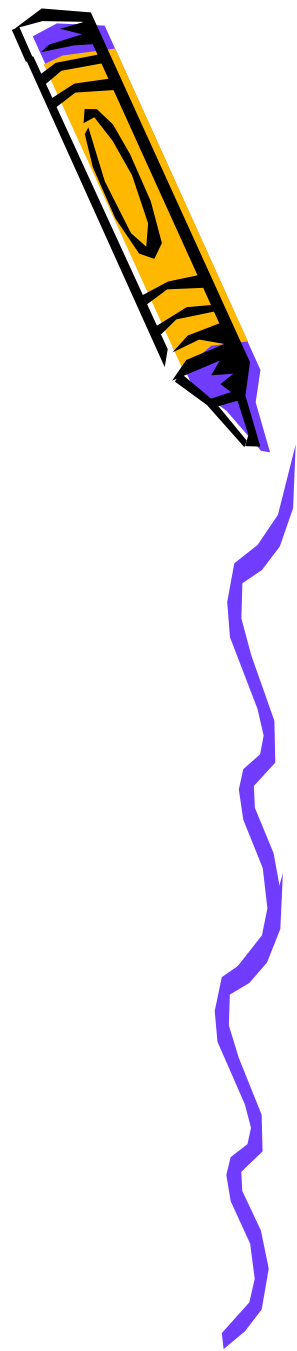
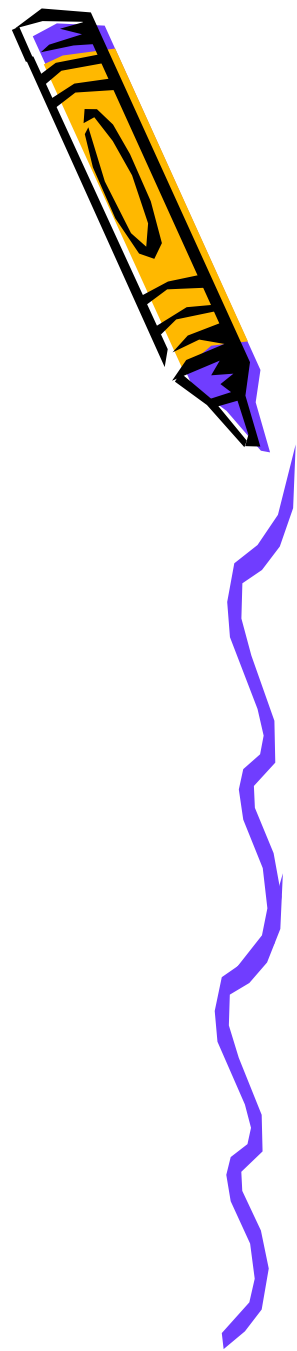


# 生物統計學

## 課程回顧



統計學是什麼？

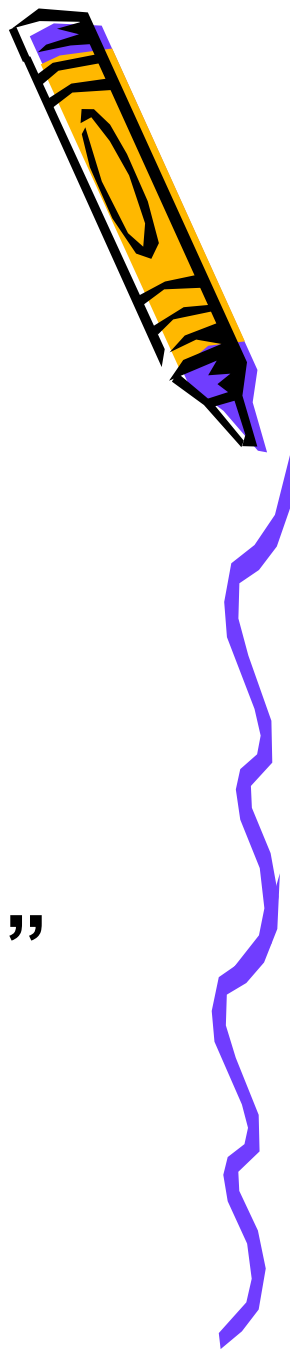


應用統計學最終目的：

比較各群資料間是否相同(似)?

例如：

“男生一群的身高”與“女生一群的身高”



## 某班級身高與性別資料如下：

座號	身高	性別
1	153	0 (女)
2	154	0
3	156	0
4	157	0
5	159	0
6	161	0
7	162	0
8	163	1 (男)
9	167	1
10	168	1
11	172	1
12	173	0
13	174	1
14	177	1
15	179	1
16	180	1

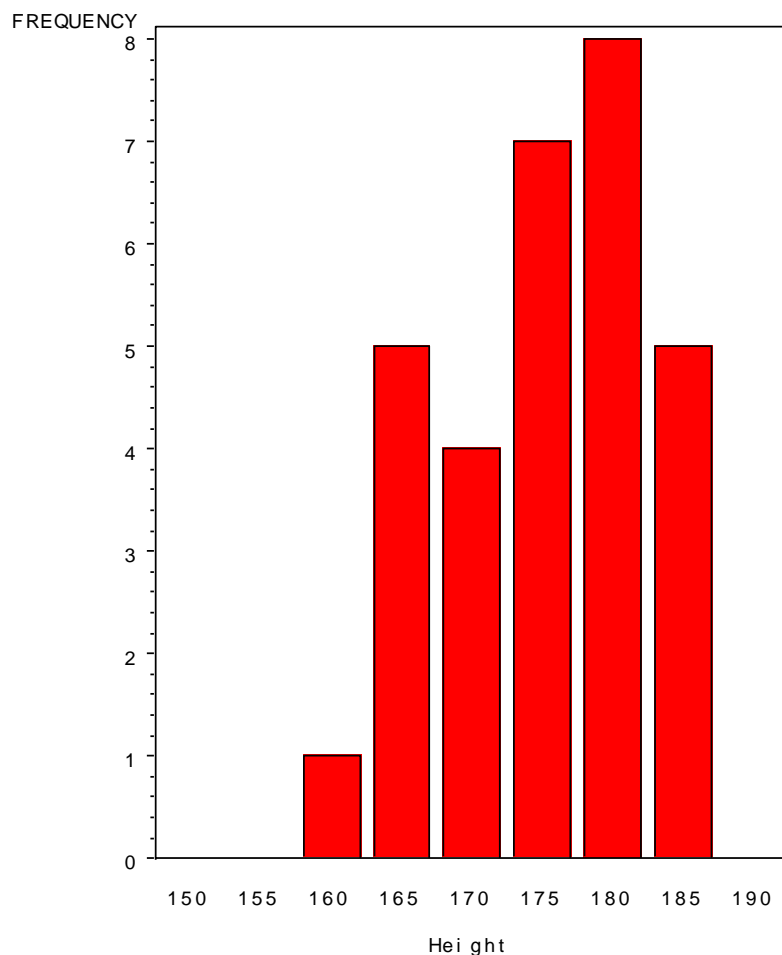
座號	身高	性別
17	181	1
18	182	1
19	184	1
20	185	1
21	150	0
22	151	0
23	153	0
24	154	0
25	155	0
26	156	0
27	159	0
28	160	1
29	162	0
30	165	1
31	167	1
32	168	1

座號	身高	性別
33	173	1
34	175	1
35	176	1
36	180	1
37	181	1
38	182	1
39	183	1
40	185	1
41	153	0
42	154	0
43	164	1
44	165	0
45	169	1
46	171	0
47	174	1
48	177	1
49	182	1
50	183	1

繪圖  
(資料分佈)

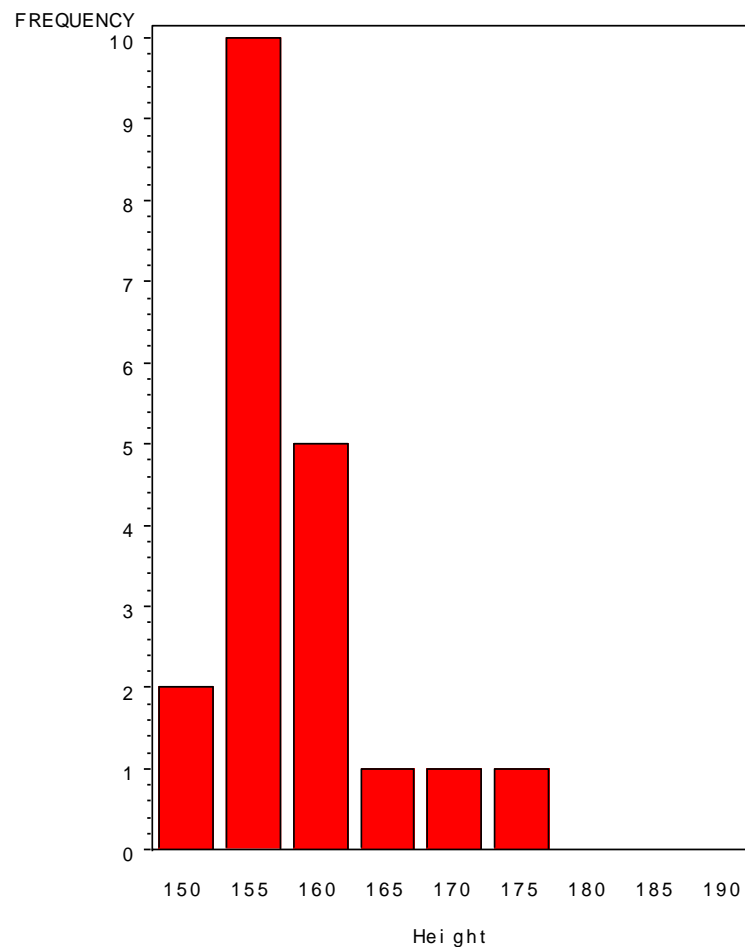
男同學  
人數 30(60%)  
身高平均 175.2  
標準差 7.4  
最高 185  
最矮 160

Sex=1



女同學  
人數 20(40%)  
身高平均 157.9  
標準差 6.2  
最高 173  
最矮 150

Sex=0



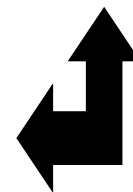
班上女生平均身高為**157.9公分**，標準差為**6.2公分**，**30名男生**平均身高**175.2公分**，標準差為**7.4公分**，請問**男、女生身高分佈是否不同？男、女生身高差別的95%信賴區間**是多少？

Variable	Sex	N	Lower CL Mean	Mean	Upper CL Mean	Lower CL Std Dev	Std Dev	Upper CL Std Dev	Std Err	Minimum	Maximum
Height	0	20	154.98	157.9	160.82	4.7454	6.2399	9.1139	1.3953	150	173
Height	1	30	172.45	175.2	177.95	5.8576	7.355	9.8875	1.3428	160	185
Height	Diff (1-2)		-21.33	-17.3	-13.27	5.7833	6.9351	8.664	2.002		

差別的95%信賴區間

T-Tests 利用統計檢定判別男、女生身高是否不同？

Variable	Method	Variances	DF	t Value	Pr >  t
Height	Pooled	Equal	48	-8.64	<.0001
Height	Satterthwaite	Unequal	45.1	-8.93	<.0001

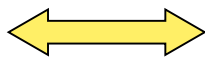


# 統計學 即



探討”因子”與”因子”間的關係。

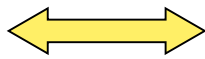
因子1



因子2



性別



身高





使用統計方法前先了解



變項(因子)的測量方式

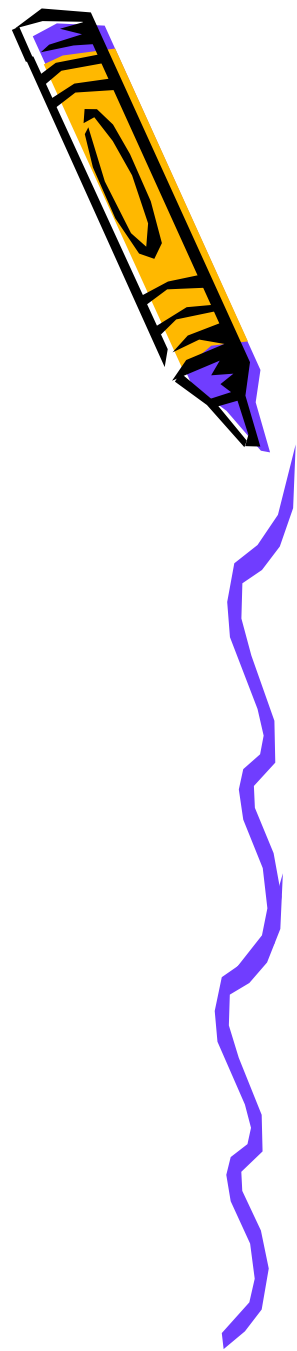




# 名詞定義

## 變項是什麼

資料中的單一項目名稱，例如性別，年齡。



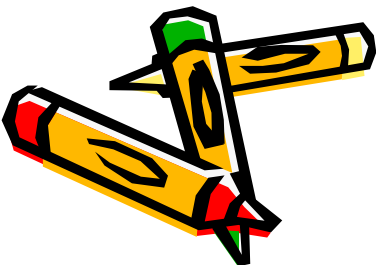
# 變項的測量



## 非連續性(Discrete) 與連續性(Continuous)變項

**非連續性(類別)**：測量單位不可加以細分，為點計變項，  
例如班級中人數。

**連續性(等距)**：測量單位可以加以細分，例如重量單位。



# 變項的測量

表 3.1 100 名抽樣資料，母體為 1969 年 7683 名檀香山心臟研究的對象

編號	教育程度	體重 (公斤)	身高 (公分)	年齡	抽菸習慣	居家活動力	血糖濃度	血中膽固醇	血壓 (收縮壓)	肥胖指標
1	2	70	165	61	1	1	107	199	102	40.0361
2	1	60	162	52	0	2	145	267	138	41.3808
3	1	62	150	52	1	1	237	272	190	37.8990
4	2	66	165	51	1	1	91	166	122	40.8291
5	2	70	162	51	0	1	185	239	128	39.3082
6	4	59	165	53	0	2	106	189	112	42.3838
7	1	47	160	61	0	1	177	238	128	44.3358
8	3	66	170	48	1	1	120	223	116	42.0663
9	5	56	155	54	0	2	116	279	134	40.5138
10	2	62	167	48	0	1	105	190	104	42.1942
11	4	68	165	49	1	2	109	240	116	40.4248
12	1	65	166	48	0	1	186	209	152	41.2862
13	1	56	157	55	0	2	257	210	134	41.0365
14	2	80	161	49	0	1	218	171	132	37.3648
15	3	66	160	50	0	2	164	255	130	39.5918
16	4	91	170	52	0	2	158	232	118	37.7951
17	3	71	170	48	1	1	117	147	136	41.0547
18	5	66	152	59	0	2	130	268	108	37.6123
19	1	73	159	59	0	2	132	231	108	38.0444
20	4	59	161	52	0	1	138	199	128	41.3563
21	1	64	162	52	1	1	131	255	118	40.5001
22	3	55	161	52	1	1	88	199	134	42.3356
23	2	78	175	50	1	1	161	228	178	40.9582
24	2	59	160	54	0	1	145	240	134	41.0995
25	3	51	167	48	1	2	128	184	162	45.0326
26	3	83	171	55	0	1	231	192	162	39.2016
27	2	66	157	49	1	2	78	211	120	38.8495

# 變項的型態



編號	教育程度	體重 (公斤)	身高 (公分)	年齡	抽菸 習慣	居家 活動力	血糖 濃度	血中 膽固醇	血壓 (收縮壓)	肥胖 指標
96	2	68	155	67	0	2	173	251	118	37.9749
97	1	58	170	62	0	1	148	187	162	43.9178
98	3	68	160	55	0	1	110	290	128	39.1998
99	5	60	159	50	0	2	188	238	130	40.6144
100	2	61	160	54	1	1	208	218	208	40.6453

變項譯碼：

教育程度：1 = 無，2 = 小學，3 = 中學，4 = 高中，5 = 技術專科  
6 = 大學

體重單位：公斤

身高單位：公分

抽菸習慣：0 = 無，1 = 有

居家活動力：1 = 幾乎坐著，2 = 適度，3 = 積極

血糖濃度單位：mg 百分比

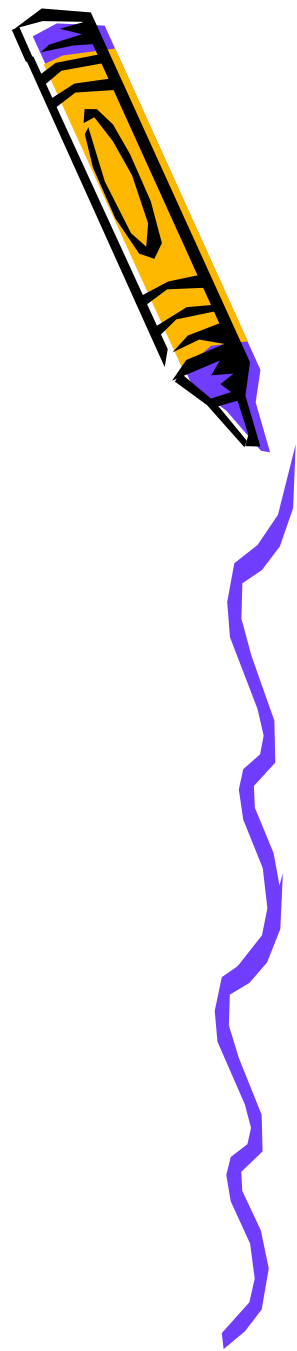
血中膽固醇單位：mg 百分比

血壓：毫米水銀柱 (mmHg)

肥胖指標：身高 /  $\sqrt[3]{\text{體重}}$



# 連續性(等距變項)的敘述



男同學  
人數 30(60%)

身高平均 175.2

標準差 7.4

最高 185

最矮 160

女同學  
人數 20(40%)

身高平均 157.9

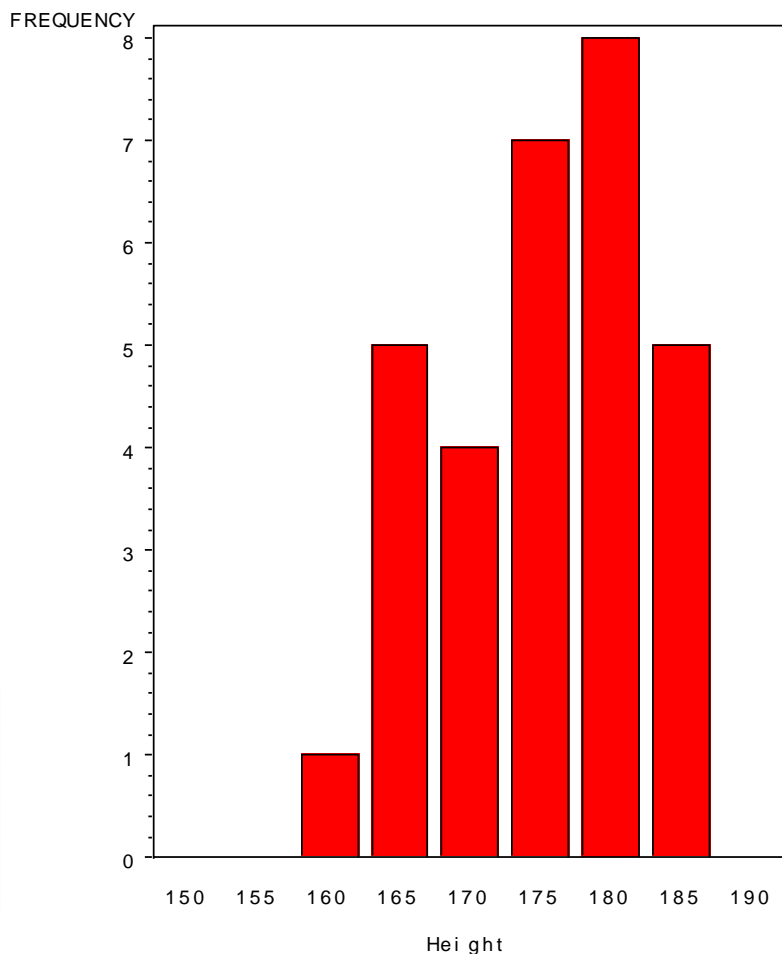
標準差 6.2

最高 173

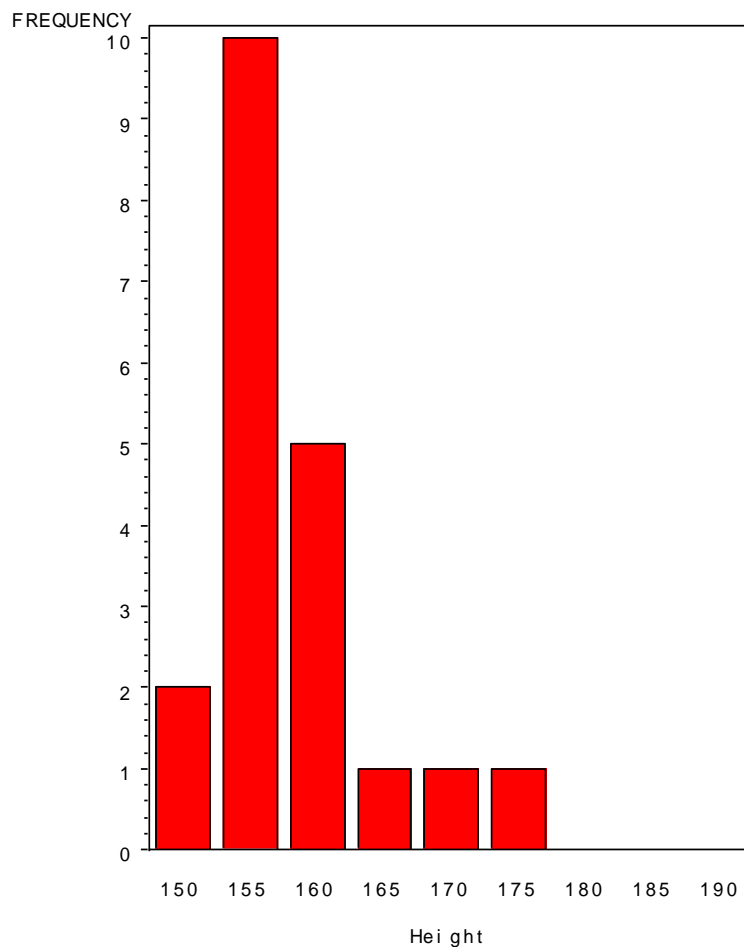
最矮 150

繪圖  
(資料分佈)

Sex=1

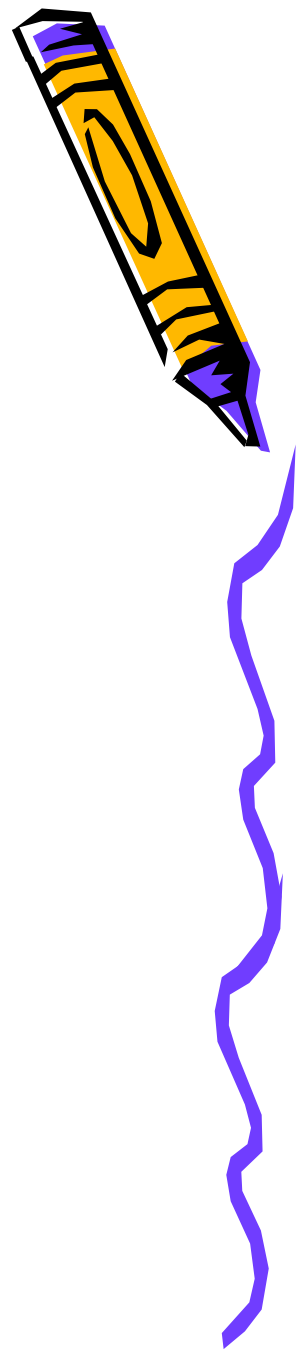


Sex=0

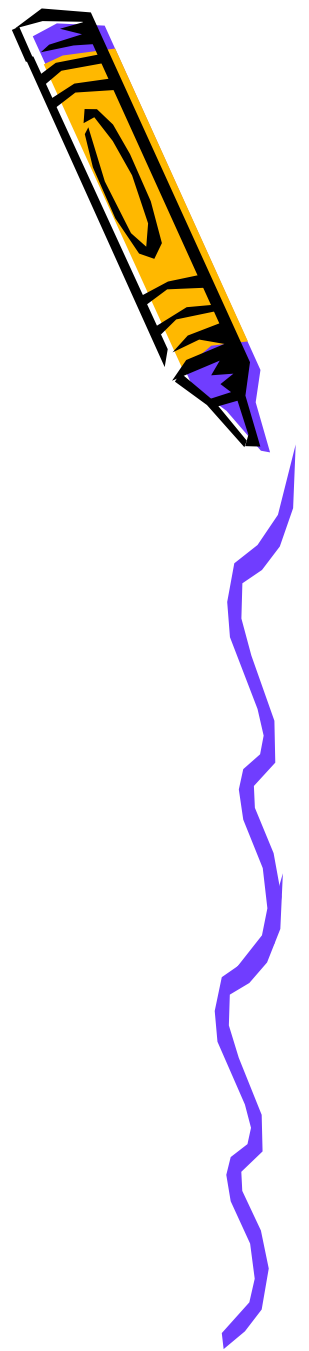


## 分析指標(即統計學運用的指標)

1. 中央趨勢性(central tendency)
2. 資料變異性(variability)



# 中央趨勢性



資料處理時，會先找出資料群集中的數值，常用到的兩種指標為平均數、中位數，即稱中央趨勢測量，意指資料群趨向某固定數值的趨勢。





## 平均數

1. 算數平均數（Mean； $\mu$ 、 $\bar{x}$ ）簡稱平均數，將所有資料加總除上資料的數目。
2. 算數平均數可視為所有資料的平衡點或重心，考慮到每個資料與平衡正和負數的差距。
3. 平均數受每一個資料影響，因此較極端的數值會影響，甚至扭曲平均數，所以並不一定能代表所有資料的數值。

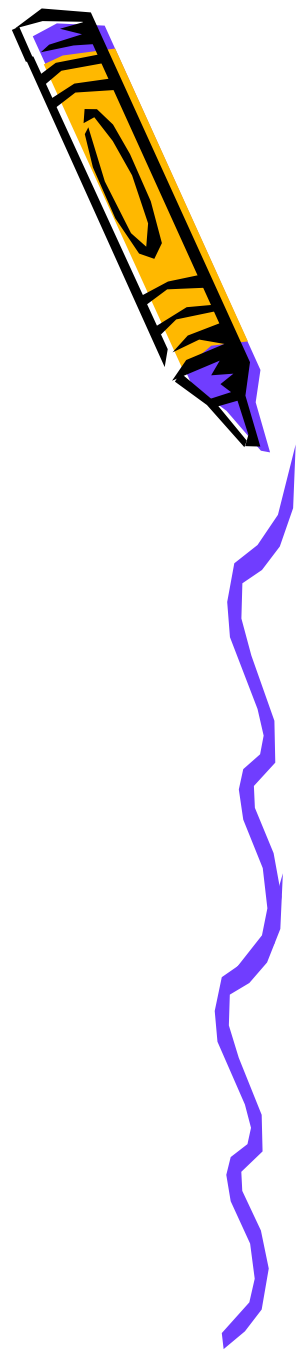


# 平均數

## 特性與公式

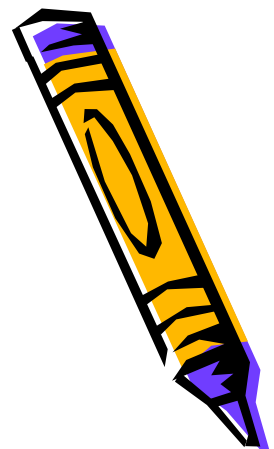
1. 每個個體的數值均包括在計算之中。
2. 每個個體的數值所佔的份量都是一樣。

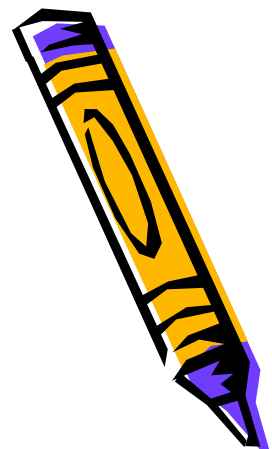
$$\mu = \frac{\sum_{i=1}^N X_i}{N} \quad \text{—} \quad \bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$



## 中位數

1. 中位數（Median； $M_d$ ）將資料數值從小排到大後，分為兩邊相等資料個案數目，其中點即中位數(median)。
2. 中位數是不受極端值影響的中央趨勢測量。





# 中位數

例子：

若一公司會計部門員工薪資26000、30000、30000、34000和70000，則此會計部門員工薪資中位數為何？

26000, 30000, 30000, 34000, 70000



中位數

若為26000, 30000, 30000, 32000, 34000, 70000，則為何？



$$\frac{30000 + 32000}{2} = 31000$$



# 資料變異性

1. 表達資料的同質性(homogeneous)與異質性(heterogeneous)情形，即描述資料的變異情形。

## 2. 常用的指標

最大值(maximum)

最小值(minimum)

差(全)距(range)

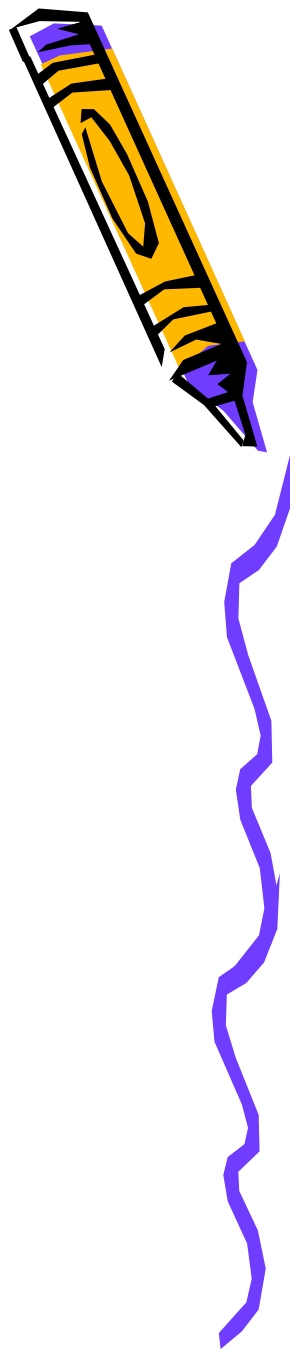
} 中位數

平均差(mean deviation)

變異數(variance)

標準差 (standard deviation)

} 平均數



# 資料變異性

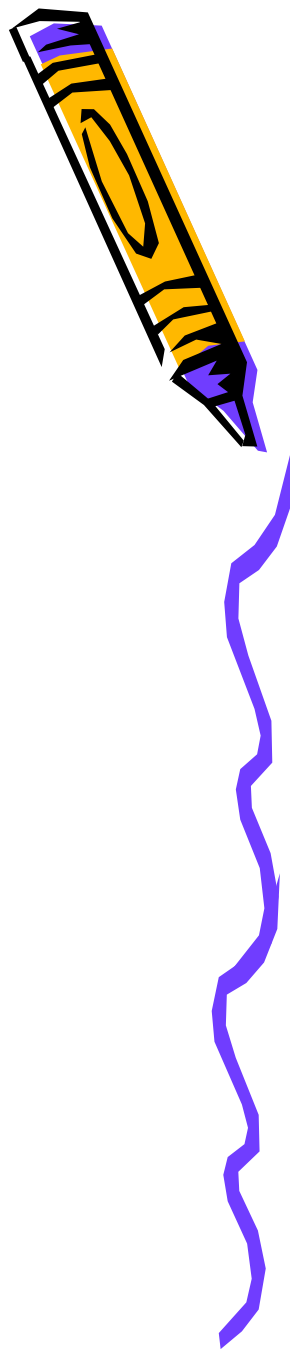
**A: 5, 10, 15, 50, 85, 90, 95**

**B: 35, 40, 45, 50, 55, 60, 65**

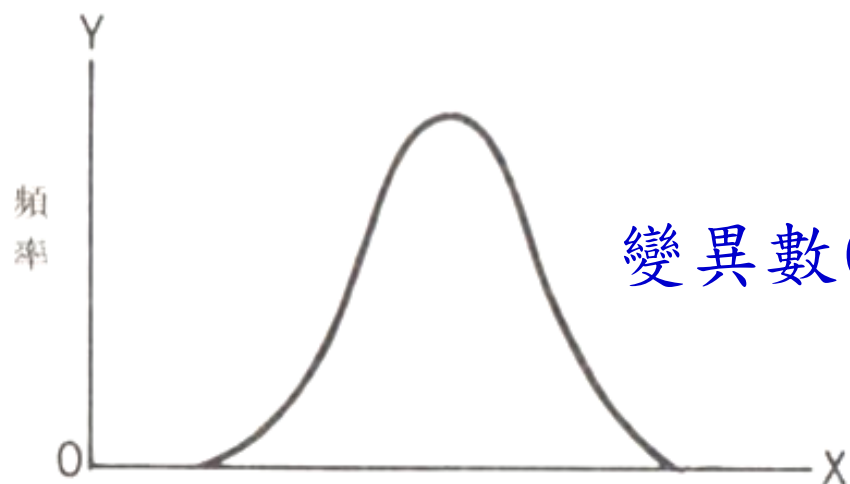
**平均值A: 50**

**平均值B: 50**

**A: heterogeneous**  
**B: homogeneous**



## 變異性不同的差異



變異數(性)小



變異數(性)大



# 資料變異性



變異數 (Variance,  $\sigma^2$ ,  $s^2$ )

標準差 (standard deviation,  $\sigma$ ,  $s$ )

1. 變異數和標準差是測量資料的變異情形，數值愈大表示觀察值愈不相同。
2. 所有觀察值減平均值的平方加總，再除上母全體數(N) [母全體]，或樣本數減一(n-1)[樣本]。
3. 標準差 ( $\sigma$ ) = 變異數的平方根 ( $\sqrt{\sigma^2}$ )。





# 資料變異性

## 變異數

若資料為母全體

$$\sigma^2 = \frac{\sum (\chi - \mu)^2}{N} = \frac{\sum \chi^2 - \frac{(\sum \chi)^2}{N}}{N} = \frac{\sum \chi^2 - N\mu^2}{N}$$

若資料為樣本

$$S^2 = \frac{\sum (\chi - \bar{\chi})^2}{n-1} = \frac{\sum \chi^2 - \frac{(\sum \chi)^2}{n}}{n-1}$$



# 資料變異性



例子：

若一公司會計部門員工薪資26000、30000、30000、34000和70000，則此會計部門員工薪資變異數和標準差為何？

	26	30	30	34	70 (千)
平均值	38				
離均差	-12	-8	-8	-4	32
離均差絕對值	12	8	8	4	32
離均差平方	144	64	64	16	1024

$$\begin{aligned}\text{變異數} = \sigma^2 &= \frac{\sum (x - \mu)^2}{N} = \frac{144 + 64 + 64 + 16 + 1024}{5} (\text{千})^2 \\ &= 262400000\end{aligned}$$



$$\text{標準差} = \sigma = \sqrt{\sigma^2} = \sqrt{262400000} = 16199$$

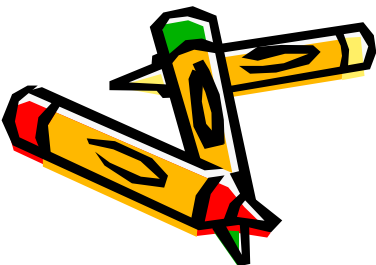
# 資料變異性

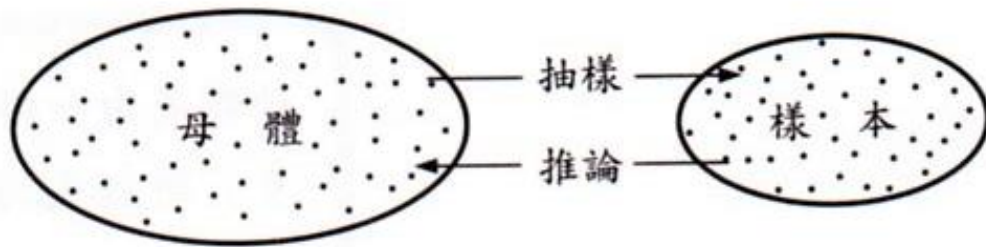


## 標準誤(Standard Error of Means ; SEM, SE)

樣本的平均值間標準差  $\hat{=}$  眾多樣本 的 平均值間 標準差

$$SEM = \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad \text{或} \quad SEM = S_{\bar{x}} = \frac{S}{\sqrt{n}}$$





樣本數目: 無限  
樣本組合: 有限組合

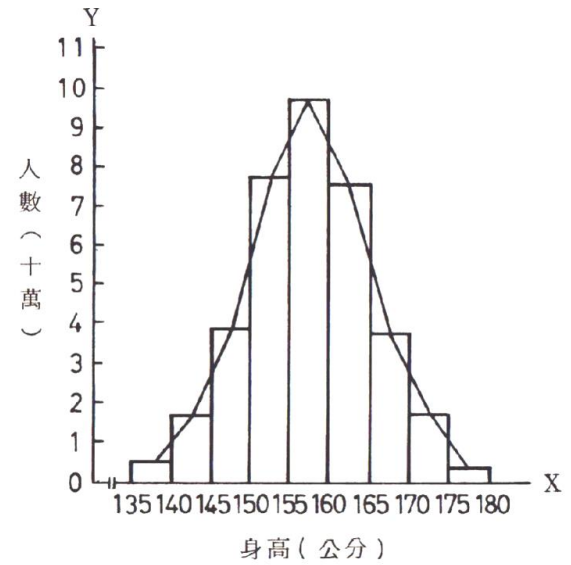


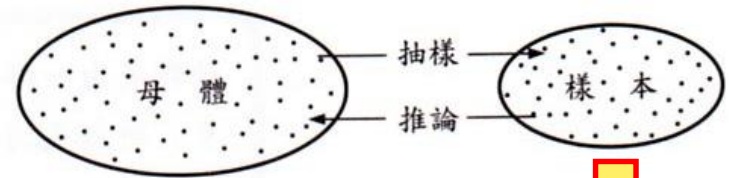
圖 4-1 台灣地區正常成年女子身高次數分佈假想圖

眾多  $\chi \Rightarrow \chi_i$

次數或百分比

$\chi_i$   
 $\sigma$

cm



眾多  $\bar{\chi} \Rightarrow \bar{\chi}_i$

平均值

$\bar{\chi}_i$   
 $\sigma_{\bar{\chi}}$

$\chi_i$

# 中央極限定理

## 樣本平均數的分布與中央極限定理(Central Limit Theorem)

對固定 $\mu$ 及 $\sigma^2$ 的母群（在一定樣本數 $n$ 之下）作隨機且每次放回的**重複抽樣**時，樣本平均值的抽樣分布有幾個特性：

1. 平均值等於母群平均值  $\mu_x = \mu$
2. 樣本平均數抽樣分佈的標準差等於母體標準差除上樣本數的平方根  $\sigma_x = \frac{\sigma}{\sqrt{n}}$
3. 若母群為常態分布，則抽樣分布亦為常態分布

若母群不為常態分布，則在樣本數夠時，其抽樣分布亦可趨近常態分布。

不管原觀察值分布是否為常態分布，隨機抽樣的平均值分布必趨近常態分布。若每次抽樣**30**個樣本以上，則形成的分布與**Z**分布相同。

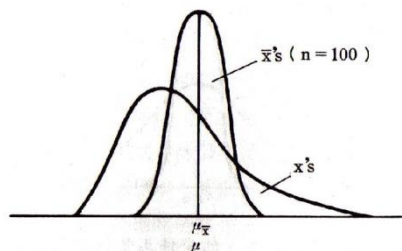


圖 5-3 母全體 ( $x$ 's) 為偏右分佈，樣本平均值的分佈為常態分佈。

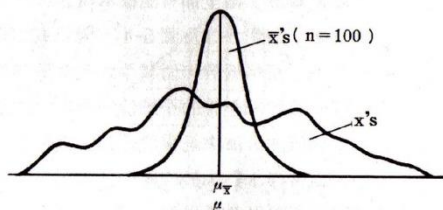
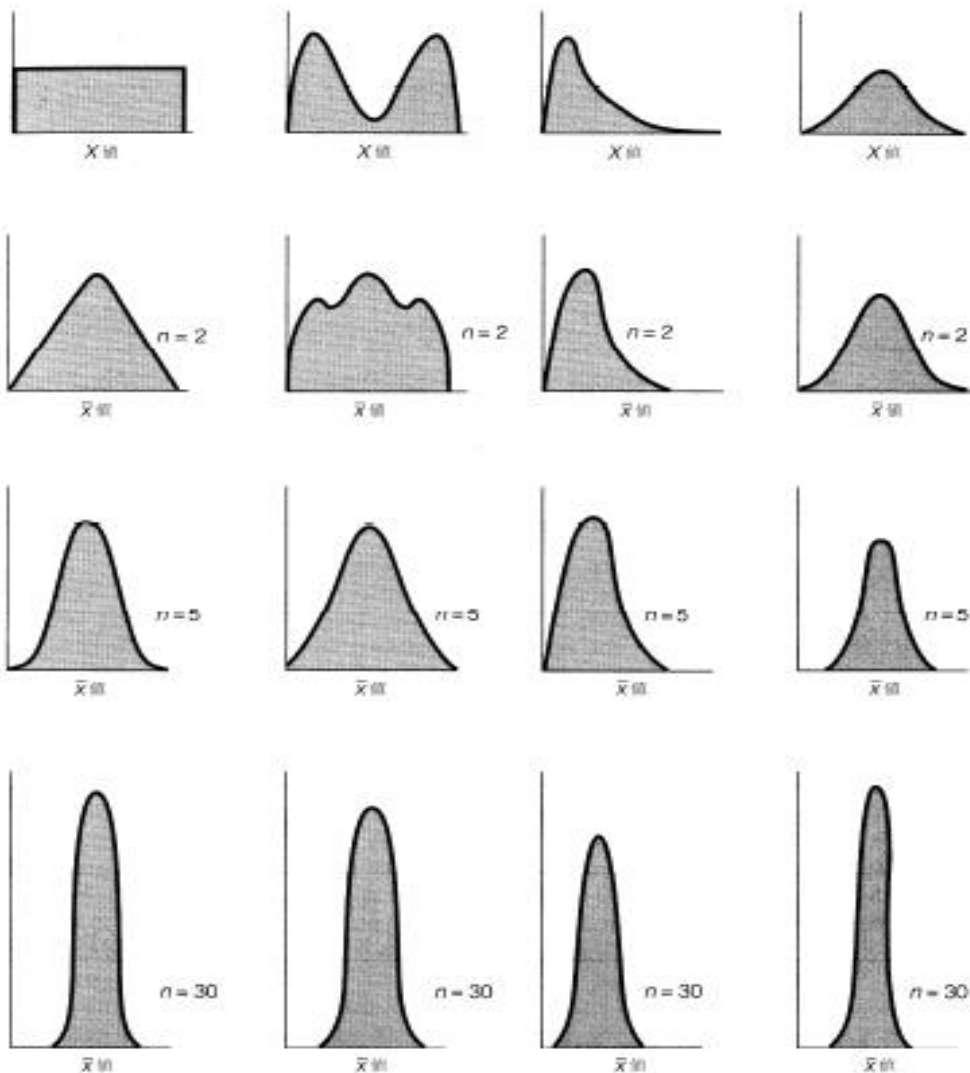
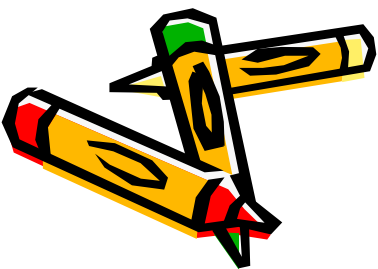
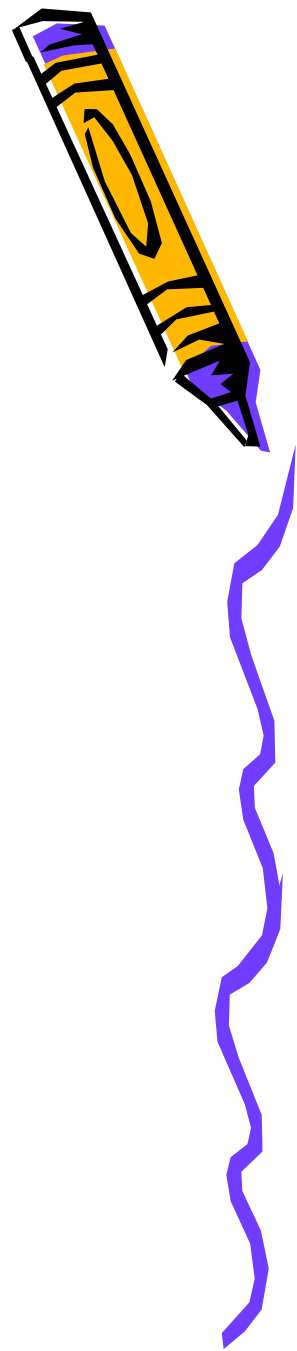


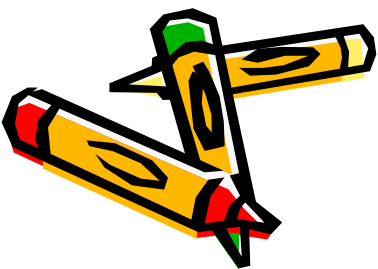
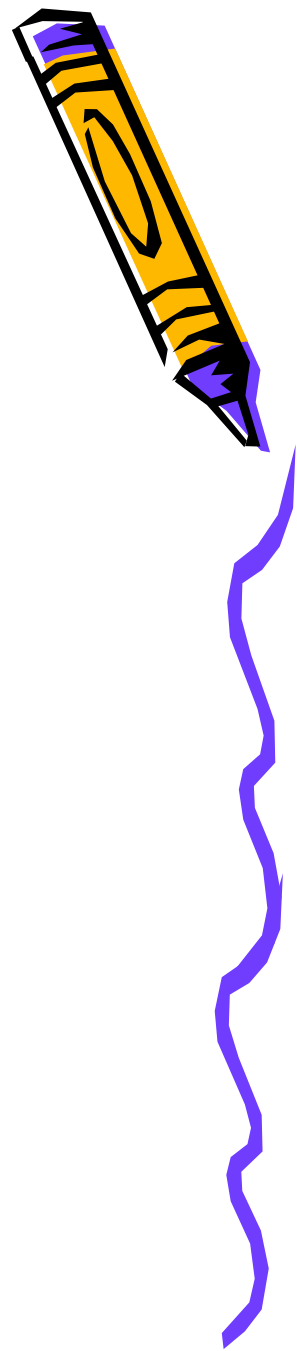
圖 5-4 母全體 ( $x$ 's) 為不規則分佈，樣本平均值的分佈為常態分佈。



# 常態分佈



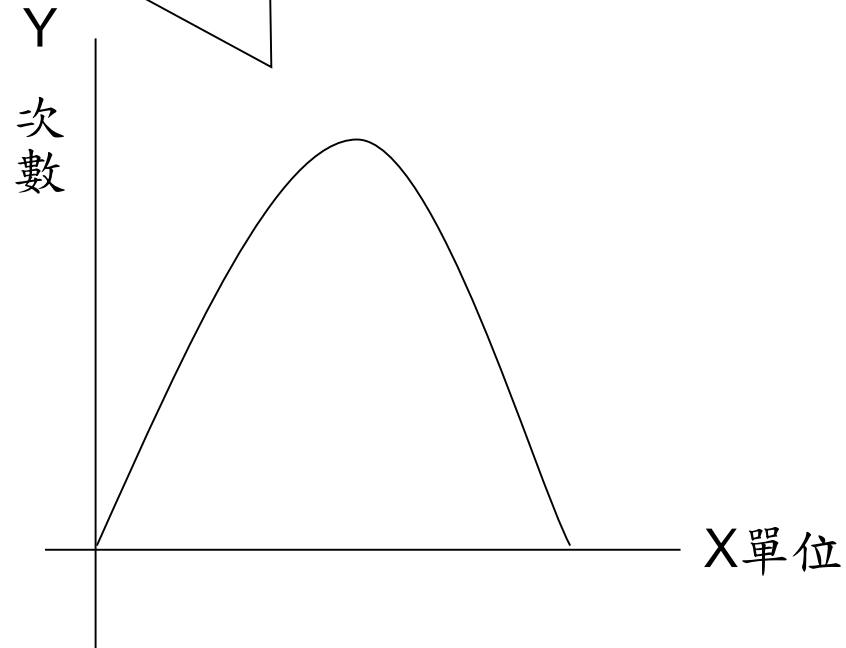
# 資料分布

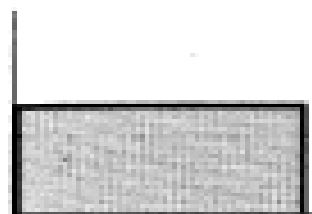




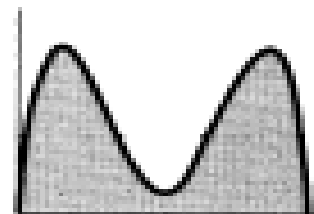
# 資料分布

假設一樣本有 $N$ 個人，測量其身高後，將資料以圖表呈現，則其分布可能為何？





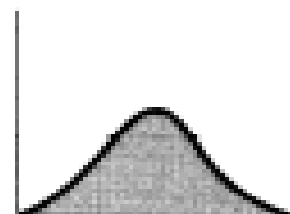
$X$  值



$X$  值

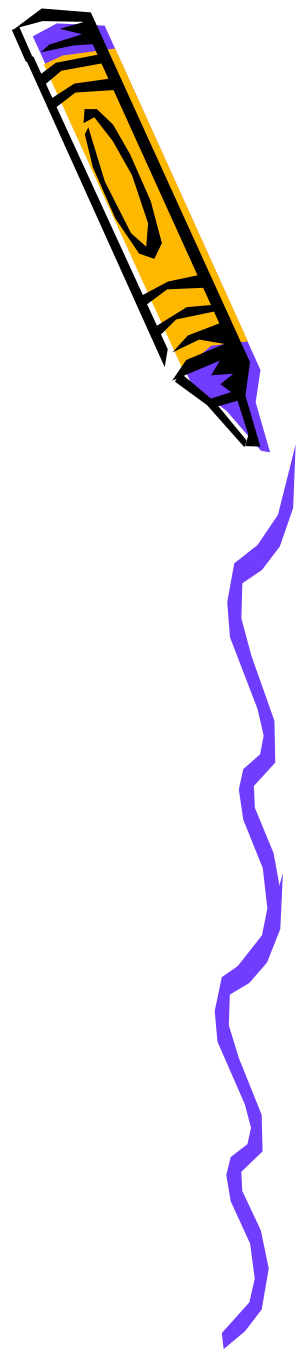


$X$  值



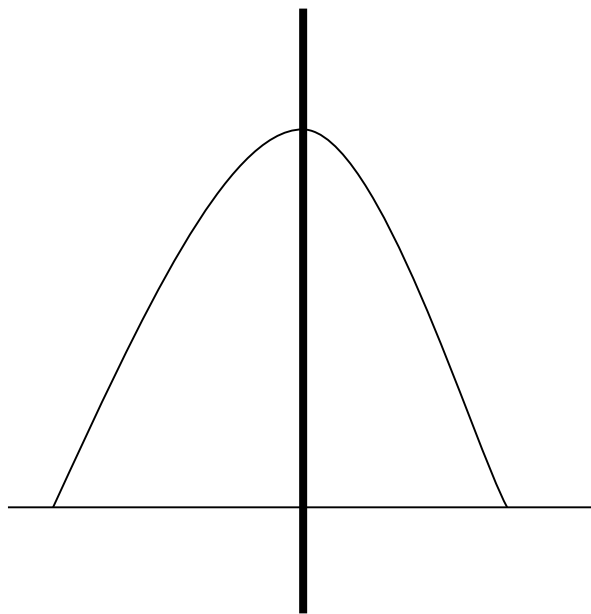
$X$  值

# 常態分佈 normal distribution



# 常態分佈

資料的分佈是一個對稱的分佈，故三個測量值都相同。



平均數 = 中位數 = 眾數



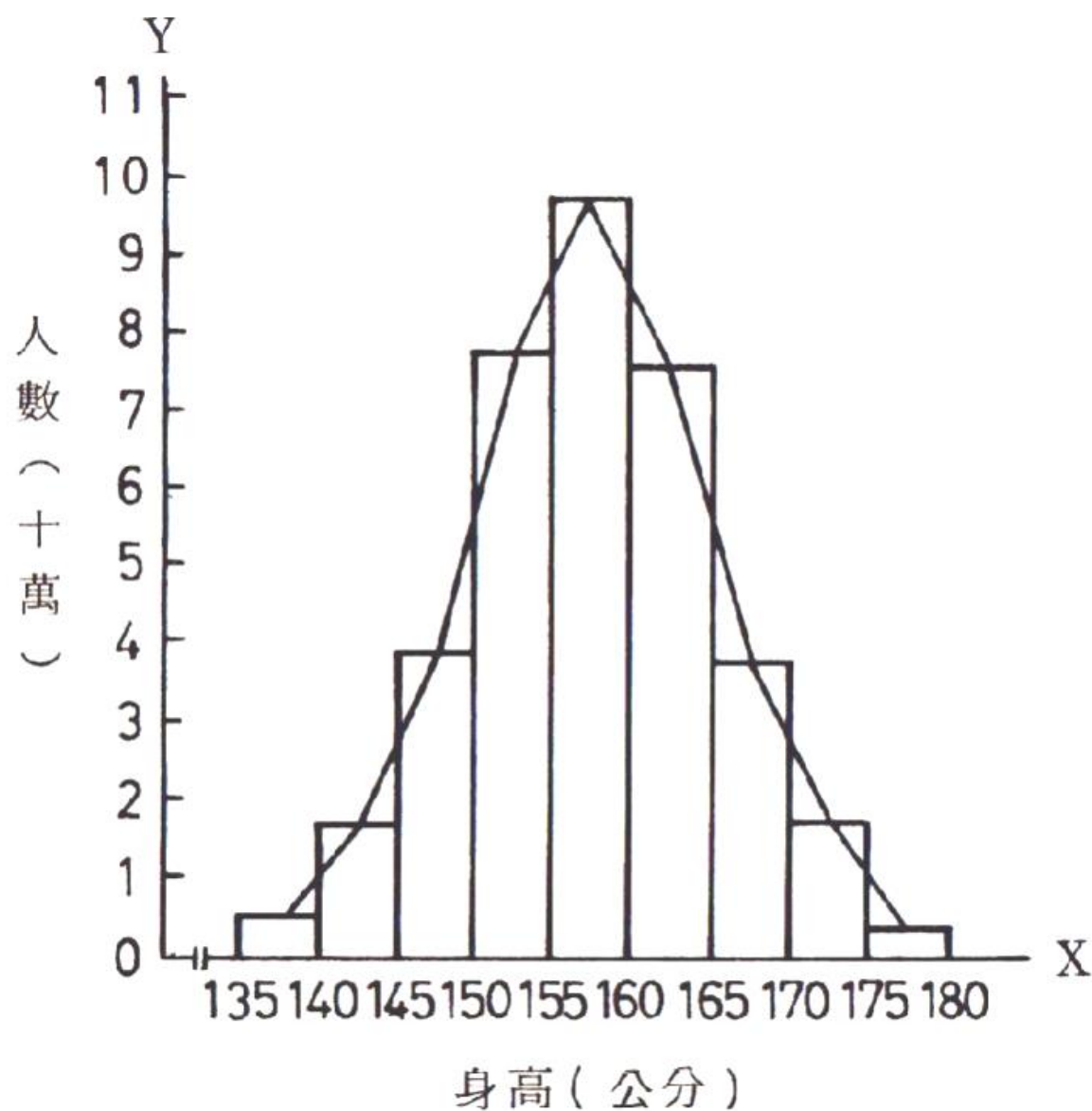


圖 4-1 台灣地區正常成年女子身高次數分佈假想圖

## 資料分佈

各個觀察或測量值出現的次數或頻率。

## 常態分佈(normal distribution)

一種現象受眾多因素同時共同影響，個案值分佈通常為常態分佈。

### 特性

1. 愈接近平均值，個體數愈多或頻率愈高。
2. 愈遠離平均值，個體數愈少或頻率愈低。
3. 向外左右延伸至負無限大或正無限大。
4. 所有常態分佈都可算出在曲線下的面積，無論平均數或標準差多大或多小，曲線下某對稱兩點間的相對面積永遠相同。
5. 曲線分佈圖呈現鐘型且對稱（又稱高斯曲線）。
6. 常態分佈的平均值、中位數與眾數皆相同。

# 常態分佈



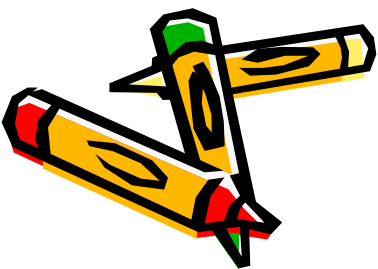
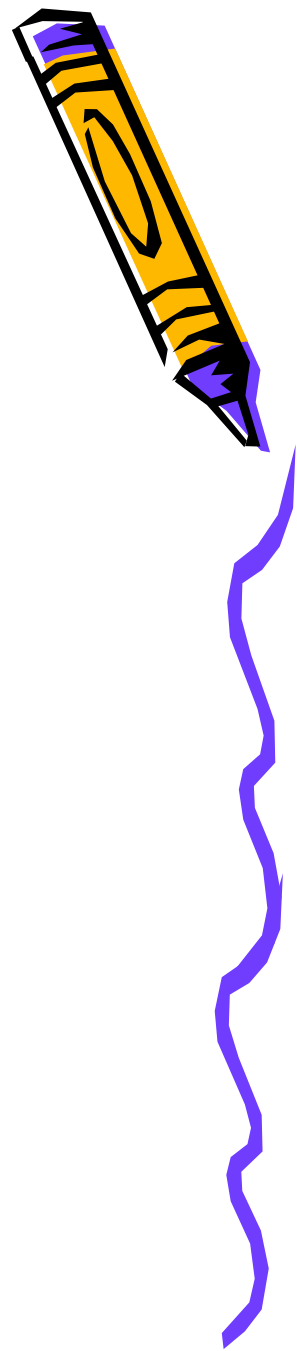
## 重要性

1. 很多變項（測量）的分佈都趨近常態分布，例如身高、膽固醇值....等
2. 許多的統計理論和方法理論之發展，均建立在該資料的分佈需接近常態分佈的假說下。



# 統計推論

## (推論性生物統計學)





# 統計推論的方法



信賴區間估計(檢定) (parameter estimation)

統計假說檢定 (hypothesis testing)

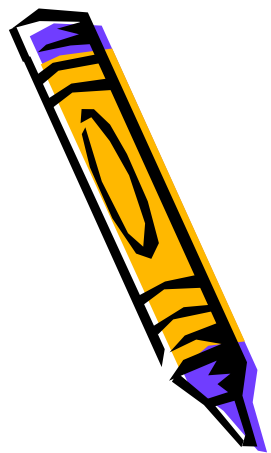


# 母數估計

## 信賴區間估計(confidence Interval estimation)

信賴區間是利用樣本平均值( $\bar{x}$ )估計母全體平均值( $\mu_x$ )，提供一個誤差範圍來標示。

**95% Confident Interval of  $\mu_x$  : 95% C. I. of  $\mu_x$**



# 母數估計

## 信賴區間估計(confidence Interval estimation)

即設定母全體平均值 $\mu$ ，在樣本平均值左右範圍內含95%的區間做為估計。

$$P(a < \mu < b) = 0.95 \text{ 或}$$

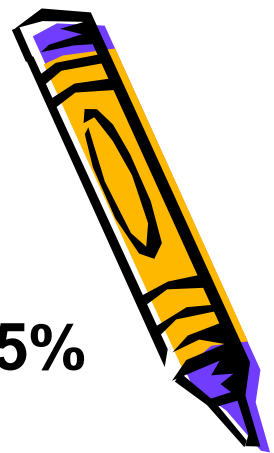
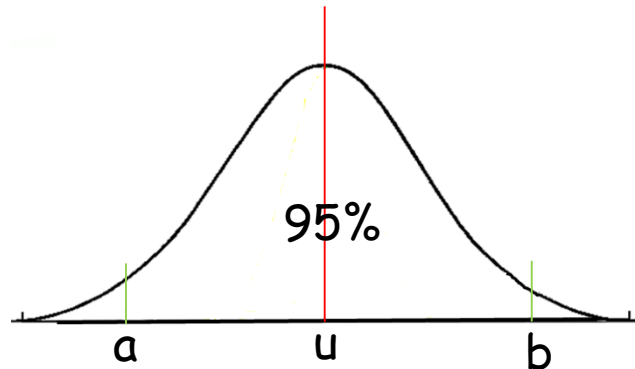
$$P(-1.96 \leq Z \leq 1.96) = 0.95$$

a: 信賴區間下限

b: 信賴區間上限

$$P(-1.96 \leq \frac{\bar{x} - \mu_x}{\sigma / \sqrt{n}} \leq 1.96) = 0.95$$

母數



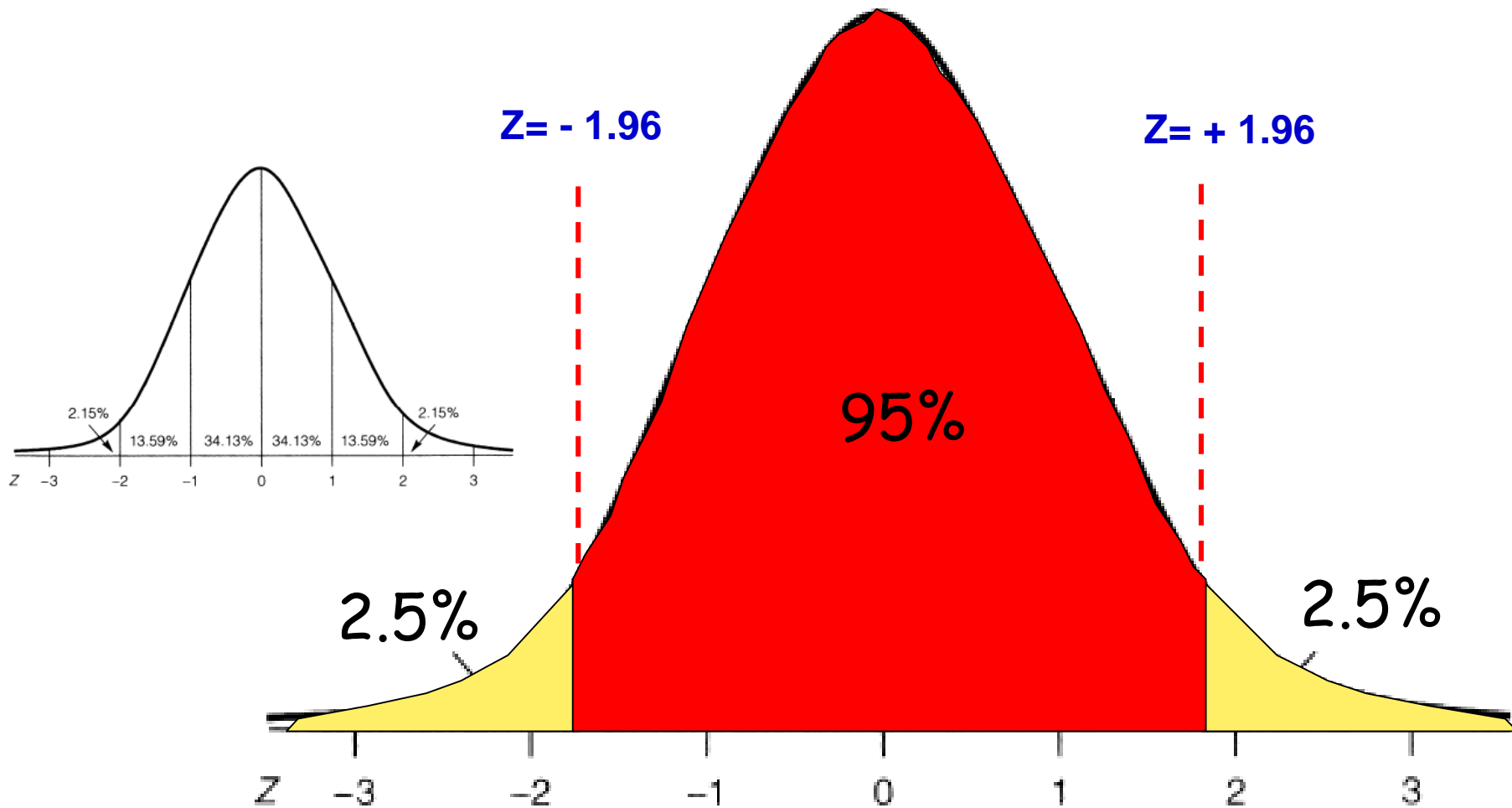
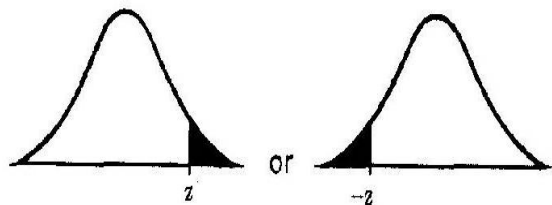


表 4-2 常態分佈表 (+ Z 以上或 - Z 以下之單尾面積)



1.5	.067	.066	.064	.063	.062	.061	.059	.058	.057	.056
1.6	.055	.054	.053	.052	.051	.049	.048	.048	.046	.046
1.7	.045	.044	.043	.042	.041	.040	.039	.038	.038	.037
1.8	.036	.035	.034	.034	.033	.032	.031	.031	.030	.029
1.9	.029	.028	.027	.027	.026	.026	.025	.024	.024	.023

# 母數估計

## 信賴區間估計(續)

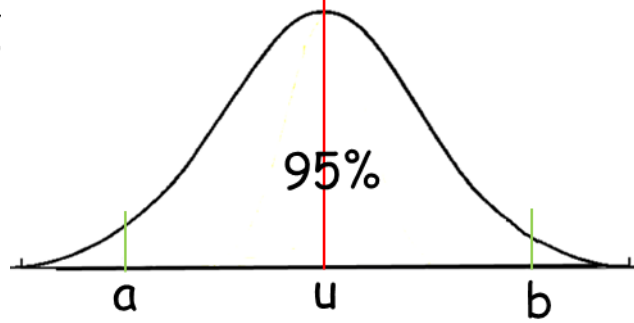
$$\therefore P(-1.96 \leq \frac{\bar{x} - \mu_x}{\sigma / \sqrt{n}} \leq +1.96) = 0.95$$

$$\therefore P(-1.96 \frac{\sigma}{\sqrt{n}} \leq \bar{x} - \mu_x \leq +1.96 \frac{\sigma}{\sqrt{n}}) = 0.95$$

$$\therefore P(1.96 \frac{\sigma}{\sqrt{n}} \geq -\bar{x} + \mu_x \geq -1.96 \frac{\sigma}{\sqrt{n}}) = 0.95$$

$$\therefore P(\bar{x} + 1.96 \frac{\sigma}{\sqrt{n}} \geq \mu_x \geq \bar{x} - 1.96 \frac{\sigma}{\sqrt{n}}) = 0.95$$

$$\therefore P(b \geq \mu_x \geq a) = 0.95$$



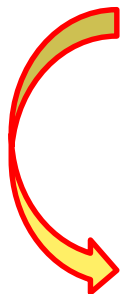
### 例題：

64 位病人收縮壓的平均值為 140 (mmHg)，若假設母體之標準差為 25，試求母全群平均值 ( $\mu_x$ ) 的 95% 信賴區間？

$$P(-1.96 \leq Z \leq 1.96) = 0.95$$



$$P(-1.96 \leq \frac{\bar{x} - \mu_x}{\sigma / \sqrt{n}} \leq 1.96) = 0.95$$



$$\therefore P(\bar{x} + 1.96 \frac{\sigma}{\sqrt{n}} \geq \mu_x \geq \bar{x} - 1.96 \frac{\sigma}{\sqrt{n}}) = 0.95$$

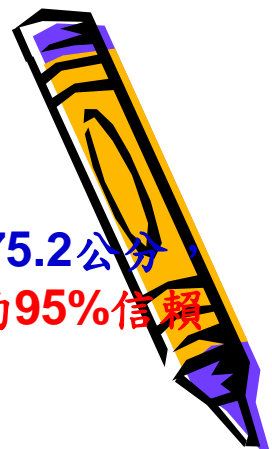
$$\therefore \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}} \geq \mu_x \geq \bar{x} - 1.96 \frac{\sigma}{\sqrt{n}}$$

$$\therefore 140 + 1.96 \frac{25}{\sqrt{64}} \geq \mu_x \geq 140 - 1.96 \frac{25}{\sqrt{64}}$$

$$\therefore 146 \geq \mu_x \geq 134$$

**即 95% C.I. of  $\mu_x = 134 \sim 146$**

班上女生平均身高為157.9公分，標準差為6.2公分，30名男生平均身高175.2公分，標準差為7.4公分，請問男、女生身高分佈是否不同？男、女生身高差別的95%信賴區間是多少？

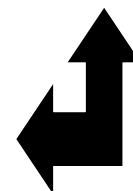


Variable	Sex	N	Lower CL		Upper CL		Lower CL		Upper CL		Std Err	Minimum	Maximum
			Mean	Mean	Mean	Std Dev	Std Dev	Std Dev					
Height	0	20	154.98	157.9	160.82	4.7454	6.2399	9.1139	1.3953	150	173		
Height	1	30	172.45	175.2	177.95	5.8576	7.355	9.8875	1.3428	160	185		
Height	Diff (1-2)		-21.33	-17.3	-13.27	5.7833	6.9351	8.664	2.002				

差別的95%信賴區間

T-Tests 利用統計檢定判別男、女生身高是否不同？

Variable	Method	Variances	DF	t Value	Pr >  t
Height	Pooled	Equal	48	-8.64	<.0001
Height	Satterthwaite	Unequal	45.1	-8.93	<.0001

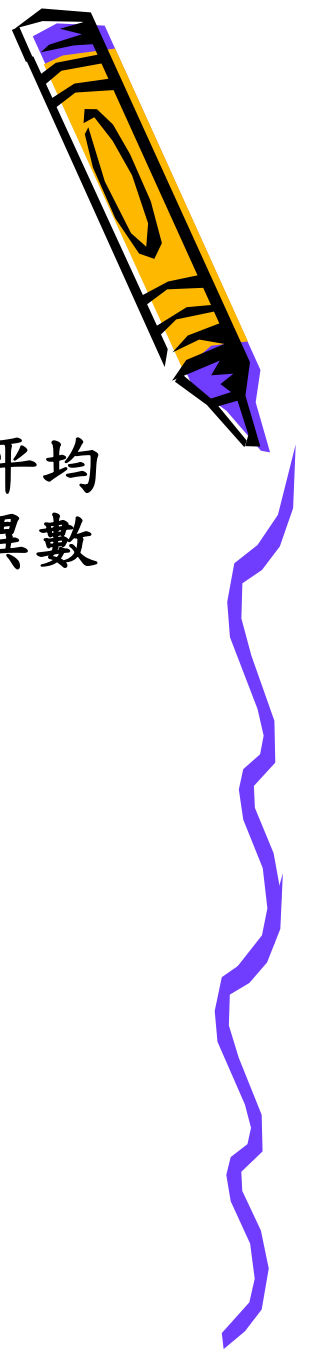


# 統計假說檢定(Hypothesis Testing)

## 檢定方法的前提

若由母全體隨機抽出之樣本平均數為 $\bar{x}$ ，則 $\bar{x}$ 為母全體平均數 $\mu$ 的不偏估計(unbiased estimate)。同樣樣本數變異數 $S^2$ 為母全體變異數 $\sigma^2$ 的不偏估計。

$$E(\bar{x}) = \mu \quad E(S^2) = \sigma^2$$





# 統計假說檢定(Hypothesis Testing)

## 統計的假說(Statistical Hypothesis)

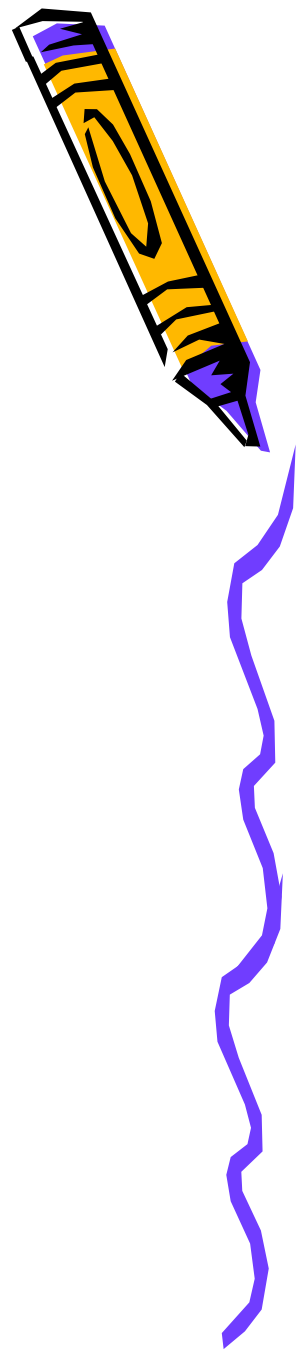
利用統計學檢定方法驗證科學性假說

虛無假說 ( Null Hypothesis,  $H_0$  )

例如: 無差異、無相關

對立假說 ( Alternative Hypothesis,  $H_1$  )

例如: 有差異、有相關

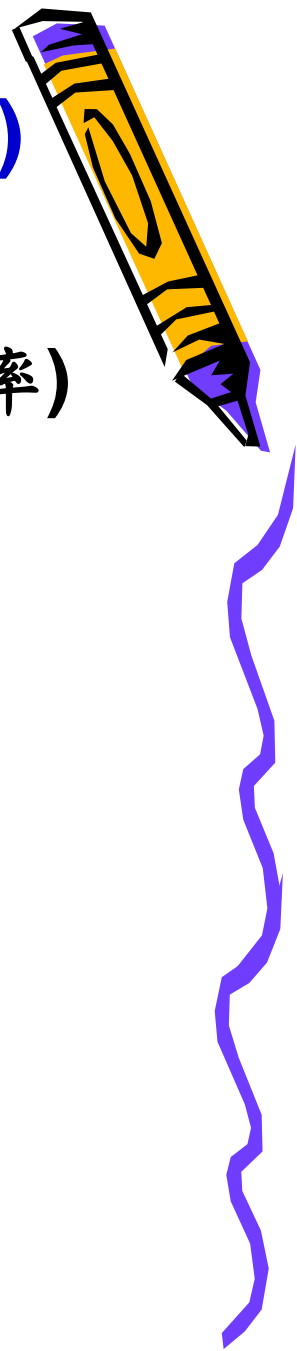


# 統計假說檢定(Hypothesis Testing)

## 執行步驟

1. 設立統計假說： $H_0$  與  $H_1$
2. 設定顯著水準值： $0.05$  ( $\alpha$ 值)
3. 根據 $\alpha$ 值，利用機率表找出臨界值。
4. 利用正確統計公式計算統計值  $\longrightarrow$  可查出機率P值
5. 利用臨界值或機率推論統計假說何者為正確  
當  $P \geq \alpha$  (即統計值  $\leq$  臨界值)，則接受虛無假說 ( $H_0$ )，  
故統計無顯著差異。  
當  $P < \alpha$  (即統計值  $>$  臨界值)，則拒絕虛無假說 ( $H_0$ )，  
接受對立假說 ( $H_1$ )，故統計有顯著差異。
6. 結論

# 統計假說檢定(Hypothesis Testing)



$\alpha$ 值：顯著水準值或拒絕虛無假設水準值(機率)

$p$  值 = 查表後所得之顯著機率  
= 接受虛無假設的機率  
= 造成第一類 (  $\alpha$ 誤差 ) 的機率



班上女生平均身高為157.9公分，標準差為6.2公分，30名男生平均身高175.2公分，標準差為7.4公分，請問男、女生身高分佈是否不同？男、女生身高差別的95%信賴區間是多少？(設  $\alpha = 0.05$ )

Variable	Sex	N	Lower CL		Upper CL		Lower CL		Upper CL		Std Err	Minimum	Maximum
			Mean	Mean	Mean	Mean	Std Dev	Std Dev	Std Dev	Std Dev			
Height	0	20	154.98	157.9	160.82	4.7454	6.2399	9.1139	1.3953			150	173
Height	1	30	172.45	175.2	177.95	5.8576	7.355	9.8875	1.3428			160	185
Height	Diff (1-2)		-21.33	-17.3	-13.27	5.7833	6.9351	8.664	2.002				

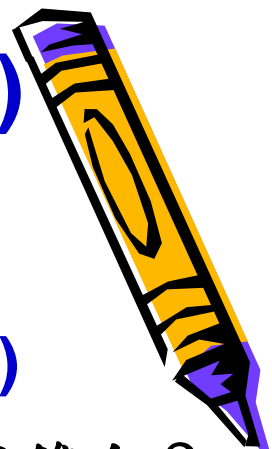
差別的95%信賴區間

T-Tests 利用統計檢定判別男、女生身高是否不同？

Variable	Method	Variances	DF	t Value	Pr >  t
Height	Pooled	Equal	48	-8.64	<.0001
Height	Satterthwaite	Unequal	45.1	-8.93	<.0001

P 值

# 統計假說檢定(Hypothesis Testing)



單尾檢定(one-tailed test)與雙尾檢定(two-tailed test)

檢定統計假說之重要問題：考慮 $\bar{x}$ 與 $\mu$ 之差異是單向或雙向？

## 單尾檢定

檢定平均年齡是否大於(小於)母體年齡平均數 $\mu$ 時

$\bar{x} - \mu$  此時**只**為正數或**只**為負數

## 雙尾檢定

檢定樣本平均年齡與母群體年齡平均數 $\mu$ 是否有差異時

$\bar{x} - \mu$  **可以**為正數，**亦可能**為負數。



# 統計假說檢定(Hypothesis Testing)



單尾檢定(one-tailed test)與雙尾檢定(two-tailed test)

適用問題

## 單尾檢定

某一新藥是否優於標準藥品？

空氣污染程度是否超過安全限制？

戒菸者之死亡率是否下降？

## 雙尾檢定

男性和女性之膽固醇含量是否有差異？



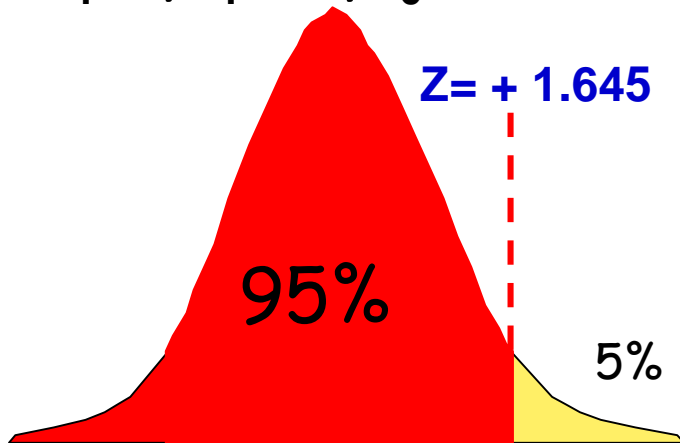
# 統計假說檢定(Hypothesis Testing)

單尾檢定(one-tailed test)與雙尾檢定(two-tailed test)

單尾檢定

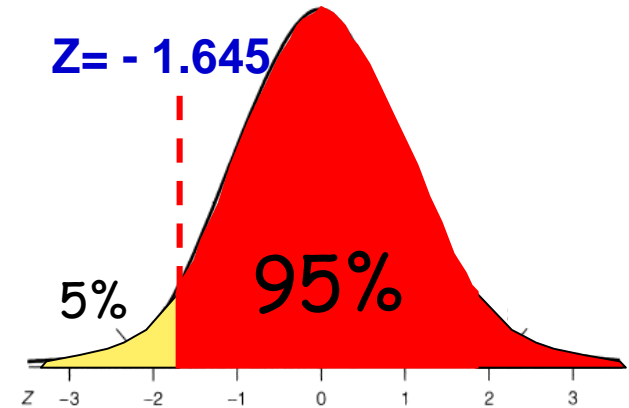
$$H_0 : \mu_1 \leq \mu_0$$

$$H_1 : \mu_1 > \mu_0$$



$$H_0 : \mu_1 \geq \mu_0$$

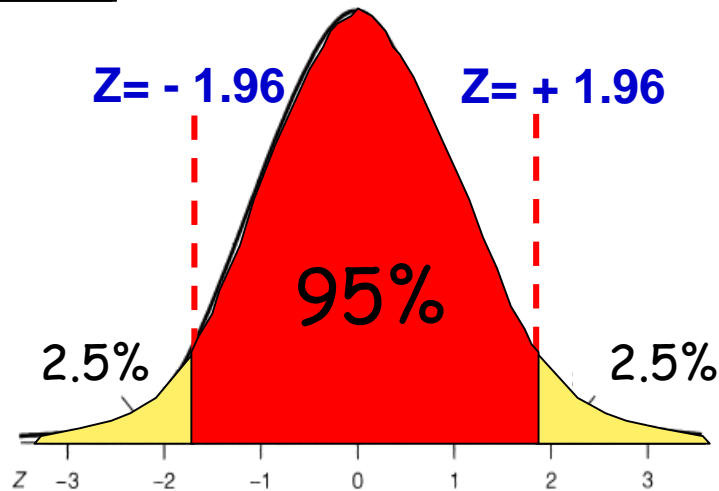
$$H_1 : \mu_1 < \mu_0$$



雙尾檢定

$$H_0 : \mu_1 = \mu_0$$

$$H_1 : \mu_1 \neq \mu_0$$



# Z分布與檢定

## 以一個樣本推論母全體

### 例 1 (雙尾檢定)

某群正常成年男子的血液平均收縮壓為125 mmHg( $\mu_0$ )，標準差為10mmHg( $\sigma$ )，今從同一年齡層的糖尿病患者中，隨機抽樣100人，其血液平均收縮壓為130mmHg( $\mu_1$ )，則這群糖尿病人血壓是否異於正常成年男子？

統計方法：單一樣本Z檢定，雙尾。

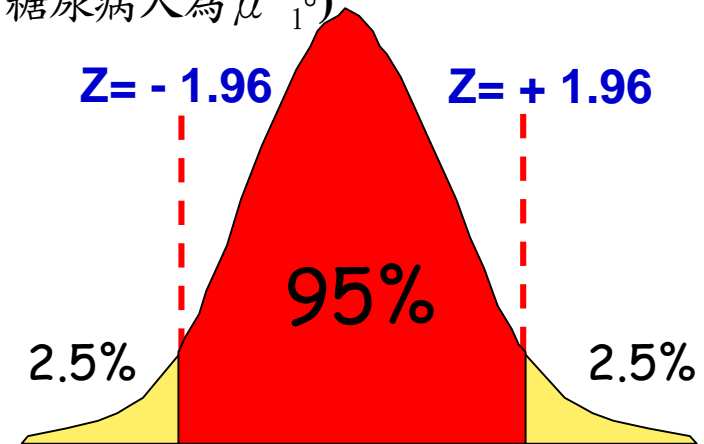
$H_0: \mu_0 = \mu_1$  (假設正常成年男子血壓 平均值為  $\mu_0$ ，糖尿病人為  $\mu_1$ )

$H_1: \mu_0 \neq \mu_1$

設  $\alpha = 0.05$

則  $Z_{0.975} = \pm 1.96$

$$Z = \frac{\bar{\chi} - \mu}{\frac{\sigma_{\bar{\chi}}}{\sqrt{100}}} = \frac{130 - 125}{\frac{10}{\sqrt{100}}} = 5$$



經查表，發現  $p < \alpha$ ，達統計顯著，推翻虛無假設  $H_0$ ，接受對立假設  $H_1$ ，故結論：糖尿病人之血壓與正常人血壓不同。



# Z分布與檢定

## 例 2 (單尾檢定)

假設美國成年男子平均身高**175公分**，標準差**20公分**。今從台灣成年男子隨機抽出**400人**，其平均身高**170公分**，請問台灣成年男子身高是否**低於**美國成年男子身高？

統計方法：單一樣本Z檢定，單尾。

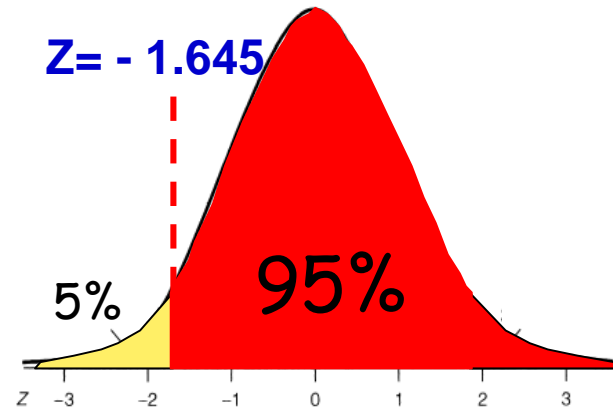
$H_0: \mu_0 \leq \mu_1$  (假設美國成年男子身高 平均值為  $\mu_0$ ，台灣男子為  $\mu_1$ )

$H_1: \mu_0 > \mu_1$

設  $\alpha = 0.05$

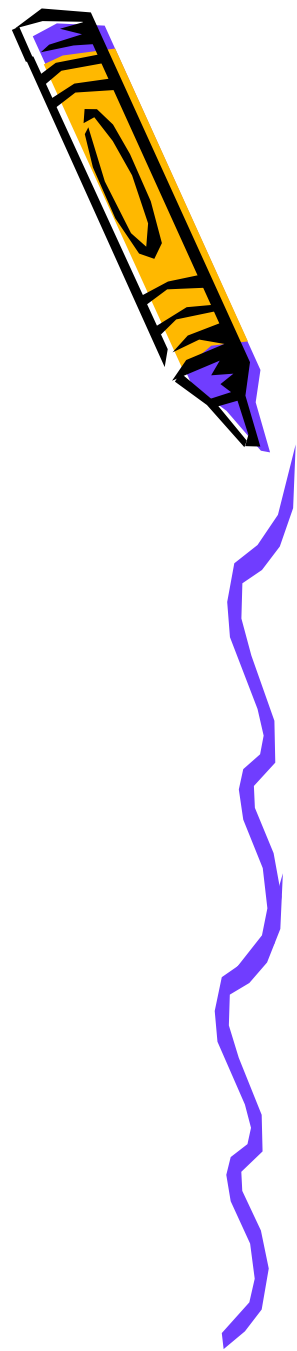
則  $Z_{0.95} = -1.645$

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma_x}{\sqrt{n}}} = \frac{170 - 175}{\frac{20}{\sqrt{400}}} = -5$$



經查表，發現  $p < \alpha$ ，達統計顯著，推翻虛無假設  $H_0$ ，接受對立假設  $H_1$ ，故結論：台灣男子身高低於美國身高。

# 選擇使用統計方法前先判別 變項(因子)的測量



## 變項(因子)種類

1. 類別或序位: 類別(非連續)
2. 等距或等比: 數字(連續)



# 變項間統計方法一覽表

1. 類別或序位
2. 等距或等比

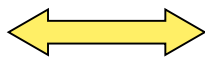
	類別序位)	等距(等比)
類別序位)	卡方檢定(ch9)	Z 檢定(ch6) $t$ 檢定(ch7) ANOVA(ch11)
等距(等比)	Z 檢定(ch6) $t$ 檢定(ch7) ANOVA(ch11)	迴歸分析(ch12) 相關分析(ch12)

# 統計學即



探討”因子”與”因子”間的關係。

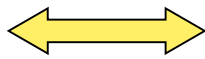
因子1



因子2



性別



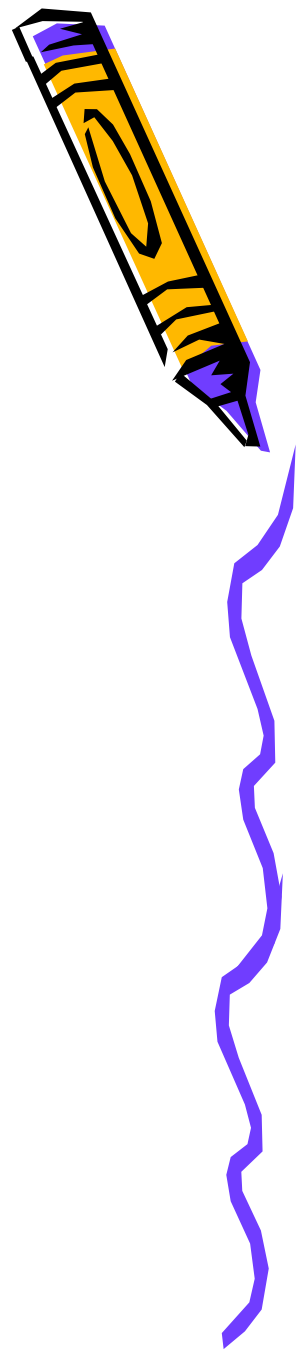
身高



# 資料型態與適用的統計方法

欲檢定的變項/ 依變項(Y)	欲進行比較的組別數/自變項(X)				
	類別				連續
	兩組		三組或以上		
	獨立樣本	相依樣本	獨立樣本	相依樣本	
連續資料	欲檢定的情形：集中趨勢 (central tendency)				相關分析
假設變項呈常態 或中央極限定理成立	Independent t-test	Paired t-test	ANOVA	Repeated measures ANOVA	Pearson's correlation / Linear regression
常態假設或中央 極限定理不成立	Wilcoxon rank-sum test	Wilcoxon signed-rank test	Kruskal-Wallis test	Friedman test	Spearman's correlation
類別資料	欲檢定的情形：關聯性(Association)				
兩個類別	Pearson's Chi-square test (with Yates' correction)	McNemar 's test	Pearson's Chi-square test	Cochran's Q test	<u>Logistic regression</u>
三類以上	Pearson's Chi-square test	Cochran's Q test	Pearson's Chi-square test	Cochran's Q test	Multinomial / Ordinal logistic regression

# t 分布與檢定

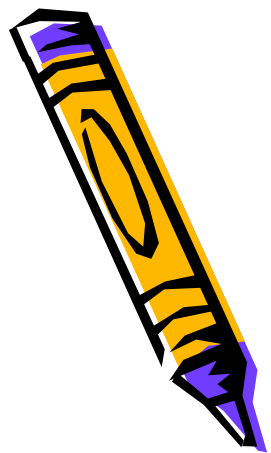


# t 分布與檢定

## t 分布的由來

英國化學家兼統計學家William S. Gossett在1908年發表，他在柏林的Guinness酒廠工作，因雇主擔心酒廠的貿易機密外露，鮮少允許雇員發表論文，所以Gossett以「學生」的假名發表，因此他發表的t分佈常又稱為學生氏t分佈 (Student's t distribution)。

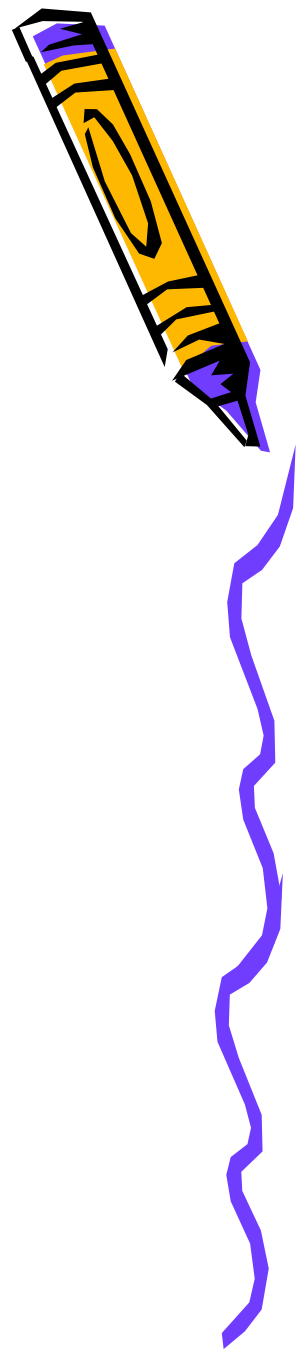
t分佈與標準常態分佈很相似，為單峰、鐘形、對稱，且兩邊無線延伸。但t分佈曲線之變異量稍高於於常態分佈，面積總合仍為1.0。



# t 分布與檢定

t 檢定必須符合的前提假設

1. 樣本必須是隨機選出的
2. 樣本分佈必須是常態分佈
3. 母全體變異數相等或**不相等**
4. 母全體**標準差未知**





# t 分布與檢定



## 點估計(point estimation)

若由母全體隨機抽出之樣本平均數為 $\bar{X}$ ，則 $\bar{X}$ 為母全體平均數 $\mu$ 的不偏估計(unbiased estimate)。同樣樣本數變異數 $S^2$ 為母全體變異數 $\sigma^2$ 的不偏估計。

t分布時，利用樣本標準差(S)預估母全體標準差( $\sigma$ )。



# t 分布與檢定

## 自由度(degree of freedom, df)的意義

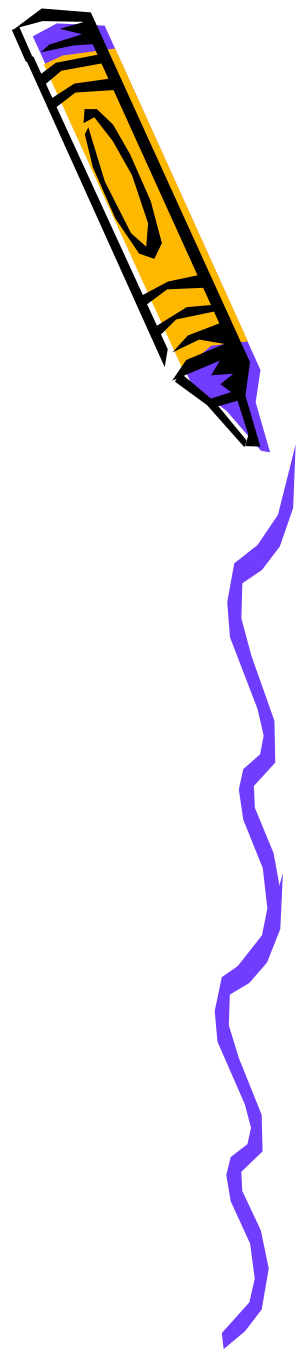
估算某一組資料中有多少個案資料可以用來估計母全體變異數 $\sigma^2$ 。

若資料為個案測量值  $\chi_i$

$$S^2 = \frac{\sum (\chi_i - \bar{\chi})^2}{n-1} \Rightarrow \sigma^2$$

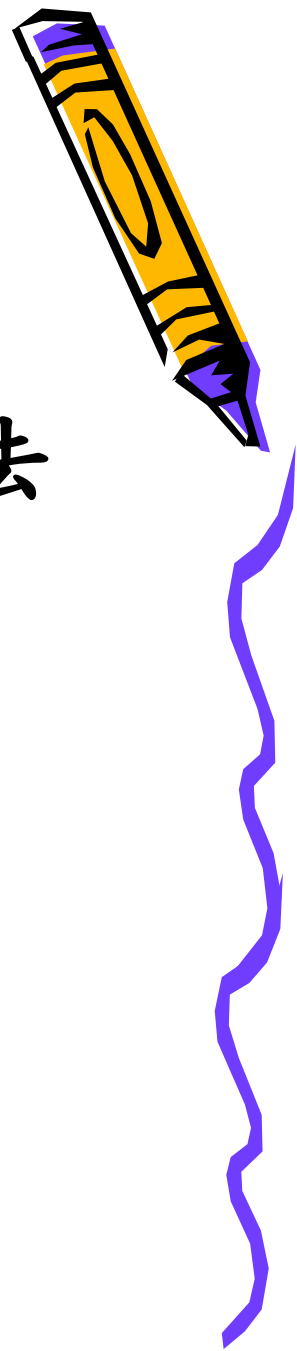


# 兩獨立樣本 t 檢定

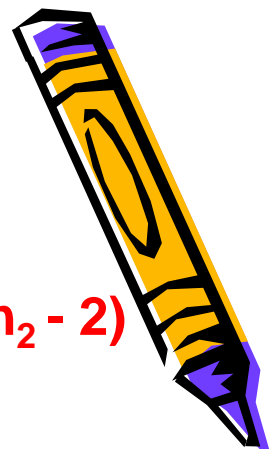


1. 點估計( $\bar{x}_1 - \bar{x}_2$ )統計檢定方法

2. 95% C.I. 檢定方法



# t 分布與檢定



以二個樣本的平均值推估二個母全體平均值 (自由度為  $n_1 + n_2 - 2$ )

統計檢定

P: pool

當  $n_1 = n_2$

$$S_p^2 = \frac{S_1^2 + S_2^2}{2}$$

則

$$S_p = \sqrt{\frac{S_1^2 + S_2^2}{2}}$$

當  $n_1 \neq n_2$

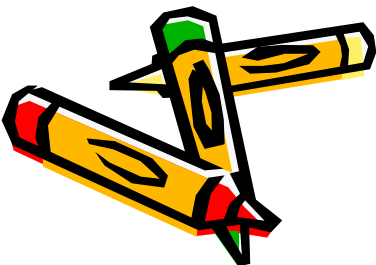
$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

則

$$S_p = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \mu_{(\bar{x}_1 - \bar{x}_2)}}{S_{(\bar{x}_1 - \bar{x}_2)}} = \frac{(\bar{x}_1 - \bar{x}_2) - \mu_{(\bar{x}_1 - \bar{x}_2)}}{S_p \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

→ ?



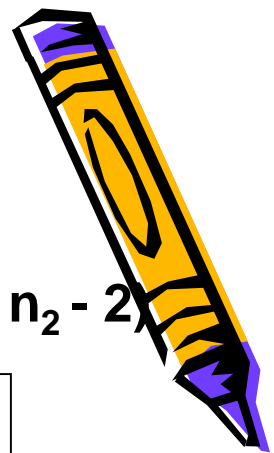


## 二個樣本推估母全體

信賴區間估計之統計檢定方法



# t 分布與檢定



以二個樣本的平均值推估二個母全體平均值 (自由度為  $n_1 + n_2 - 2$ )  
統計檢定

當  $n_1 = n_2$

$$S_p^2 = \frac{S_1^2 + S_2^2}{2}$$

則

$$S_p = \sqrt{\frac{S_1^2 + S_2^2}{2}}$$

當  $n_1 \neq n_2$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

則

$$S_p = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

$\mu_{(\bar{x}_1 - \bar{x}_2)}$  的 95% C.I.

$$95\% \text{ C.I. of } \mu_{(\bar{x}_1 - \bar{x}_2)} = (\bar{\chi}_1 - \bar{\chi}_2) \pm t_{0.975(\text{df})} \times S_{(\bar{\chi}_1 - \bar{\chi}_2)}$$

$$= (\bar{\chi}_1 - \bar{\chi}_2) \pm t_{0.975(\text{df})} \times S_p \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$



# t 分布與檢定

以二個樣本的平均值推估二個母全體平均值

點估計統計檢定

信賴區間估計之統計檢定

24名隨機抽樣之女生平均身高為154.8公分，標準差為4.7公分，另有 10名隨機抽樣獲得之男生，平均身高168.6公分，標準差為7.5公分，請問該校男、女生身高是否不同？男、女生身高差別的95%信賴區間是多少？

統計方法：兩獨立樣本t檢定，雙尾。

$$H_0: \mu_1 = \mu_2 = \mu_0$$

$$H_1: \mu_1 \neq \mu_2$$

$\mu_1$ : 該校抽樣女生之身高平均值

$\mu_2$ : 該校抽樣男生之身高平均值

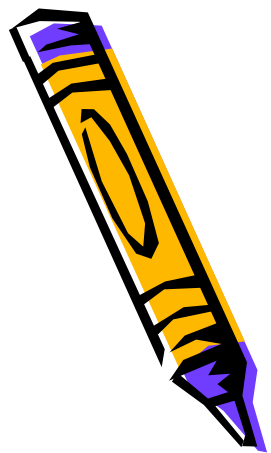
$\mu_0$ : 該校全體男女之身高平均值

$$n_1=24, \bar{X}_1=154.8, S_1=4.7$$

$$n_2=10, \bar{X}_2=168.6, S_2=7.5$$

$$\text{設 } \alpha=0.05$$

$$t_{0.975(32)} \approx \pm 2.04$$





# t 分布與檢定



(上續)  $\because n_1 \neq n_2$

$$S_p = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{(24 - 1)4.7^2 + (10 - 1)7.5^2}{24 + 10 - 2}} = 5.6$$

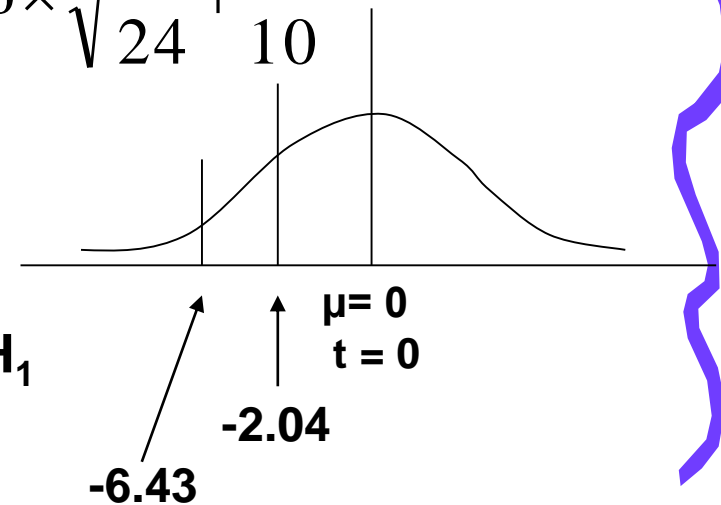
$$t = \frac{(\bar{\chi}_1 - \bar{\chi}_2) - \mu_{(\bar{\chi}_1 - \bar{\chi}_2)}}{S_{(\bar{\chi}_1 - \bar{\chi}_2)}} = \frac{(154.8 - 168.6) - 0}{5.6 \times \sqrt{\frac{1}{24} + \frac{1}{10}}} = -6.43$$

經查表  $t_{0.975(32)} \doteq \pm 2.04 < \pm 6.43$

$\therefore p < \alpha$ ，達統計顯著

推翻虛無假設  $H_0$ ，接受對立假設  $H_1$

結論：該校男、女生身高有差別



# t 分布與檢定



(上續) 計算身高差別之95%CI

設  $\alpha = 0.05$

$$t_{0.975(32)} = \pm 2.04$$

$\mu_{(\bar{x}_1 - \bar{x}_2)}$  的 95% C.I.

$$95\% \text{ C.I. of } \mu_{(\bar{x}_1 - \bar{x}_2)} = (\bar{\chi}_1 - \bar{\chi}_2) \pm t_{0.975} \times S_{(\bar{\chi}_1 - \bar{\chi}_2)}$$

$$= (\bar{\chi}_1 - \bar{\chi}_2) \pm t_{0.975(32)} \times S_p \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$= (154.8 - 168.6) \pm 2.04 \times 5.6 \times \sqrt{\frac{1}{24} + \frac{1}{10}}$$

$$= -18.077 \sim -9.373$$

∴ 身高差異之95%CI不包含 0

∴ 結論: 該校男、女生身高有差別



# t 分布與檢定

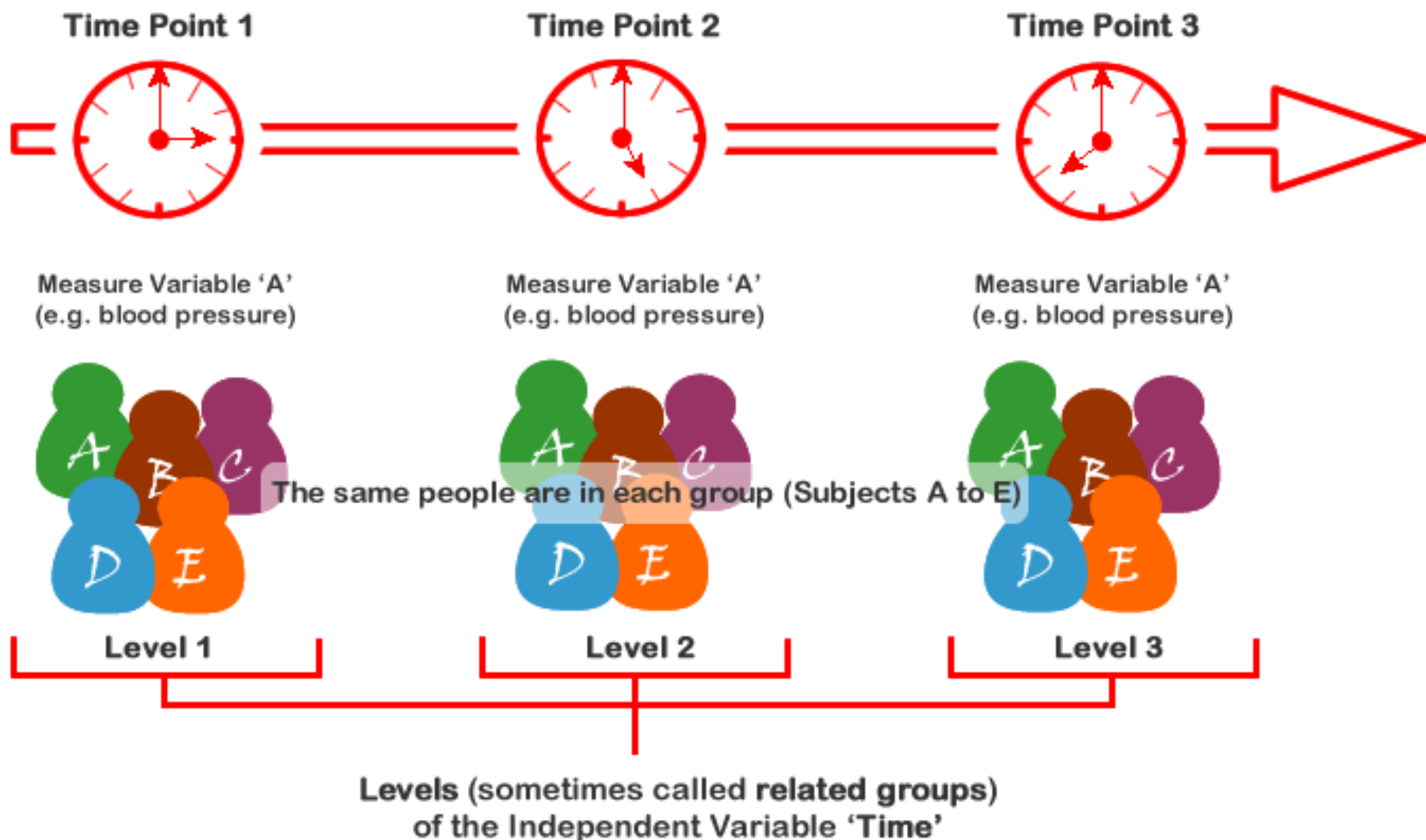
## 配對(相依) t 檢定

當兩組樣本非”各自獨立樣本“，而是”前” ”後“測量，目的可增加檢定的敏感度。

### 使用時機

1. 同一個體，不同部位測量。
2. 同一個體，前後測量。
3. 同卵雙胞胎，不同處理。
4. 同一胎老鼠，做隨機分派。





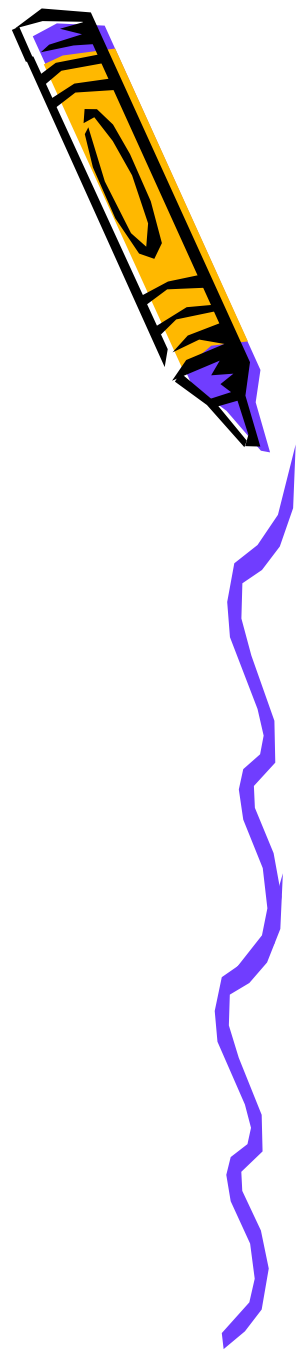
# t 分布與檢定

## 配對 t 檢定 公式

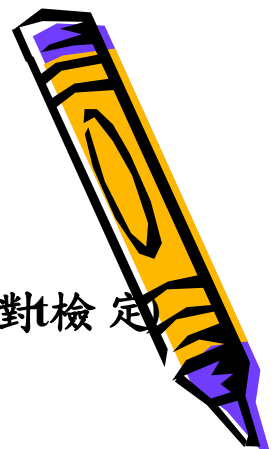
$$t = \frac{\bar{d} - \mu_{\bar{d}}}{S_{\bar{d}}} = \frac{\bar{d} - 0}{S_{\bar{d}}} = \frac{\bar{d}}{S_{\bar{d}}}$$

其中

$$\bar{d} = \frac{\sum_{i=1}^n (x_{\text{後測}} - x_{\text{前測}})}{n}$$



# t 分布與檢定



## 配對 t 檢定

例：

	治療前 (X <sub>1</sub> )	治療後 (X <sub>2</sub> )	前後差別 (d)
1	132	129	3
2	140	128	12
3	150	154	- 4
4	167	157	10
5	143	136	7
6	139	133	6
7	158	162	- 4
8	149	141	8
9	156	152	4
10	145	145	0
11	150	140	10

$$\bar{x}_1 = 148.091 \quad \bar{x}_2 = 143.364 \quad \bar{d} = 4.727$$

$$S_1 = 9.843 \quad S_2 = 11.595 \quad S_d = 5.515$$

統計方法兩相依樣本檢定(配對檢定)  
，雙尾。

$$H_0: \mu_d = 0$$

(服藥前後血壓值的數平均值為0)

$$H_1: \mu_d \neq 0$$

$$\text{設 } \alpha = 0.05$$

$$t_{0.975(10)} = 2.228$$

$$t = \frac{\bar{d} - \mu_d}{S_{\bar{d}}} = \frac{\bar{d} - 0}{S_{\bar{d}}} = \frac{\bar{d}}{S_{\bar{d}}} \\ = \frac{4.727 - 0}{\frac{5.515}{\sqrt{11}}} = 2.843 > 2.228$$

$\therefore p < \alpha$ ，達統計顯著，  
拒絕 $H_0$ ，接受 $H_1$

結論：血壓在服藥前後不同



# t 分布與檢定

配對 t 檢定

但

若視為兩獨立樣本 → 降低檢定的敏感性

