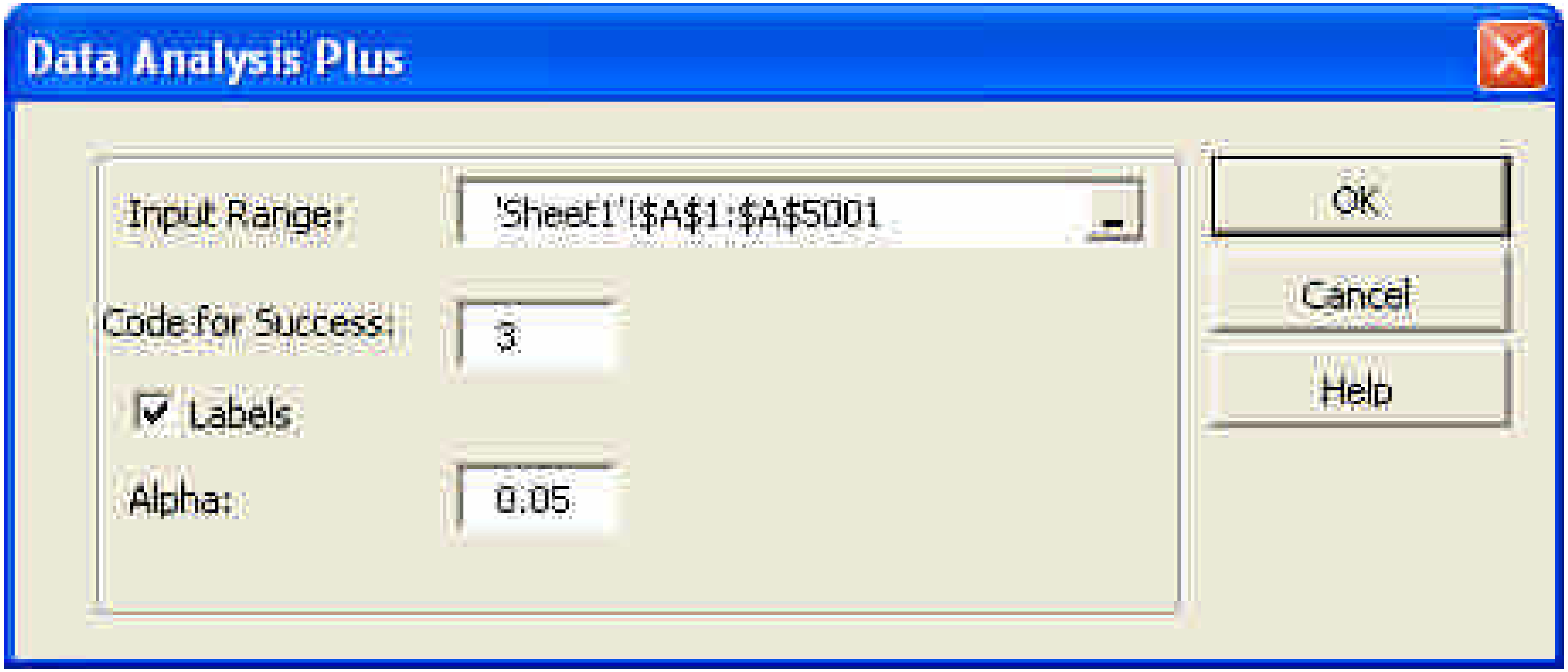
比例的信賴區間估計量為

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

尼爾森排行範例

計算

點選 **Add-Ins、Data Analysis Plus，與z-estimate: Proportion**



尼爾森排行範例

計算

	A	B
1	**z-Estimate: Proportion**	
2		*Program*
3	Sample Proportion	0.0836
4	Observations	5000
5	LCL	0.0759
6	UCL	0.0913

詮釋

尼爾森排行

我們估計，所有擁有電視的家庭收看“*Deal or No Deal*” 的比例在 7.59% 與 9.13% 之間

如果我們將這些數字乘以所有電視機的總數—1.1億台，我們產生一個收看“ *Deal or No Deal* ”節目數量的區間估計值。

詮釋

尼爾森排行

所以，收看“ *Deal or No Deal* ”節目的電視總數之 95%信賴區間估計值位於

$$LCL = .0759 \times 110,000,000 = 8,349,000$$

以及

$$UCL = .0913 \times 110,000,000 = 10,043,000$$

贊助公司因此可以決定任何出現在節目中的商業廣告的價值。

選擇估計比例的樣本大小

當我們在 10.3 節中介紹估計一個平均數的樣本大小選擇方法時，我們指出樣本大小依賴信賴水準及統計實作人員能夠容忍的估計誤差範圍。

當要估計的參數是一個比例時，此一估計誤差範圍是

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

選擇估計比例的樣本大小

為了求 n 的解，我們產生必要的樣本大小以估計 p，其中 B 是估計的誤差範圍

$$n = \left(\frac{Z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})}}{B} \right)^2$$

遺憾的是，這個數值通常是未知的 $\hat{p}$。

選擇估計比例的樣本大小

- 兩種方法——每一個案例我們可以為 $\hat{p}$ 選擇一個值，然後解 n 的方程式。
  - 方法1：如果我們甚至連 $\hat{p}$ 的近似值都沒有的時候，這是“最差的情況”所以我們以 $\hat{p}$ = .50 取代之。
  - 方法2：我們對 $\hat{p}$ 的值有些概念。這是比較好的情況，並且我們代入估計值 $\hat{p}$。

選擇估計比例的樣本大小 (p449-450)

方法1：沒有 $\hat{p}$ 的值，使用50%：

$$n = \left(\frac{1.96\sqrt{(.5)(.5)}}{.03}\right)^2 = (32.67)^2 = 1{,}068$$

方法2：對 $\hat{p}$ 的值有些概念：

$$n = \left(\frac{1.96\sqrt{(.5)(.5)}}{.01}\right)^2 = (98)^2 = 9{,}604$$

在開始之前，若我們已經有母體比例的合理估計值，我們可以只抽樣少數人。

12.4 行銷上的應用：市場區隔 (optional)

大量行銷 (*Mass marketing*) 是一間公司大量製造與行銷一種單一產品到整個市場。

大量行銷對於日用必需品如汽油之類的商品特別有效，除了透過價格與供貨的方便性之外，這類商品很難從競爭中分出差異。

大量行銷已經逐漸轉型到目標市場的鎖定，著重滿足整體市場中某一特定區隔的需求。

行銷上的應用：市場區隔

因為區隔市場沒有單一的方法，所以經理人員必須考慮幾種可能被用來辨識區隔的不同變數(或特色)。

顧客調查可以用來蒐集有關市場各種層面的資料，以及用來定義區隔的統計方法。

市場區隔區別一項產品的顧客到不同的群組，在這種方式下，每一群組中的成員是相類似的，且群組之間是有差異的。

行銷場上的應用：市場區隔

有許多方法可用來區隔市場。

表12.1列出數種不同的區隔變數以及它們的市場區隔。

表12.1　市場區隔

區隔變數	區隔
地理	
國家	巴西、加拿大、中國、法國、美國
國家區域	中西部、東北部、西南部、東南部
人口統計	
年齡	5 歲以下、5-12 歲、13-19 歲、20-29 歲、30-50 歲、50 歲以上
教育	高中肄業、高中畢業、大專肄業、大專或大學畢業
收入	$20,000 以下、$20,000-$29,999、$30,000-$49,999、$50,000 以上
婚姻狀態	單身、已婚、離婚、鰥寡
社會	
宗教	天主教、基督徒、猶太教、回教、佛教
階層	上層、中層、勞工層、下層
行為	
媒體的使用	電視、網路、報紙、雜誌
付款方式	現金、支票、Visa、Mastercard

行銷上的應用：市場區隔

對行銷經理而言，了解市場的規模是重要的，因為規模大小(參數之一)決定其利潤。

並非所有的區隔都值得追求。在一些實例中，市場區隔的規模太小或是滿足一個區隔的成本可能過高。

行銷上的應用：市場區隔

有幾種方法可以決定市場規模。

人口普查提供有用的資訊。

例如，我們可以使用不同的年齡類型或各地理居住地的大小來決定美國人的人數

對於其他的區隔，我們可能必須調查一般母體的成員，並且使用前一節中介紹的推論方法，其中我們指出如何估計成功的總數。

行銷上的應用：市場區隔

我們可以檢視大的母體來估計落在每一個區隔的母體的比例。

由此計算，我們將使用信賴區間估計值來估計市場規模。

$$N\left(\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

範例 12.6

在區隔早餐穀類食物市場時，一家食品製造廠使用健康與節食意識作為區隔變數。發展出四種區隔：

1. 關心食用健康食物
2. 主要關心的是體重
3. 關心健康問題是因為疾病
4. 不關心

範例 12.6

為了區隔群組，進行了調查。以問卷為基礎，民眾被歸類為上面四種群組之一。

一項最近的調查，要求1,250位美國成人(20歲和以上)的隨機樣本完成這份問卷。各類別使用編碼來記錄。
Xm12-06

最近的普查顯示有207,347,000位美國人在20歲和20歲以上。

以95%的信心估計關心食用健康食物的美國成人人數。

範例 12.6

辨識方法

此問題的目的是描述美國成人將自己歸類為關健康與日常飲食。

資料是名目的。因此，我們希望估計的參數是將自己歸類為關心吃得健康之美國人的比例 p (編碼 = 1)。

我們需要使用的信賴區間估計量為

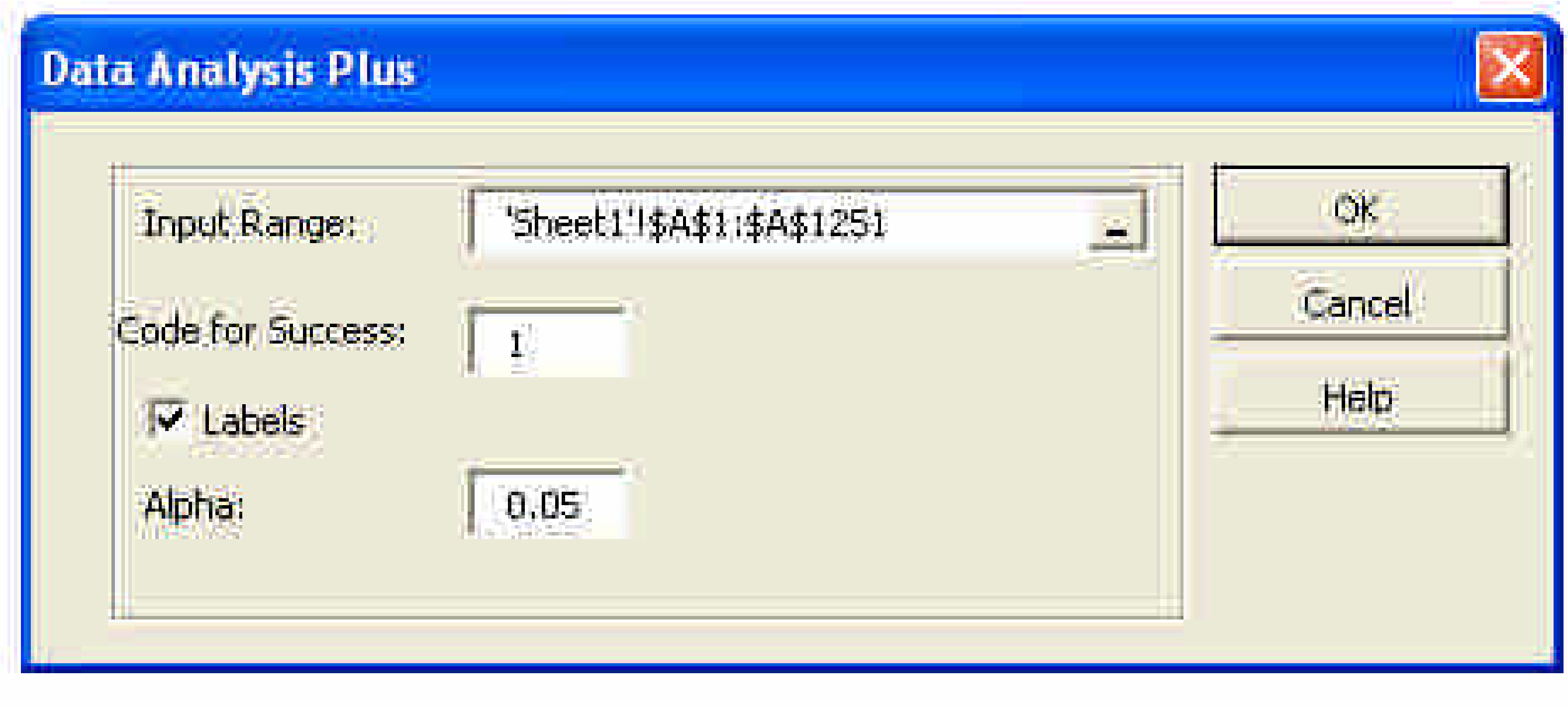
$$\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

由此計算，我們將產生市場區隔規模的估計值。

範例 12.6

計算

點選 **Add-Ins、Data Analysis Plus，與Z-Estimate: Proportion**



範例 12.6

計算

	A	B
1	**z-Estimate: Proportion**	
2		*Group*
3	Sample Proportion	0.2152
4	Observations	1250
5	LCL	0.1924
6	UCL	0.2380

範例 12.6

計算

我們藉由母體大小來增加其上限與下限

$$\text{LCL} = N\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) = 194{,}506{,}000(.1924) = 37{,}422{,}954$$

且

$$\text{UCL} = N\left(\hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) = 194{,}506{,}000(.2380) = 46{,}292{,}428$$

詮釋

範例 12.6

我們計算市場區隔的規模介於 39,893,563 和 49,348,586 之間。