T Testing
T 檢定

Jong Yih Kuo

jykuo@ntut.edu.tw
Department of Computer Science and Information Engineering
National Taipei University of Technology

推論統計

- 敘述統計學(descriptive statistics)
  - 92.5%的受訪者表示今年暑假會去打工。
  - 打工主要動機，有76.09%的人是為了賺錢，20.13%的人為了獲得工作經驗
  - 這些資料僅說明資料收集整理後的概略結果。
- 推論統計學(inferential statistics)
  - 近代統計學不僅收集資料，且以少數資料推論該資料全體性質。
  - 統計檢定最好能取得母體資訊，但母體太龐大，所以由樣本著手。

母體變異數: $\sigma^2 = \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{\mathrm{N}}$　　樣本變異數: $S^2 = \frac{\sum_{i=1}^{N}(x_i-\bar{x})^2}{n-1}$

無母數統計

- 無母數統計的特性
  - 母體數值分布型態，是不限定的。
  - 此統計方法的假設條件較少。
- 無母數統計的缺點
  - 若母體分布已知，與母數統計法相比，分析效益較差。
  - 缺乏相關機率表格可使用。
  - 若對象是常態分布資料，檢定力較差。

樣本變異數使用要n-1

- 若體重計有時秤重、有時秤輕，但秤許多次後，平均起來等於真實體重，相當於有不偏性質。但若體重計無法正確歸零，常把體重「加碼」，量很多次，平均多秤半公斤，如此稱系統性高估，而非「不偏」。
- 假設樣本變異數公式的分母用n而非n-1，將它當作母體變異數估計，常會低估，不符合「不偏」條件，把 n 改成n-1，樣本變異數會是母體變異數的不偏估計
- 例：母體{1, 2, 3, 4, 5}，一次抽取樣本數量4個(n=4)，抽取不放回，所有可能：(已知母體變異數為2)
- n-1 較接近

	$\frac{1}{4}\sum_{i=1}^{4}(x_i-\bar{x})^2$	$\frac{1}{3}\sum_{i=1}^{4}(x_i-\bar{x})^2$
1,2,3,4	0.625	0.8333
1,2,3,5	2.1875	2.9166
1,2,4,5	2.5	3.3333
1,3,4,5	2.1875	2.9166
2,3,4,5	0.625	0.8333
總和/5	1.625	2.16662

T-test介紹

- 用於小樣本（樣本量小於30）的兩個平均值差異程度的檢驗方法。
- 藉由實驗組與對照組視為兩母群體樣本，使用其樣本平均數進行假設檢定。
- T-test是高斯特(William Sealy Gosset)為觀測釀酒品質而發明，於1908年公佈，因其老闆認為是商業機密而被迫使用筆名(學生)，所以T-test亦稱Student's t test。
- T-test假設：
  - 兩樣本不會互相影響(是獨立樣本)
  - 要符合常態分布(不符合就要用無母數統計)
  - 兩樣本的變異數、標準差須已知

範例

- 10個病人隨機分配到2組中的其中一組。
- 每組都含有5個病人。
- Group1：服用抗憂鬱藥(antidepressant drug) 6個月
- Group2：服用安慰劑(placebo drug) 6個月
- 之後由心理醫生記錄這10位病人的憂鬱程度 (0~12)
- Group1：11, 1, 0, 2, 0.
- Group2：11, 11, 5, 8, 4.
- 分數越高，受試者憂鬱程度越嚴重
- 檢定: 服用抗憂鬱藥是否有效？

虛無假設與對立假設

- 虛無假設(null hypothesis) H_0：假設兩種方法成效相同。
  - u1 = u2
- 對立假設(alternative hypothesis)H_1：假設兩種方法成效不同。
  - u1 ≠ u2
    - 雙尾檢定，抗憂鬱藥(Group1)的成效可能比安慰劑(Group2)好或差
  - u1 > u2
    - 右邊單尾檢定，抗憂鬱藥(Group1)的成效比安慰劑(Group2)差
  - u1 < u2
    - 左邊單尾檢定，抗憂鬱藥(Group1)的成效比安慰劑(Group2)好

樣本變異數

樣本變異數：$S^2 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{n-1} = \frac{1}{n-1}\left[\sum_{i=1}^{N} {x_i}^2 - n\bar{x}^2\right]$

Table 11.1 Data for Example 11.1

	Group 1			Group 2	
	X_1	X_1^2		X_2	X_2^2
Subject 1,1	11	121	Subject 1,2	11	121
Subject 2,1	1	1	Subject 2,2	11	121
Subject 3,1	0	0	Subject 3,2	5	25
Subject 4,1	2	4	Subject 4,2	8	64
Subject 5,1	0	0	Subject 5,2	4	16
	$\Sigma X_1 = 14$	$\Sigma X_1^2 = 126$		$\Sigma X_2 = 39$	$\Sigma X_2^2 = 347$

Subject i, j：第j個group裡的第i個資料

計算T值(平均數跟樣本變異數)

- 樣本均值(mean)與樣本變異數(estimated population variance)

$$\bar{X}_1 = \frac{\Sigma X_1}{n_1} = \frac{14}{5} = 2.8 \qquad S^2 = \frac{\Sigma X_1^2 - \frac{(\Sigma X_1)^2}{n_1}}{n_1 - 1} = \frac{126 - \frac{(14)^2}{5}}{5 - 1} = 21.7$$

$$\bar{X}_2 = \frac{\Sigma X_2}{n_2} = \frac{39}{5} = 7.8 \qquad S^2 = \frac{\Sigma X_2^2 - \frac{(\Sigma X_2)^2}{n_2}}{n_2 - 1} = \frac{347 - \frac{(39)^2}{5}}{5 - 1} = 10.7$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left[\frac{(n_1 - 1)\tilde{s}_1^2 + (n_2 - 1)\tilde{s}_2^2}{n_1 + n_2 - 2}\right]\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\tilde{s}_1^2}{n_1} + \frac{\tilde{s}_2^2}{n_2}}} = \frac{2.8 - 7.8}{\sqrt{\frac{21.7}{5} + \frac{10.7}{5}}} = \frac{-5}{2.55} = -1.96$$

t.05

- t.05(顯著水準, the level of significant)：
  - alpha(α)，通常設為0.05。
  - 有95%的信心水準，答案是對的。
    - 這兩組資料存在顯著性差異的機率是95%。
  - 錯誤只有5%。
    - 這兩組數據有5%的機率是沒有差異的。
    - 有5%的機率，計算出可拒絕虛無假設，但實際上虛無假設卻成立(Type I error)。
    - 5%的差異是隨機誤差造成。

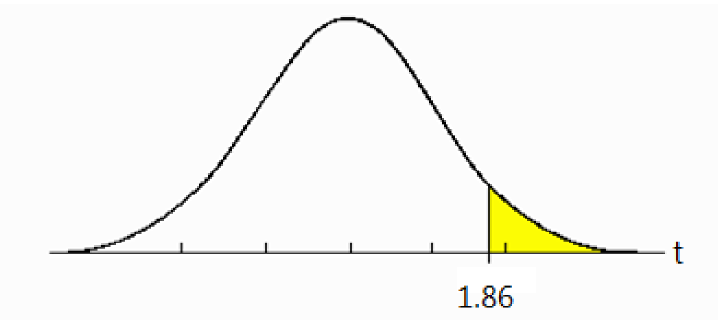
測試結果說明(t.05)

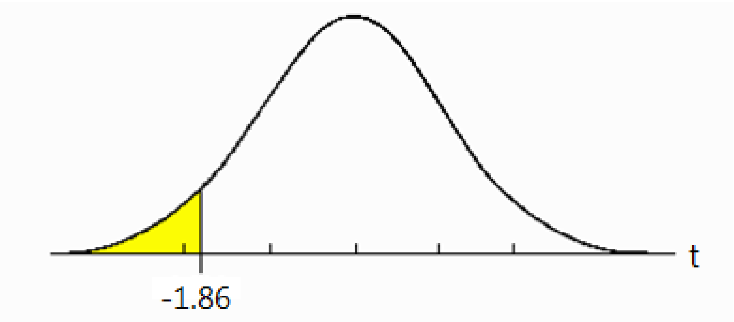
- 計算自由度df (degree of freedom)
  - df = n1 + n2 – 2 = 8，兩群體，樣本為N-1，所以須「– 2」
  - 查表，根據.05與one-tailed跟two-tailed查表，得到值：
  - Two-tailed values, t.05 = 2.31；One-tailed values t.05=1.86

	雙尾檢定	0.5	0.2	0.1	0.05	0.02	0.01	0.005	0.002	0.001
	單尾檢定	0.25	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
df										
1		1	3.078	6.314	12.706	31.821	63.657	127.321	318.309	636.619
2		0.816	1.886	2.92	4.303	6.965	9.925	14.089	22.327	31.599
3		0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.215	12.924
4		0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.61
5		0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6		0.718	1.44	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7		0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8		0.706	1.397	1.86	2.306	2.896	3.355	3.833	4.501	5.041
9		0.703	1.383	1.833	2.262	2.821	3.25	3.69	4.297	4.781
10		0.7	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587

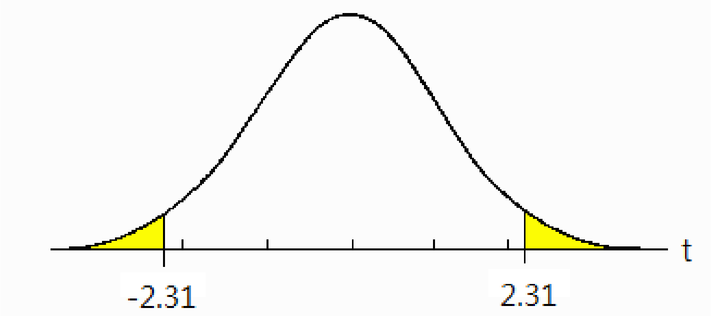
臨界值&單雙尾檢定(t.05)

- 臨界值(critical value)：虛無假設拒絕域與接受域間的點。
- 單尾檢定(one-tailed)：對立假設為單一方向。
  - 尾檢定：(t.05的情況)





- 雙尾檢定(two-tailed)：對立假未指出一個方向。
  - 雙尾檢定：(t.05)



測試結果說明(t.05)

- 本範例計算所得 t 值是 -1.96、|t| 值是1.96，三種可能：
  - u1≠u2：若 |t| ≧ two-tailed的值(1.96 ≧ 2.31) 【無方向H1】
  - u1>u2：若t是正數；t≧one-tailed的值(t > 0 and t≧1.86)
  - u1<u2：若t是負數；|t|≧one-tailed的值(t<0 and |t|≧ 1.86)
- 上述三個對立假設，只有u1<u2成立 (|-1.96| ≧ 1.86)。
  - 可推斷服用抗憂鬱藥的成效比使用安慰劑好。

t.01

- t.01(顯著水準, the level of significant)：
  - 有時也會設為0.01。
  - 有99%的信心水準，答案是對的。
    - 這兩組數據存在顯著性差異的機率是99%。
  - 錯誤只有1%。
    - 這兩組數據有1%的機率是沒有差異的。
    - 有1%的機率，計算出來可拒絕虛無假設，但實際上虛無假設卻成立(Type I error)。
    - 這1%的差異是隨機誤差造成的。

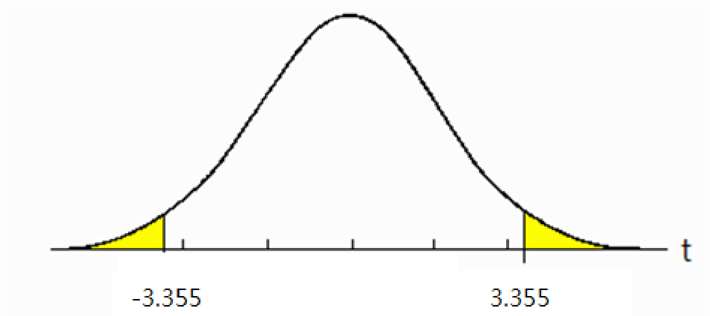
測試結果說明(t.01)

- 計算df值(自 由 度)
  - df = n1 + n2 – 2 = 8
  - 查表，根據.01與one-tailed跟two-tailed查表：

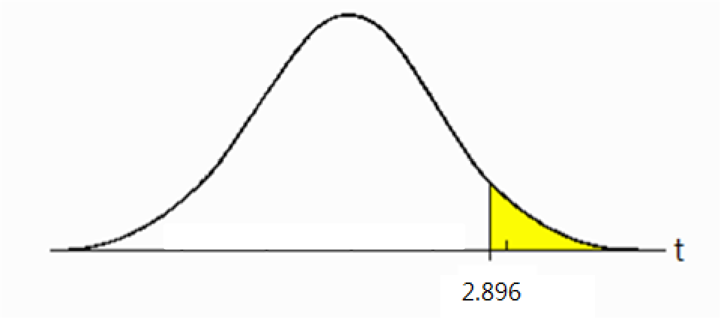
	雙尾檢定	0.5	0.2	0.1	0.05	0.02	0.01	0.005	0.002	0.001
	單尾檢定	0.25	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
df										
1		1	3.078	6.314	12.706	31.821	63.657	127.321	318.309	636.619
2		0.816	1.886	2.92	4.303	6.965	9.925	14.089	22.327	31.599
3		0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.215	12.924
4		0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.61
5		0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6		0.718	1.44	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7		0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8		0.706	1.397	1.86	2.306	2.896	3.355	3.833	4.501	5.041
9		0.703	1.383	1.833	2.262	2.821	3.25	3.69	4.297	4.781
10		0.7	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587

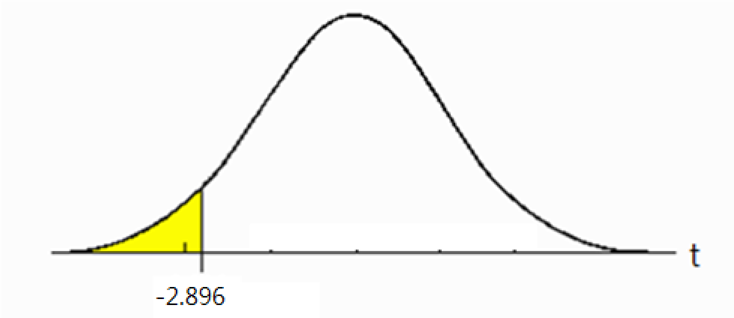
臨界值&單雙尾檢定(t.01)

- 雙尾檢定：(t.01的情況下)



- 單尾檢定：(t.01的情況下)





測試結果說明(t.01)

- 本範例的t值是 -1.96、|t|值是1.96
- u1 ≠ u2：如果|t| ≧ two-tailed的值(如果 1.96 ≧ 3.355)
- u1 > u2：t是正數；如果 t ≧ one-tailed的值(如果t>0且 1.96 ≧ 2.896)
- u1 < u2：t是負數；如果|t| ≧ one-tailed的值(如果t<0且1.96 ≧ 2.896)
- 上述三個對立假設，均不接受。
- 因此可推斷服用抗憂鬱藥的成效與使用安慰劑並無差別(接受虛無假設)。

T值的物理意義

- T值越小，代表兩樣本差異越小；越大代表兩樣本差異越大。
  - T值是否超越臨界值，判定是否有顯著性差異。
  - 分子是兩樣本平均數互減，數值越大代表兩樣本差異越大。
  - 變異數越大，代表資料離散程度越大，越不準確。
  - 兩變異數越大，T值越小，越不容易接受虛無假設，因此變異數越小越明確，使T值越大，越容易拒絕虛無假設。

Exercise1

- 數學老師為了驗證新的教學方法(電腦上課)是否與傳統教學方法(黑板上課)不同，分成兩個班級上課，教學一段時間後，讓他們去考學測模擬考，而得到的級分數如下：
  - Group1(電腦上課)：13, 2, 4, 0, 1
  - Group2(黑板上課)：11, 12, 8, 9, 10
- The level of significant(α值): 0.05 (t.05)

回前面

Exercise1

- 平均值：

$$X1\ 平均值 = \frac{\Sigma X1}{n1} = 4$$

$$X2\ 平均值 = \frac{\Sigma X2}{n2} = 10$$

標準差：

$$X1\ 標準差 = \frac{\Sigma X1^2 - \frac{(\Sigma X_1)^2}{n_1}}{n1-1} = \frac{190 - \frac{20^2}{5}}{5-1} = 27.5$$

$$X2\ 標準差 = \frac{\Sigma X2^2 - \frac{(\Sigma X_2)^2}{n_2}}{n2-1} = \frac{510 - \frac{50^2}{5}}{5-1} = 2.5$$

- T值：

$$T = \frac{X1\ 平均值 - X2\ 平均值}{\sqrt{\frac{X_1\ 標準差}{n_1} + \frac{X_2\ 標準差}{n_2}}} = \frac{4-10}{\sqrt{\frac{27.5}{5} + \frac{2.5}{5}}} = -2.449489742783178(可取近似值)$$

- 雙尾的值藉由查表得知是2.31
- 如果|T| ≧ two-tailed的值，可接受研究假設(拒絕虛無假設)
- 2.45大於2.31，因此可拒絕虛無假設

Exercise2

- Owens 公司主要從事製造與組裝除草機。現在有兩種組裝與架設除草機引擎的程序，問題是：使用這兩種啟動引擎的平均時間是否有差異？第一種程序由公司員工Herb Welles 所開發（稱為程序 1），而另一種由公司副總裁 William Atkins 所開發（稱為程序 2）。公司決定進行研究組裝引擎時間的評估比較。
- 現在抽選 5 位員工為樣本，使用程序 1 來組裝引擎；另外再抽選 6 位員工使用程序 2 來組裝引擎，結果如下表所示（以分鐘為單位），請問組裝引擎的平均時間是否有差異？請使用 0.05 顯著水準。
- Welles：2, 4, 9, 3, 2
- Atkins：3, 7, 5, 8, 4, 3

Exercise2

- X1平均值 = 4 X2平均值 = 5
- X1樣本變異數 = 8.5 X2樣本變異數 = 4.4
- T值 = -0.44
- df(自 由 度) = 5+6-2 = 9
- 因α值為0.05，所以查表得知臨界值是2.262
- u1 ≠ u2：如果|t| ≧ two-tailed的值(如果 0.44 ≧ 2.262)
- 所以不能拒絕虛無假設

Exercise 3

- Rosenthal and Jacobson (1968) informed classroom teachers that some of their students showed unusual potential for intellectual gains. Eight months later the students identified to teachers as having protentional for unusual intellectual gains showed significantly greater gains performance on a test said to measure IQ than did children who were not so identified. Below are the data for the students in the first grade:(The level of significant is 0.05)

Experimental	**Comparison**
35, 40, 12, 15, 21, 14, 46, 10, 28, 48, 16, 30, 32, 48, 31, 22, 12, 39, 19, 25	02, 27, 38, 31, 01, 19, 01, 34, 03, 01, 02, 03, 02, 01, 02, 01, 03, 29, 37, 02

- https://www.easycalculation.com/statistics/t-distribution-critical-value-table.php

Exercise3

- X1平均值 = 27.15　X2平均值 = 11.95
- X1樣本變異數 = 156.5　X2樣本變異數 = 213.74
- T值 = 3.54
- df(自由度) = 20+20-2 = 38
- 因α值為0.05，所以查表得知臨界值是2.03
- u1 ≠ u2：如果|t| ≧ two-tailed的值(如果 3.54 ≧ 2.03)
- 所以可以拒絕虛無假設