

圖 10.2 現有與新軟體技術的假設檢定之 Minitab 輸出

```
Two-sample T for Current vs New

      N      Mean      StDev      SE Mean
Current  12      325.0      40.0         12
New      12      286.0      44.0         13

Difference = mu Current - mu New
Estimate for difference:  39.0000
95% lower bound for difference = 9.4643
T-Test of difference = 0 (vs >):  T-Value = 2.27  P-Value = 0.017  DF = 21
```

為 Minitab 的現有與新軟體技術之比較輸出。結果最後一行顯示 $t = 2.27$ 及 p 值 = 0.017。注意 Minitab 採用式 (10.7) 計算出的自由度為 21。

實用忠告

在任何情形下，均應盡可能地使樣本大小相等，即 $n_1 = n_2$ 。

本節中介紹之區間估計與假設檢定程序為穩健，且可用於極小之樣本大小。大多數應用上，雖然母體不是常態分配，但若兩樣本大小相等或幾乎相等，且總樣本數目 $n_1 + n_2$ 至少為 20，則可期望有較佳結果。若母體分配極度偏斜或包含離群值，則應使用較大的樣本數目。若分析者對母體分配至少是近似常態分配感到滿意，方可使用較小的樣本數。

評註

另一種當 σ_1 與 σ_2 未知時，兩母體平均數之差的推論係基於兩母體具有相同標準差 ($\sigma_1 = \sigma_2 = \sigma$) 的假定。在此假定下，兩樣本標準差結合成下列混合樣本變異數：

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

此時 t 檢定統計量則變成

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

且具有 $n_1 + n_2 - 2$ 個自由度。此時， p 值的計算與抽樣結果的解釋與本節前述討論之程序相同。

此程序的困難之處是較難證實兩母體標準差相等。通常，不同的母體標準差較常出現。使用混合樣本變異數也許無法滿足結果，尤其是當兩樣本大小 n_1 與 n_2 相當不同時。

本節介紹的 t 程序並不需要母體標準差相同之假定，且無論母體變異是否相同均可運用，故其為較普遍之程序，亦被多方運用。

習題

方法

SELF test

9. 考慮下列取自兩母體的獨立隨機樣本的結果。

SELF test

樣本 1

$n_1=20$

$\bar{x}_1=22.5$

$s_1=2.5$

樣本 2

$n_2=30$

$\bar{x}_2=20.1$

$s_2=4.8$

- 此兩母體平均數之差的點估計值為何？
 - t 分配之自由度為何？
 - 在 95% 信賴水準下，邊際誤差為何？
 - 此兩母體平均數之差的 95% 信賴區間為何？
10. 考慮下列假設檢定。

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_a: \mu_1 - \mu_2 \neq 0$$

下列為取自兩母體的獨立樣本結果。

樣本 1

$n_1=35$

$\bar{x}_1=13.6$

$s_1=5.2$

樣本 2

$n_2=40$

$\bar{x}_2=10.1$

$s_2=8.5$

- 檢定統計量之值為何？
 - t 分配之自由度為何？
 - p 值為何？
 - 在 $\alpha=0.05$ 時，你的結論為何？
11. 考慮下列取自兩常態母體的獨立隨機樣本資料：

| 樣本 1 | 10 | 7 | 13 | 7 | 9 | 8 |
|------|----|---|----|---|---|---|
| 樣本 2 | 8 | 7 | 8 | 4 | 6 | 9 |

- 試求兩樣本平均數。
- 試求兩樣本標準差。
- 此兩母體平均數之差的點估計值為何？
- 此兩母體平均數之差的 90% 信賴區間估計值為何？

應用

SELF test

12. 在美國運輸部所蒐集的資料中，可看到美國 75 個大都會區的居民每天在車上的哩程數。假定一個包含 50 位水牛城居民的簡單隨機樣本，其平均數為一天 22.5 哩，標準差為一天 8.4 哩；而另一包含 40 位波士頓市居民的獨立簡單隨機樣本，其平均數為一天 18.6 哩，標準差為一天 7.4 哩。
- 水牛城與波士頓居民每天在車上的平均哩程數之差的點估計值為何？
 - 此兩母體平均數之差的 95% 信賴區間為何？



13. FedEx 與 United Parcel Services (UPS) 在數量與獲利上是世界上兩大貨運公司 (*The Wall Street Journal*, January 27, 2004)。根據國際機場協會，曼菲斯國際機場 (FedEx) 與路易斯維爾國際機場 (UPS) 是世界 10 大貨運機場中的兩個。下面為由這些機場每日處理成噸貨物的隨機樣本，以千噸為單位。

曼菲斯

| | | | | | |
|-----|------|-----|------|------|------|
| 9.1 | 15.1 | 8.8 | 10.0 | 7.5 | 10.5 |
| 8.3 | 9.1 | 6.0 | 5.8 | 12.1 | 9.3 |

路易斯維爾

| | | | | |
|-----|-----|-----|-----|-----|
| 4.7 | 5.0 | 4.2 | 3.3 | 5.5 |
| 2.2 | 4.1 | 2.6 | 3.4 | 7.0 |

- 請計算各個機場的樣本平均數與樣本標準差。
 - 此兩母體平均數之差的點估計值為何？由較高容量機場的觀點解釋此值，並比較兩機場的容量差異。
 - 試建立兩個機場的每日母體平均分數之差的 95% 信賴區間。
14. 在 1990 年代，包括鱈魚角、外灘群島、卡羅來納地區，以及海灣海岸等美國沿海區域，即有非常高的人口成長率。全美國居住在沿海與非沿海地區居民的資料均被蒐集 (*USA Today*, July 21, 2000)。假設下列樣本結果分別來自兩個母體之個人年齡：

沿海地區

$n_1 = 150$
 $\bar{x}_1 = 39.3$ 歲
 $s_1 = 16.8$ 歲

非沿海地區

$n_2 = 175$
 $\bar{x}_2 = 35.4$ 歲
 $s_2 = 15.2$ 歲

欲檢定 $\alpha = 0.05$ 下，此兩母體平均數間無差異。

- 請建立虛無與對立假設。
 - 檢定統計量之值為何？
 - p 值為何？
 - 你的結論為何？
15. 近幾年來美國職棒大聯盟球員受傷情形持續增加。在 1992 年至 2001 年期間，因聯盟擴張造成大聯盟球員名單增加 15%。然而，在同一時期，球員因受傷而被列入傷兵名單人數亦增加 32% (*USA Today*, July 8, 2002)。一項研究報導，在 2001 年大聯盟球員被放在傷兵名單的時間較 10 年前長。
- 請利用在傷兵名單中球員母體平均天數，建立檢定上述研究問題之虛無與對立假設。
 - 設若運用下列資料：



| | 2001 年球季 | 1992 年球季 |
|-------|--------------------|--------------------|
| 樣本大小 | $n_1 = 45$ | $n_2 = 38$ |
| 樣本平均數 | $\bar{x}_1 = 60$ 天 | $\bar{x}_2 = 51$ 天 |
| 樣本標準差 | $s_1 = 18$ 天 | $s_2 = 15$ 天 |

2001 年球季傷兵名單與 1992 年球季的母體平均天數之差的點估計值為何？
傷兵名單增加天數的比例為何？

- c. 使用 $\alpha = 0.01$ ，你對傷兵名單的天數結論為何？ p 值為何？
d. 這些資料可否建議職棒大聯盟更應多關心此種情形？

16. 大學委員會提供基於受測者父母最高教育程度的學術性向測驗 (SAT) 比較分數，研究假設為父母教育程度較高的學生在 SAT 上有較高的平均成績。2003 年時整體平均 SAT 言語分數為 507 (*The World Almanac 2004*)。學生獨立樣本的 SAT 測驗分數亦同。第一組樣本為父母具學士學位的學生之 SAT 言語測驗分數，第二組樣本則是父母高中畢業但不具大學學歷的學生之 SAT 言語測驗分數。

| 學生的父母 | | | |
|-------|-----|------|-----|
| 大學畢業 | | 高中畢業 | |
| 485 | 487 | 442 | 492 |
| 534 | 533 | 580 | 478 |
| 650 | 526 | 479 | 425 |
| 554 | 410 | 486 | 485 |
| 550 | 515 | 528 | 390 |
| 572 | 578 | 524 | 535 |
| 497 | 448 | | |
| 592 | 469 | | |

- a. 請建立假設以決定樣本資料是否支持父母教育程度較高的學生，會有較高的母體平均 SAT 言語測驗分數之檢定。
b. 此兩母體平均數之差的點估計值為何？
c. 請計算此假設檢定的 p 值。
d. 在 $\alpha = 0.05$ 時，你的結論為何？
17. Merrill Lynch 公司會週期性地要求其顧客對該公司之財務顧問與服務等作評價 (2000 年 Merrill Lynch 顧客滿意度調查)。顧客滿意度調查指標愈高表示其服務愈好，7 則為最高之服務等級。針對兩位財務顧問之獨立抽樣服務評比資料如下，其中顧問 A 有 10 年經驗；而顧問 B 有 1 年經驗。請使用 $\alpha = 0.05$ ，檢定是否較具經驗的顧問有較高之平均服務評價。

| 顧問 A | 顧問 B |
|--------------------|--------------------|
| $n_1 = 16$ | $n_2 = 10$ |
| $\bar{x}_1 = 6.82$ | $\bar{x}_2 = 6.25$ |
| $s_1 = 0.64$ | $s_2 = 0.75$ |



- a. 請建立虛無假設與對立假設。
- b. 請計算檢定統計量之值。
- c. p 值為何？
- d. 你的結論為何？

18. 一些教育補習班提供家教、課堂學習，以及模擬測驗等，以協助學生能在一些如學術性向測驗 (SAT) 等考試中獲得較好的成績。補習班業者宣稱其課程可提升 SAT 成績平均約 120 分 (*The Wall Street Journal*, January 23, 2003)。某一研究人員不認同此種說法，並認為 120 分係言過其實，僅為慫恿學生參加其考前預習課程而已。為調查某一補習班之成效，該研究人員蒐集了 35 位參加考前預習課程的學生與 48 位未參加此課程學生的 SAT 成績。本研究成績收錄於 CD 光碟中，檔案名為 SAT。

- a. 請建立可用於檢定該研究人員之 SAT 分數改善少於平均 120 分看法的假設。
- b. 使用 $\alpha=0.05$ ，你的結論為何？
- c. 參加考前預習課程而使平均 SAT 成績改善的點估計值為何？請建立此改善之 95% 信賴區間之估計值。
- d. 得知此信賴區間後，你對該研究人員的建議為何？



10.3 兩母體平均數之差的推論：配對樣本

假定某製造商欲生產某一特定產品可以採用兩種不同的方法。為了使產出最大，該公司想知道哪一種方法其完成一單位產品的平均時間最短。令 μ_1 及 μ_2 分別代表生產方法 1 及生產方法 2 的母體平均數完成時間，由於事先並不知道哪一種方法較好，我們先假定這兩種方法的平均完成時間相等。因此，虛無假設為 $H_0: \mu_1 - \mu_2 = 0$ 。如果此一假設被拒絕，即顯示其母體平均數完成時間確有差異。而在本案例中，平均完成時間較短的方法將被公司所採用。此時，虛無及對立假設可表示如下。

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_a: \mu_1 - \mu_2 \neq 0$$

而在設計適當的抽樣程序，以蒐集生產時間的資料並對上述假設進行檢定時，有兩種方法：獨立樣本 (independent samples) 及配對樣本 (matched samples) 可供選擇。

1. 獨立樣本設計：從作業員中選取一簡單隨機樣本，且每位作業員使用方法 1 來生產；另從作業員中選取一獨立簡單隨機樣本，且每位作業員使用方法 2 來生產。並以 10.2 節中所介紹的程序，檢定母體平均數的差異。
2. 配對樣本設計：從作業員中選取一簡單隨機樣本，每位作業員先使用其中一種方法來生產，再使用另一方法來生產。至於使用方法的順序則為隨機指派，因此，某些作業員先使用方法 1，某些則先使用方法 2。如此下來，每一位被選取的作業員均提供了一對觀測值，其中一個是方法 1 的，另一個是方法 2 的。