

SPSS 市場調查與研究

大數據分析

- R/Python/Julia/SQL 程式設計與應用
(R/Python/Julia/SQL Programming and Application)
- 資料視覺化 (Data Visualization)
- 機器學習 (Machine Learning)
- 統計品管 (Statistical Quality Control)
- 最佳化 (Optimization)



李明昌博士

alan9956@gmail.com

<http://rwepa.blogspot.com/>

大綱

Day 1

- 1. SPSS軟體簡介
 - 2. 敘述統計分析
- 補充篇：圖表建置器

Day 2

- 3. 卡方檢定應用
 - 4. 平均數檢定(t檢定)與變異
數分析
- 補充篇：R平均數檢定
補充篇：Python平均數檢定

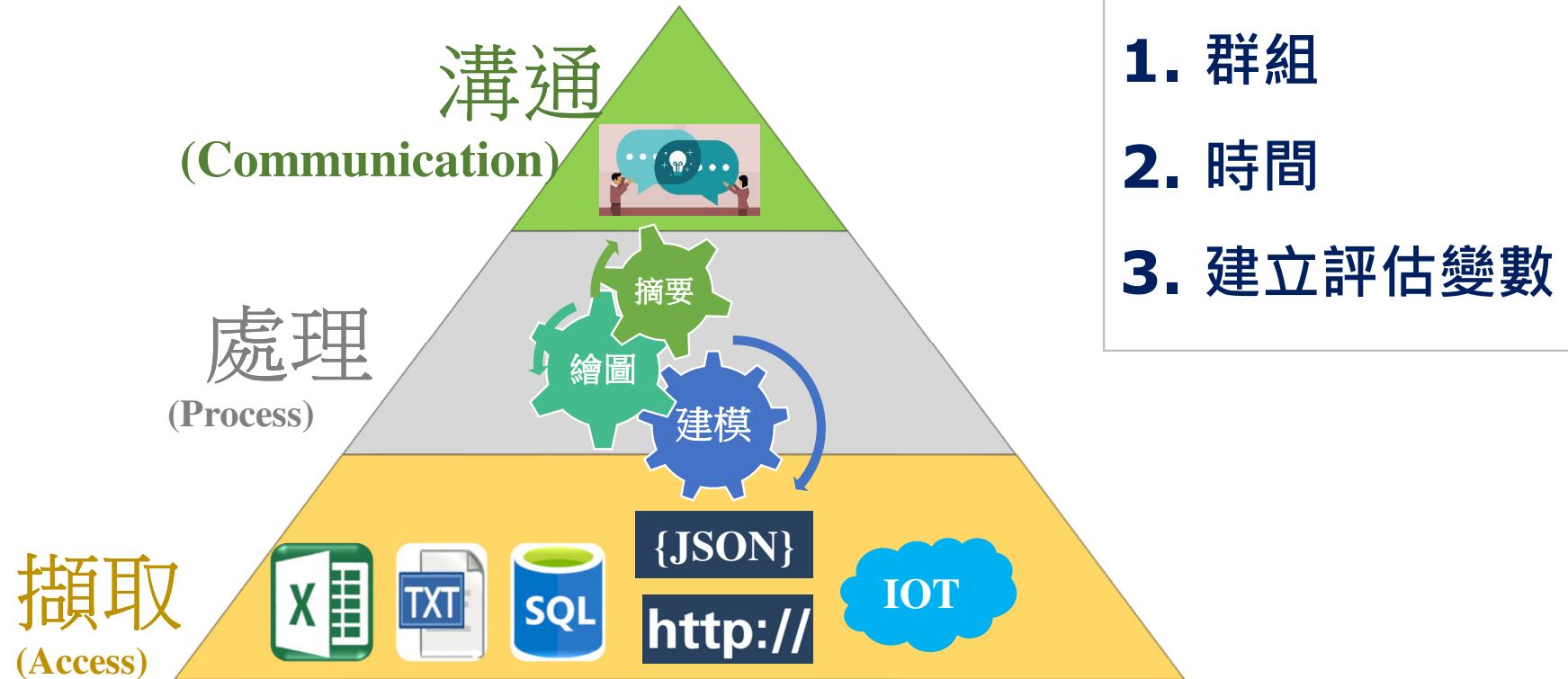
個人簡介 <http://rwepa.blogspot.com/>

- 姓名：李明昌 (ALAN LEE)
- 現職：中華R軟體學會 常務理事
臺灣資料科學與商業應用協會 常務理事
- 學歷：中原大學 工業與系統工程所 博士
- 經歷：
 - 淡江大學 兼任教師
 - 育達科技大學 兼任教師
 - 佛光大學 兼任教師
 - 國立台北商業大學 兼任教師
 - 東吳大學 兼任教師
 - 育達科技大學 資訊管理系(所) 專任助理教授
 - 崇友實業 行銷企劃專員
 - 國航船務代理股份有限公司 海運市場運籌管理員
- 大專院校、資策會、工業技術研究院、國家發展委員會、中央氣象局、公平交易委員會、各縣市政府與日本名古屋產業大學等公營單位演講達300餘場，2900小時以上。
- 連絡資訊：*alan9956@gmail.com*



- iPAS 巨量資料分析師 證照推廣
- iPAS 營運智慧分析師 證照推廣

★★★資料分析架構→APC方法



Day 2

3. 卡方檢定應用

交叉分析

交叉分析

- 單一個類別變數可以使用次數分配表理解資料的分佈情形。
- 進行市場調查或問卷調查，可以使用**交叉分析表**來探討**兩個或多個類別變數**之關聯性。
- 交叉分析表（Contingency table, Cross-tabulation, Crosstab, Two-way table）。
 - 地區別與某政策之贊成與否
 - 性別與政黨偏好
 - 教育程度與使用品牌
 - 品牌與購買原因
 - 所得與是否有購買數位相機

- 檔案：星巴克.sav
- 資料筆數：200
- 變數個數：5
- 次數（每週購買次數）：
 - 1: ~3次
 - 2: 4~9次
 - 3: 10次~。
- 零用金值標籤：
 - 1: ~5000
 - 2: 5001~10000
 - 3: 10001~
- 性別：
 - 1: 男
 - 2: 女
- 目標：建立「次數」對「零用金」之交叉分析表。

	編號	次數	零用金	性別	年齡
1	1	~3次	5001~10...	女	21~30
2	2	4~9次	5001~10...	女	21~30
3	3	4~9次	5001~10...	男	~20
4	4	~3次	~5000	女	21~30
5	5	~3次	5001~10...	男	21~30

- 步驟1. 執行「分析(A)/敘述統計(E)/交叉資料表(C)...」



- 步驟2. 選「次數」，按 鈕，將其送到右側之「列(O)」方塊。
- 步驟3. 選「零用金」，按 鈕，將其送到右側之「欄(C)」方塊。



- 步驟4.按 **資料格(E)...** 鈕，設定顯示內容：

- 計數：觀察值(O)
- 百分比：欄(C)



- 步驟5. 按〔繼續〕鈕，回步驟3對話方塊。
- 步驟6. 按〔確定〕鈕。
- 結論：
 - 受訪者中，每月到星巴克之次數以『~3次』者最多（72.5%）。
 - 其次依序為每月『4~9次』（20.0%）與每月『10次~』（7.5%）。

次數*零用金 交叉列表

		零用金				總計
次數	~3次	~5000	5001~10000	10001~		
		計數 零用金內的 %	50 83.3%	70 75.3%	25 53.2%	145 72.5%
4~9次	計數 零用金內的 %	7 11.7%	17 18.3%	16 34.0%	40 20.0%	40
	計數 零用金內的 %	3 5.0%	6 6.5%	6 12.8%	15 7.5%	
總計	計數 零用金內的 %	60 100.0%	93 100.0%	47 100.0%	200 100.0%	200

- 結論：

- 經由與零用金交叉分析後，可發現受訪者每月到星巴克之次數與零用金高低應存有明顯關係。
- 零用金愈低每月到星巴克之次數愈低；隨零用金逐步增加，每月到星巴克之次數則同步增加。
- 每月到星巴克未滿3次者，主要集中於較低零用金（~5000）顧客（83.33%）。
- 每月到星巴克4~9次及10次以上者，均以高零用金者（10001~）顧客的為主（34.04%與 12.77%）。

卡方檢定 - 簡介

卡方檢定

- 卡方檢定(Chi-Squared Test 或 χ^2 Test)主要是用於屬性資料(類別資料)之資料分析。
- 比較各類別抽樣所得到之「觀察次數」(observed frequency, O_i) 與虛無假設 (H_0)成立之條件下之「期望次數」(expected frequency, E_i) 之差異，計算檢定之統計量：

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

- 卡方檢定應用有三種情形：
 1. 適合度檢定：檢定資料是否來自於某一特定之機率分配。【Demo】
 2. 獨立性檢定：檢定兩個屬性是否有相關性或互為獨立。【Demo】
 3. 齊一性檢定：檢定兩個或兩個以上不同的母體是否具有相同分配或相同比例。

參考: 統計學：觀念、方法、應用 3/e , 前程出版, 2008. [[LINK](#)]

適合度檢定

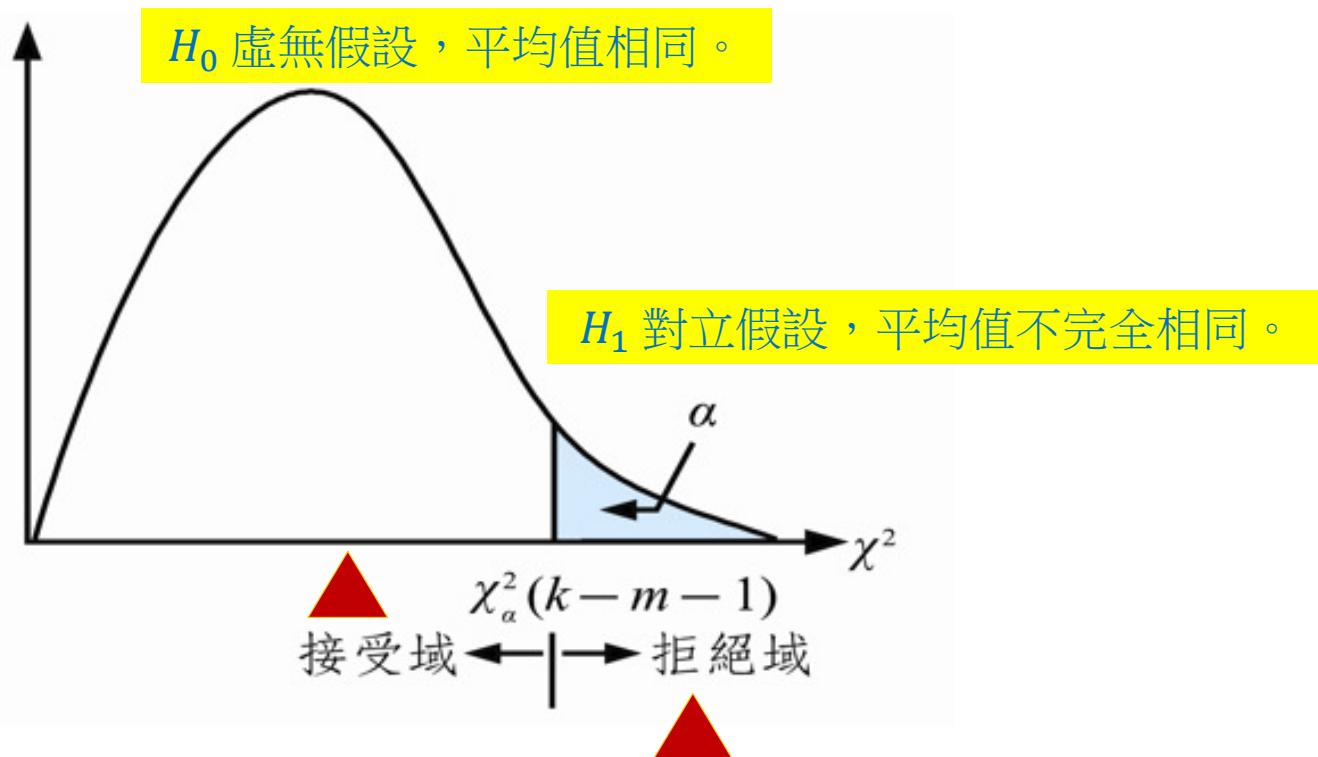
- 適合度之概念：檢定某一資料之次數分配或相對次數分配是否符合一個特定之機率分配。
- 定理：若 O_i 、 E_i 分別表適合度檢定之觀察次數與期望次數，且 k 為資料之類別個數， m 為需要估計母體之參數個數，則當 $n \rightarrow \infty$ 時，在 H_0 成立條件下：

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \sim \chi^2(k - m - 1) \text{，其中 } E_i = n \times p_i.$$

- 母體是否具有一致分配。例：公平骰子，則 $m = 0$ 。
- 母體是否具有常態分配，當母體平均數 μ 及母體變異數 σ 均為已知時，則 $m=0$ 。
- 母體是否具有常態分配，當母體平均數 μ 及母體變異數 σ 均為未知時，則 $m=2$ 。

適合度檢定(續)

- 當決定顯著水準為 α 時，則其拒絕域為 $\chi^2_\alpha \geq \chi^2_\alpha(k - m - 1), m = 0$
- 如下圖所示，卡方檢定為右尾檢定。



適合度檢定 - 範例

適合度檢定 - 範例

某公司想瞭解消費者對其所生產三種不同口味的產品之喜好程度是否一致，於是隨機訪問此產品消費者120人，得其結果如下：

產品類別	1	2	3
喜好人數	35	42	43

請以 $\alpha = 0.05$ 檢定該產品消費者對三種產品之喜好程度是否有顯著地差異？

適合度檢定 - 範例 (續)

【解】

- 令 p_i 表示第 i 種產品受喜好的比例
- 假設 $H_0: p_1 = p_2 = p_3 = \frac{1}{3}$, $H_1: p_1$ 、 p_2 、 p_3 不完全相等
- 拒絕域 $\{\chi_0^2 \geq \chi_{0.05}^2(3 - 1) = \chi_{0.05}^2(2) = 5.991\}$
- 在 H_0 成立條件下， $E_1 = E_2 = E_3 = 120 \times \frac{1}{3} = 40$
$$\chi_0^2 = \sum_{i=1}^3 \frac{(O_i - E_i)^2}{E_i} = \frac{(35-40)^2 + (42-40)^2 + (43-40)^2}{40} = 0.95 < 5.9915$$
，計算統計值 0.59 小於查表 5.9915 結果。
- 結論：接受 H_0 ，即三種不同口味之產品受消費者喜好程度無顯著地差異。

卡方分配表

- 下載 <https://github.com/rwepa/DataDemo/blob/master/prob.pdf>
- 參考 pdf P.19
 - 題目 $\alpha = 0.05$
 - 橫列自由度 $v = 2$
 - 直行 $\alpha = 0.05$
 - 二者合併為 5.9915

卡方分配之臨界值(續)										
v 表自由度(<i>df.</i>)	α									
	0.3	0.25	0.2	0.1	0.05	0.025	0.02	0.01	0.005	0.001
1	1.0742	1.3233	1.6424	2.7055	3.8415	5.0239	5.4119	6.6349	7.8794	10.8274
2	2.4079	2.7726	3.2189	4.6052	5.9915	7.3778	7.8241	9.2104	10.5965	13.8150
3	3.6649	4.1083	4.6416	6.2514	7.8147	9.3484	9.8374	11.3449	12.8381	16.2660
4	4.8784	5.3853	5.9886	7.7794	9.4877	11.1433	11.6678	13.2767	14.8602	18.4662
5	6.0644	6.6257	7.2893	9.2363	11.0705	12.8325	13.3882	15.0863	16.7496	20.5147
6	7.2311	7.8408	8.5581	10.6446	12.5916	14.4494	15.0332	16.8119	18.5475	22.4575

獨立性檢定 - SPSS

H_0 : 兩個變項間無關, 即二個變數是獨立.

H_1 : 二個變數是相關的, 兩個變數間有顯著差異.

- 檔案：星巴克.sav
- SPSS\分析\敘述統計\交叉資料表\統計資料：「卡方檢定」打勾



次數*零用金 交叉列表

次數*零用金 交叉列表

		零用金			零用金(行數=3)
		~5000	5001~10000	10001~	總計
次數	~3次	計數	50	70	25
		零用金 內的 %	83.3%	75.3%	53.2%
4~9次	計數	7	17	16	40
	零用金 內的 %	11.7%	18.3%	34.0%	20.0%
10次~	計數	3	6	6	15
	零用金 內的 %	5.0%	6.5%	12.8%	7.5%
總計	計數	60	93	47	200
	零用金 內的 %	100.0%	100.0%	100.0%	100.0%

次數(列數=3)

- 卡方值為12.716，自由度為4，顯著性（雙尾）為 0.013。
- 判斷檢定結果：判斷顯著性（ p 值， p value）是否小於所指定顯著水準之 α 值，本例 $\alpha=0.05$ 。
 - 如果 p 值 $\leq \alpha$ ，即接受 H_1 ，即兩個變數間有顯著差異，即二個變數是相關的。
 - 如果 p 值 $> \alpha$ ，即接受 H_0 ，兩個變項間無關，即二個變數是獨立。
- 本例，由於其 p 值 $0.013 \leq \alpha$ ，故應捨棄購買次數與零用金無關之虛無假設，而接受 H_1 。即每月到星巴克之次數，會因其零用金不同而有顯著差異。

卡方檢定

	值	df	漸近顯著性 (兩端)
Pearson 卡方檢定	12.716 ^a	4	.013
概似比	12.248	4	.016
線性對線性關聯	9.490	1	.002
有效觀察值個數	200		

a. 2 單元 (22.2%) 預期計數小於 5。預期的計數下限為 3.53。

自由度之計算：
 $df = (\text{列數} - 1) \times (\text{行數} - 1)$
 $= (3 - 1) \times (3 - 1) = 4$

卡方檢定注意事項

- 卡方檢定僅適用於類別資料（名目變數），如：性別、地區、政黨傾向、宗教信仰、是否有手機。
- 各儲存格之期望次數不應少於5。通常要有80%以上的儲存格期望次數 ≥ 5 ，否則會影響其卡方檢定的效果。若期望次數小於5時，可將其合併計算。

4. 平均數檢定與變異數分析

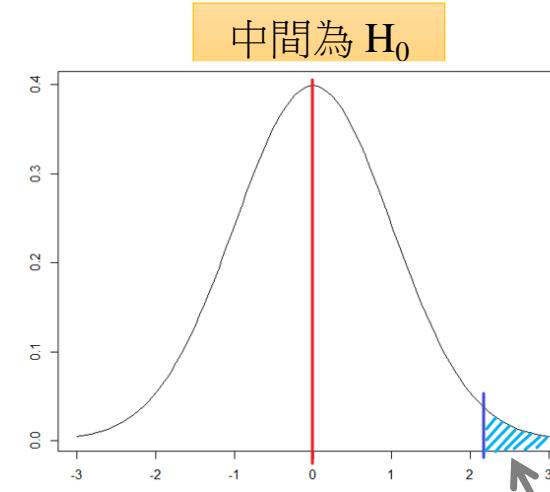
平均數檢定 - 簡介

單一樣本 t 檢定(小樣本)

$H_0 : \mu = \mu_0$ 虛無假設 null hypothesis

$H_1 : \mu \neq \mu_0$ 對立假設 alternative hypothesis

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n - 1),$$
$$\bar{x} \rightarrow \bar{X}, s \rightarrow S$$



左右二側為 H_1

假設檢定

- 假設檢定的意義：首先對母體特性作一假設描述，然後利用抽樣得到的樣本資料，來驗證是否支持此假設。因此假設檢定即為利用所抽出的樣本，對母體的特性作推論。
- 建立假設：
 - 虛無假設 (null hypothesis)：對母體的未知參數所作之假設，以 H_0 表示。
 - 對立假設 (alternative hypothesis)：與虛無假設相反之假設，以 H_1 表示。

誤差

- 假設檢定的兩種誤差
 - 型 I 誤差：當 H_0 為真，檢定結果為拒絕 H_0 ，所犯的錯誤
 - 型 II 誤差：當 H_0 為誤，檢定結果為接受 H_0 ，所犯的錯誤

		真實狀況	
		H_0 真	H_0 假
檢定結果	接受 H_0	正確的決策	犯型 II 誤差
	拒絕 H_0	犯型 I 誤差	正確的決策

- 發生型 I 誤差之最大機率，稱為顯著水準 (significance level) ，記為 α 。

- 假設檢定的三種類型

- 左尾檢定 :
$$\begin{cases} H_0 : \theta \geq \theta_0 \\ H_1 : \theta < \theta_0 \end{cases}$$

- 右尾檢定 :
$$\begin{cases} H_0 : \theta \leq \theta_0 \\ H_1 : \theta > \theta_0 \end{cases}$$

- 雙尾檢定 :
$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases}$$

θ : 表示欲檢定之母體參數，例:平均值

θ_0 : 表示特定參數數值

檢定方法

- 臨界值檢定 (critical value test)
- Z 值檢定(常態分配，樣本數 ≥ 30)或 t 值檢定(t 分配，樣本數 < 30)
- 信賴區間檢定
- p 值檢定

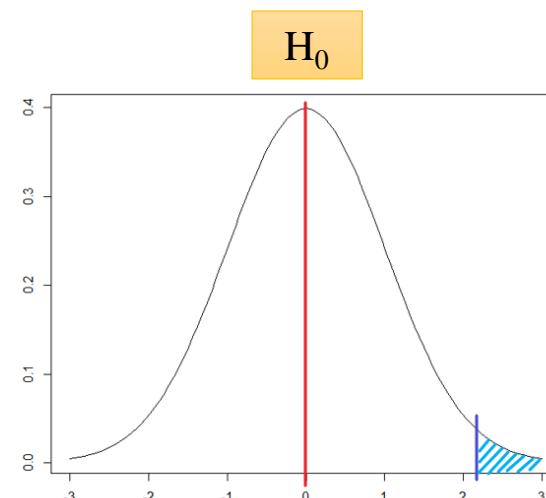
上述四種方法雖不同，但檢定出來的結果會一致。

p 值檢定法

- 近年來，電腦軟硬體快速發展，許多統計電腦軟體都會在輸入樣本資料後，*p* 值可以被快速計算，並列印在報表中。
- *p* 值為在 H_0 為真的條件下，以樣本觀察值作為新的拒絕域邊界，所計算出來的機率值。
- *p* 值即衡量樣本觀察值與 H_0 相符的程度：
 - *p* 值越小表示拒絕 H_0 的證據越強烈。

p 值檢定之步驟 (以單尾為例)

1. 建立虛無假設
2. 抽取一組隨機樣本並計算欲檢定之統計量
3. 在 H_0 為真的條件下，計算 p 值
4. 將 p 值與顯著水準 α 作比較
 1. 若 p 值 $\leq \alpha$ ，則拒絕 H_0
 2. 若 p 值 $> \alpha$ ，則接受 H_0
5. 根據檢定結果下決策



H_1

平均數檢定 - 手動計算

雙尾檢定

- 考量全班隨機抽取40位學生之成績：

75	85	78	70	80	80	54	78
85	88	85	85	80	85	56	82
25	85	78	75	78	82	47	83
60	80	78	70	82	78	60	75
80	88	78	83	90	90	49	82

- 樣本數 $n = 40$ 為大樣本，使用Z值檢定。
- $\alpha=0.05$ 之顯著水準，是否可接受全班成績為70分之假設？

平均值檢定

- 步驟1.設定虛無假設 $H_0: \mu = 70$ ，設定對立假設 $H_1: \mu \neq 70$ ，本例為雙尾檢定。
- 步驟2.決定顯著水準 (α) ， $\alpha=0.05$ 。
- 步驟3.選擇適當的檢定統計資料，並決定拒絕域：由於是大樣本，以Z檢定統計資料，採雙尾檢定，應查詢 $\alpha=0.025$ 之表格。
- 步驟4.查詢『標準常態分配表』，累積機率為0.475時，其拒絕域之臨界點為1.96。

如果 $\{Z\text{檢定統計值} \leq -1.96 \text{ 或 } Z\text{檢定統計} \geq 1.96\}$ ，判定為捨棄虛無假設(H_0)而接受對立假設(H_1)。

標準常態分配表

- 下載 <https://github.com/rwepa/DataDemo/blob/master/prob.pdf>
- 參考 pdf P.16

- 題目 $\alpha = 0.05$
- 雙尾檢定 $\frac{0.05}{2} = 0.025$
- 0.025 對照橫列為 1.9
- 0.025 對照直行為 0.06
- 二者合併為 1.96

<i>z</i>	標準常態分配之累積機率值(續)									
	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817

平均值檢定 (續)

- 步驟5. 計算所選之觀察值，計算樣本平均數 ($\bar{x} = 75.55$) 與樣本標準差 ($s = 13.54$) 及已知之 $\mu=70$ ，代入Z檢定統計資料之公式：

$$Z = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} = \frac{75.55 - 70}{\frac{13.54}{\sqrt{40}}} = 2.592$$

- 步驟6. 結論：檢定統計資料 $Z=2.592 \geq 1.96$ 之臨界值，已經位於拒絕域，故捨棄虛無假設 $H_0 : \mu = 70$ 。也就是應接受其對立假設： $H_1 : \mu \neq 70$ 。結論：**無法接受全班成績為70分之假設。**

平均數檢定 - SPSS

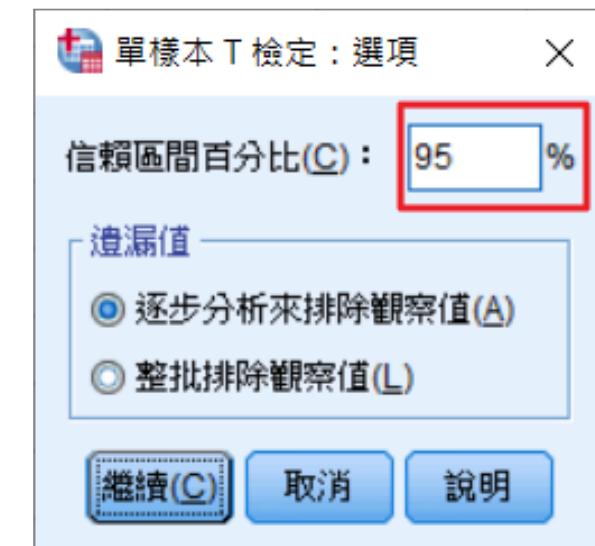
SPSS - 平均值檢定

- 檔名：單一母體平均數檢定.sav
- 分析(A) / 比較平均數(M) / 單一樣本T檢定(S)...



SPSS - 平均值檢定 (續)

- 檢定變數：成績
- 檢定值：70



SPSS - 平均值檢定 (續)

單一樣本統計量

	N	平均值	標準差	標準誤平均值
成績	40	75.55	13.540	2.141

單一樣本檢定

檢定值 = 70

	t	自由度	顯著性 (雙尾)	平均值差異	差異的 95% 信賴區間
成績	2.592	39	.013	5.550	1.22 9.88

SPSS - 平均值檢定 (續)

- 平均數與標準差分別為75.55與13.54。
- 檢定結果之t值2.592，其結果與先前計算之Z值相同。
- 判斷檢定結果於雙尾檢定只須看此『顯著性（雙尾）』是否小於所指定之 α 值；於單尾檢定則看此『顯著性（雙尾）』除以2是否小於所指定之 α 值。
- 本例是雙尾檢定，『顯著性（雙尾）』為 $0.013 < \alpha = 0.05$ ，即表示在 $\alpha = 0.05$ 時，此檢定結果要捨棄虛無假設 H_0 ，接受對立假設 H_1 。
- 結論：捨棄 $H_0 : \mu = 70$ ，即接受 $H_1 : \mu \neq 70$ 。所以，無法接受全班成績為70分之假設。

變異數分析 - SPSS

變異數分析 (Analysis of variance, ANOVA)

- 變異數分析是用來檢定多組 (>2) 之平均數是否相等？亦即，Z 與t檢定是用於兩組資料比較平均數差異時；而比較二組以上的平均數是否相等時，就須使用變異數分析。
- 其虛無假設與對立假設為：
 - $H_0: \mu_1 = \mu_2 = \dots = \mu_k$ (每組之平均數相等)
 - H_1 : 至少有兩個平均數不相等

變異數分析 - SPSS

- 目標：某大飯店之餐廳於 facebook 上進行廣告，不同廣告方式當日所獲得之回應人數如下表所示，考量 $\alpha = 0.05$ 之顯著水準，檢定不同廣告方式之回應人數是否存有顯著差異？

全版	半版	1/4 版	小廣告
1250	1083	850	660
1324	1400	755	605
1600	1385	623	580
890	680	600	856
926	868	701	964
1051		782	
		760	

行銷結果輸入到 SPSS

- 檔案：廣告效果.sav

全版	半版	1/4 版	小廣告
1250	1083	850	660
1324	1400	755	605
1600	1385	623	580
890	680	600	856
926	868	701	964
1051		782	
		760	



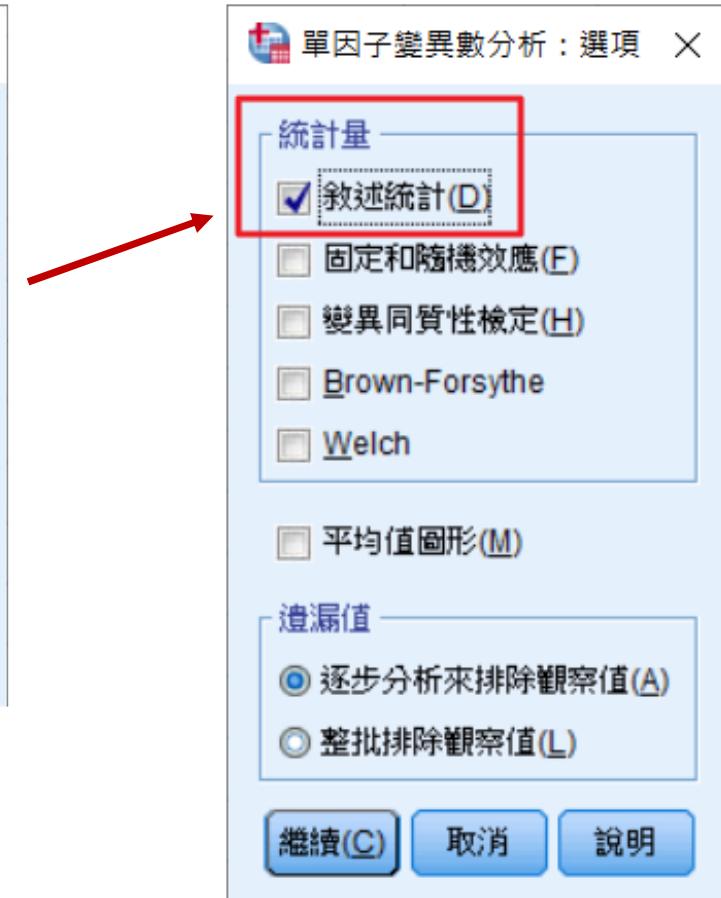
	廣告類別	回應人數
1	全版	1250
2	全版	1324
3	全版	1600
4	全版	890
5	全版	926
6	全版	1051
7	半版	1083
8	半版	1400
9	半版	1385
10	半版	680
11	半版	868
12	1/4 版	850
13	1/4 版	755
14	1/4 版	623
15	1/4 版	600
16	1/4 版	701
17	1/4 版	782
18	1/4 版	760
19	小廣告	660
20	小廣告	605
21	小廣告	580
22	小廣告	856
23	小廣告	964

單因數變異數分析

- SPSS \ 分析(A) \ 比較平均數(M) \ 單因數變異數分析(O)...



單因子變異數分析 (續)



單因數變異數分析 (續)

- 結果： $F=6.556$ ， $df=3,19$ ，顯著性為0.03小於 $\alpha=0.05$ 。
- 不同方式廣告之回應人數間存有顯著差異。全版與半版廣告之回應人數（1173.5與1083.2）高於1/4版與小廣告（724.4與733.0）。
- 實際分析時，最簡單方式為判斷『顯著性』是否小於所指定之 α 值即可。
- 若『顯著性』 $< \alpha$ ，即應捨棄每組均數相等之虛無假設。

敘述統計							
回應人數	N	平均值的 95% 信賴區間					
		平均值	標準差	標準誤	下限	上限	最小值
全版	6	1173.50	270.939	110.610	889.17	1457.83	890
半版	5	1083.20	316.357	141.479	690.39	1476.01	680
1/4版	7	724.43	89.071	33.666	642.05	806.81	600
小廣告	5	733.00	168.532	75.370	523.74	942.26	580
總計	23	921.43	293.092	61.114	794.69	1048.18	580
							1600

變異數分析					
回應人數	平方和	自由度	均方	F	顯著性
					.003
群組之間	961279.638	3	320426.546	6.556	
群組內	928580.014	19	48872.632		
總計	1889859.652	22			

補充篇：R平均數檢定

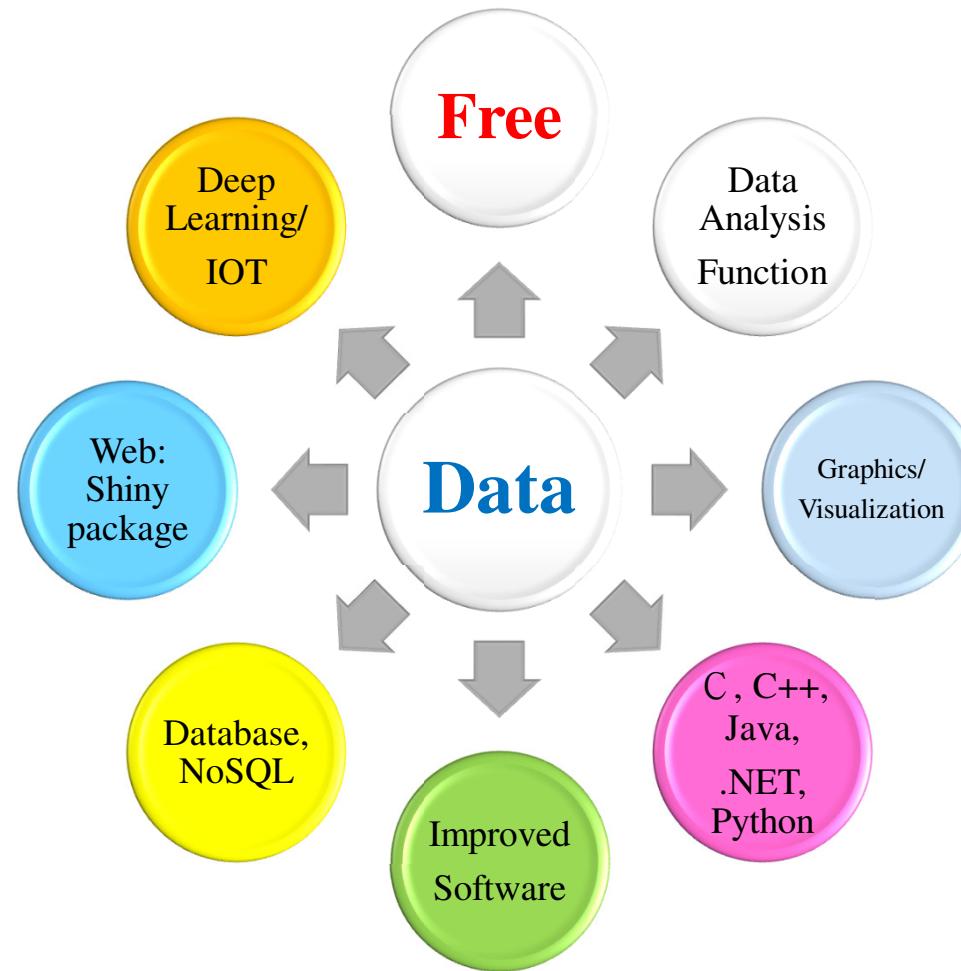
R 簡介

認識R

- 1976 - 貝爾實驗室 John Chambers, Rick Becker, and Allan Wilks 研發S語言。
- 1993 - Ross Ihaka and Robert Gentleman, University of Auckland, New Zealand 將S語言研發為R 語言。
 - R 是一種基於 S 語言所發展出具備統計分析、繪圖與資料視覺化的程式語言。
 - R 提供現代資料科學免費實務應用工具。
- 1997年－R的核心開發團隊 (R development core team) 成立，專責R原始碼的修改與編寫。
 - 2000年2月 – R 1.0.0
 - 2013年3月 – R 2.15.3
 - 2022年10月 – R 4.2.2



R-八大功能



R 下載

R官方網頁

[\[Home\]](#)[Download](#)[CRAN](#)**下載****繪圖**[R Project](#)[About R](#)[Logo](#)[Contributors](#)[What's New?](#)

The R Project for Statistical Computing

Getting Started

統計計算

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred CRAN mirror.

If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

R-下載

- 官網: <http://www.r-project.org/>
- 選取左側 Download \ CRAN
- 選取 Taiwan CRAN

Taiwan
<https://cran.csie.ntu.edu.tw/>

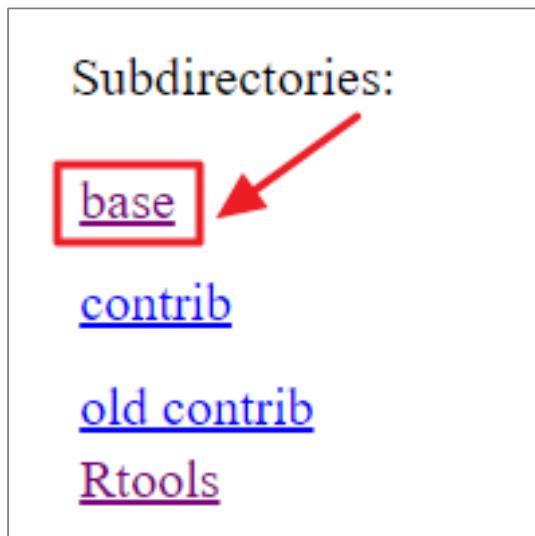
- 選取 Download R for Windows

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)



R-下載 (續)

- 選取 base → 下載 [R-4.2.2-win.exe]

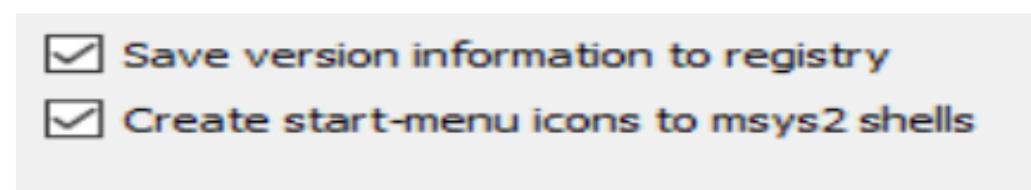
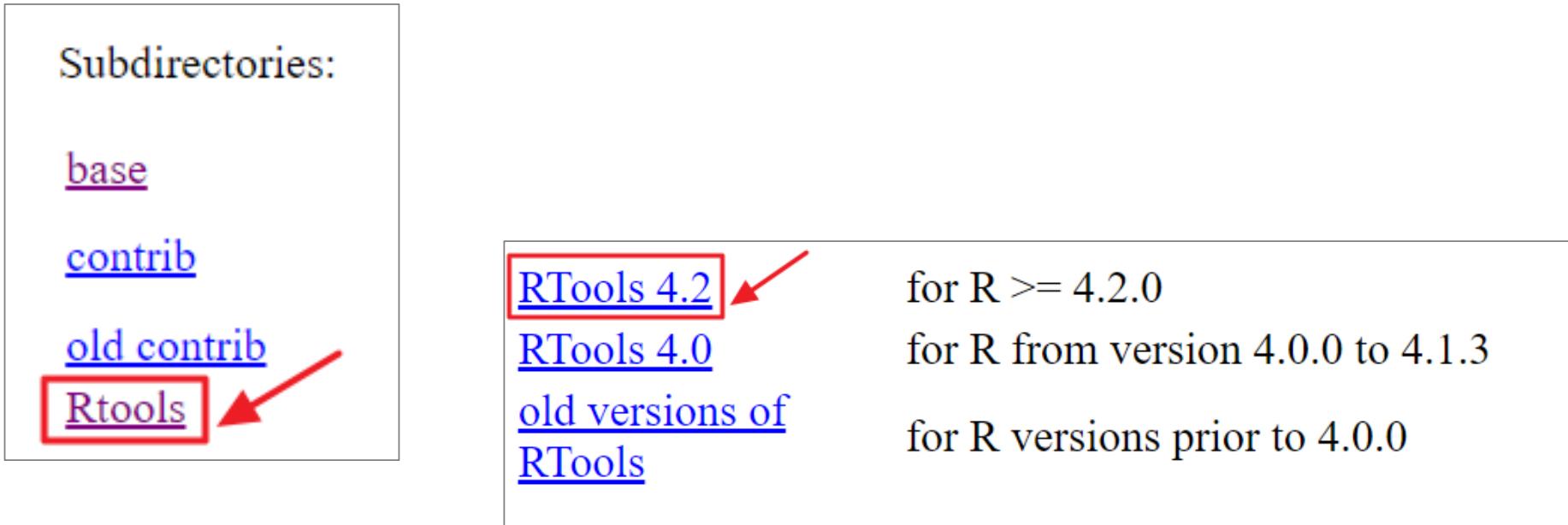


[Download R-4.2.2 for Windows](#) (76 megabytes, 64 bit)
[README on the Windows binary distribution](#)
[New features in this version](#)

- R安裝路徑: 保留原路徑,不要修改
- 安裝參考說明, 2006
https://github.com/rwepa/DataDemo/blob/master/windows_intall_R.pdf

Rtools 下載與安裝

- Rtools for Windows: 一定要保留原路徑 **C:\rtools42**





安裝R



安裝 Rtools

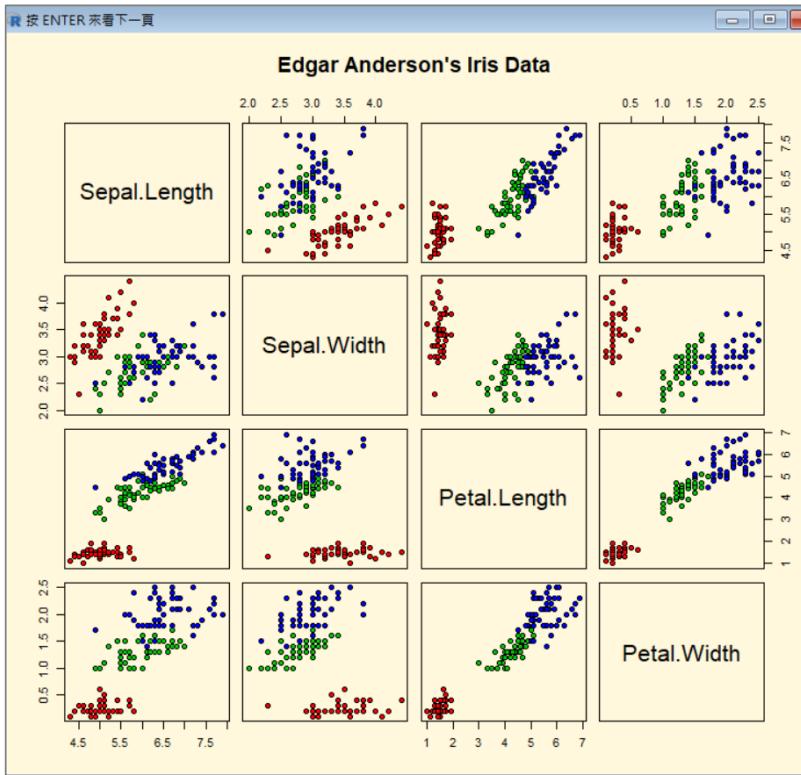
安裝R, Windows 登入名稱:

1. 不要使用空格
2. 不要使用中文字型

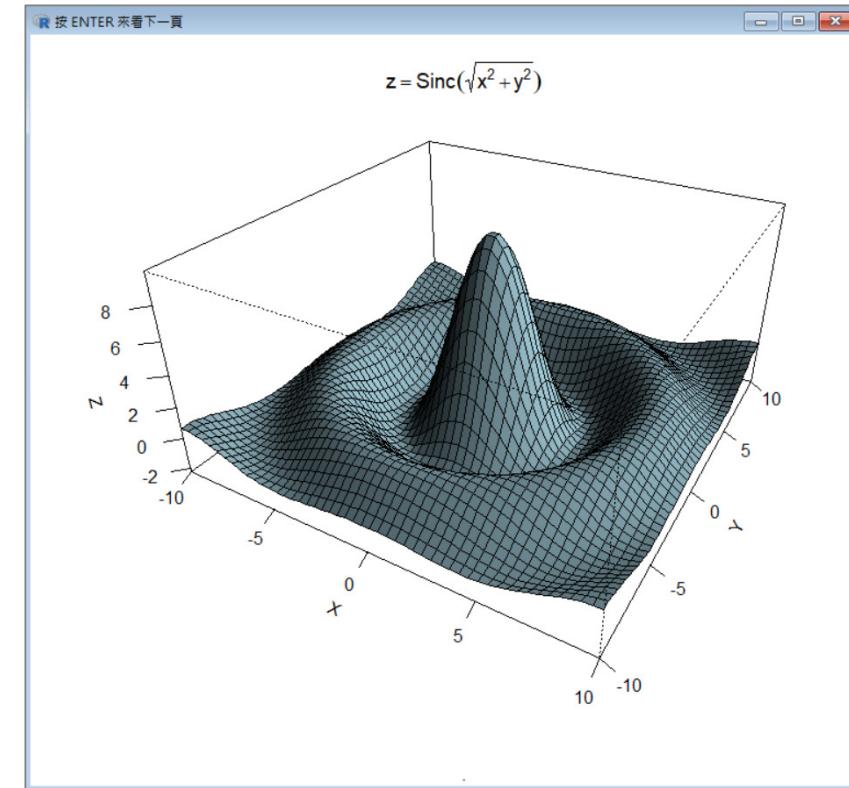


R 視覺化

- demo(graphics)

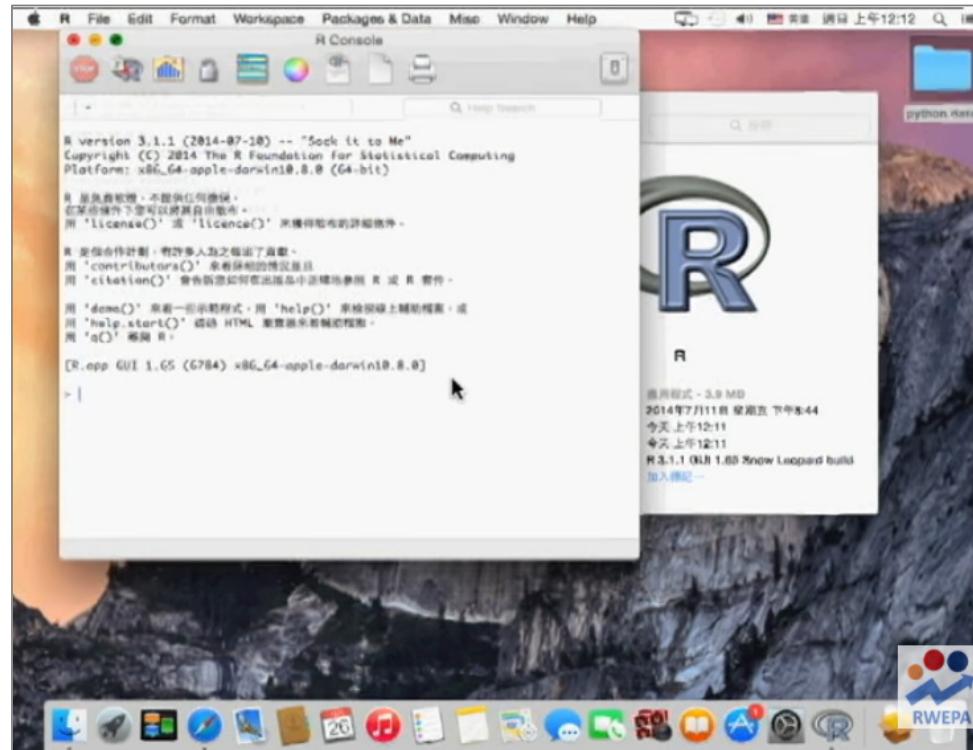


- demo(persp)



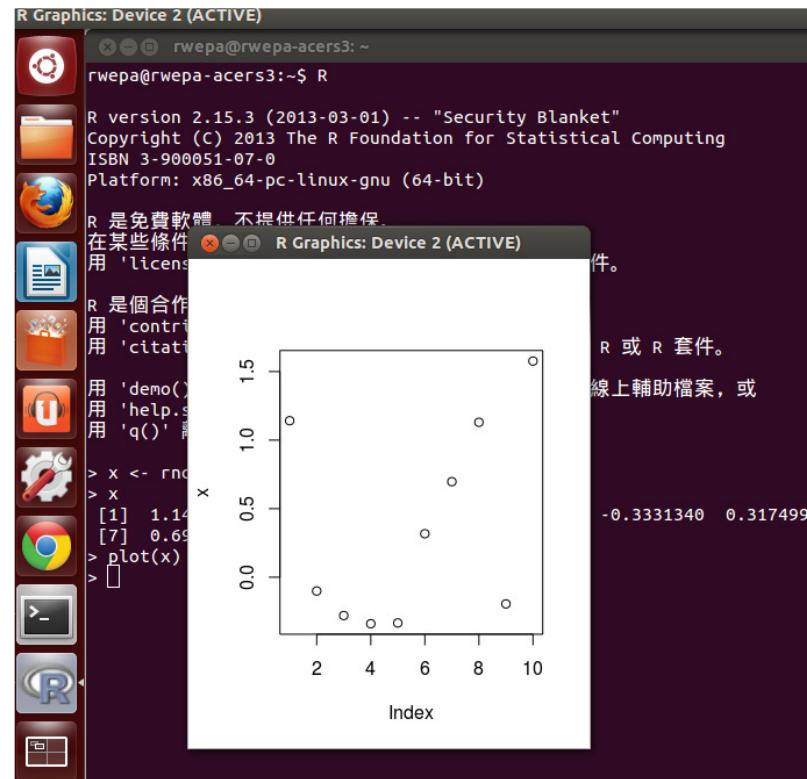
R for Mac

- <https://youtu.be/72MYRBNo5Bk>



R for Ubuntu

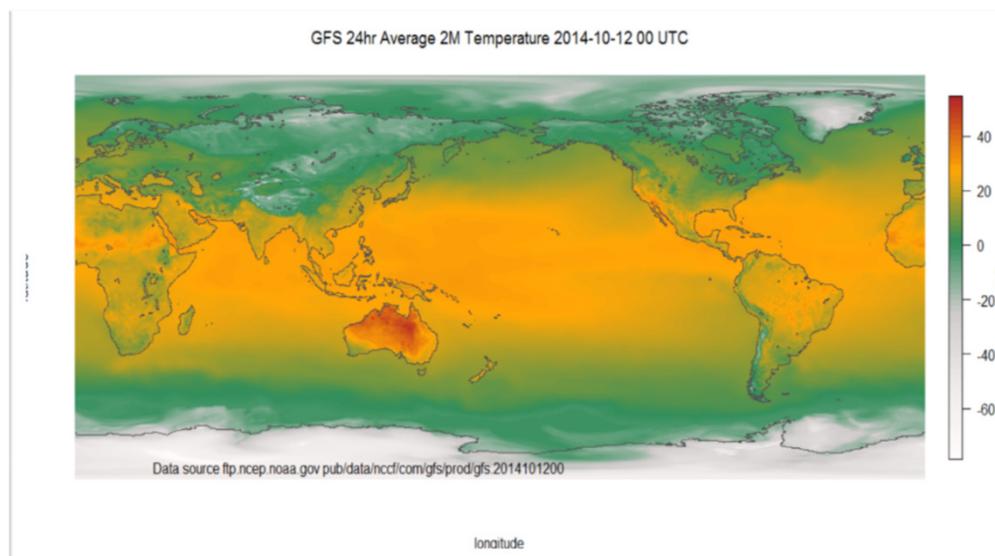
- <http://rwepa.blogspot.com/2013/05/ubuntu-r.html>



RStudio 簡介

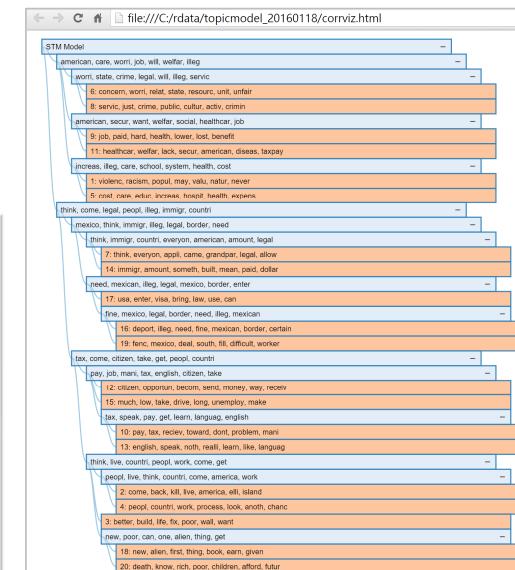
整合式開發環境 - RStudio

- <http://www.rstudio.com/>



視覺化應用

(全球2M氣溫圖)



主題模型

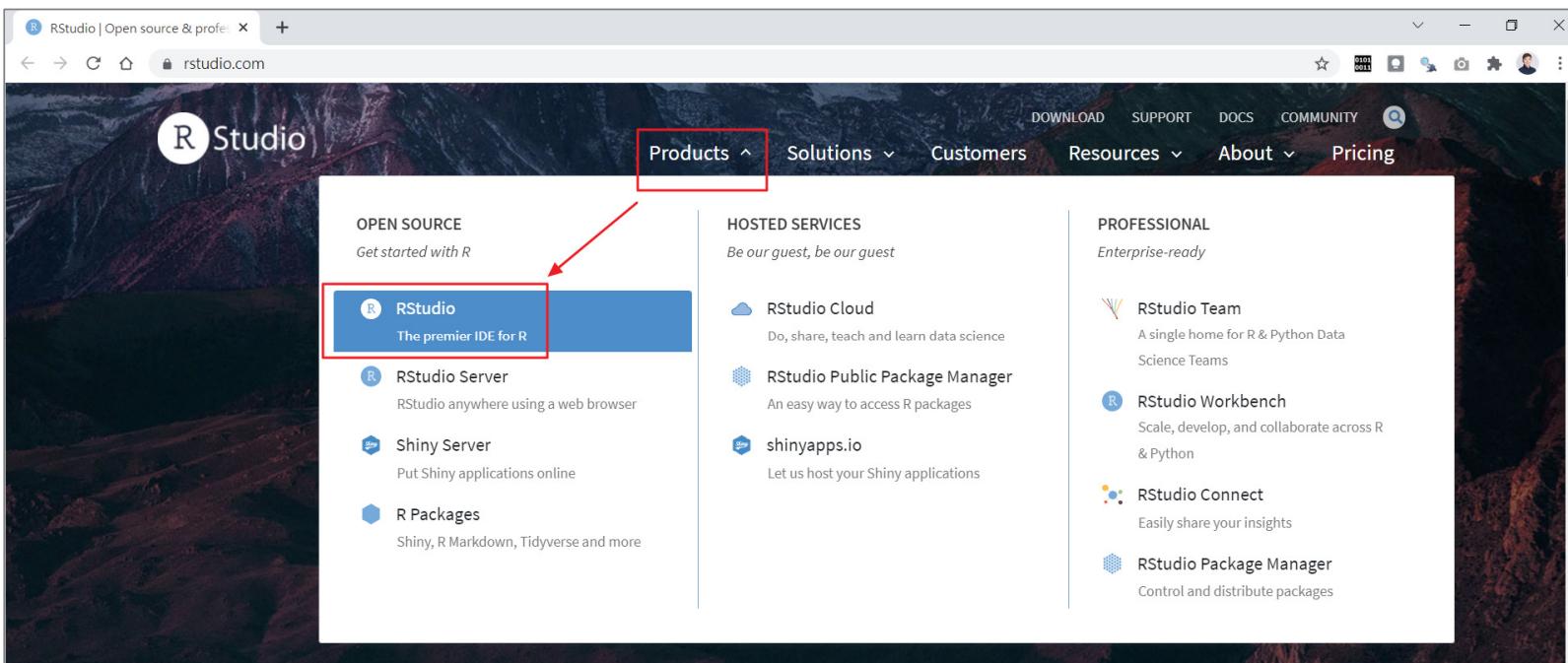
RStudio – 特性

- 支援智慧輸入 (按Tab)
- 高亮度顯示程式碼
- 整合R程式, 控制台, 變數清單, 繪圖視窗
- 整合資料庫匯入 SQL, Spark
- 整合R套件: shiny, rmarkdown
- 支援 RStudio外掛程式 (Addins)
- 安裝注意:
 - 先安裝R, 再安裝 RStudio
 - 安裝 RStudio時, 請先關閉R

RStudio 下載

RStudio 下載

- <http://www.rstudio.com/>



RStudio 下載 (續)

RStudio Desktop	RStudio Desktop Pro	RStudio Server	RStudio Workbench
Open Source License	Commercial License	Open Source License	Commercial License
Free	\$995	Free	\$4,975
單機版	/year	伺服器版本	/year (5 Named Users)
DOWNLOAD	BUY	DOWNLOAD	BUY
Learn more	Learn more	Learn more	Evaluation Learn more
Integrated Tools for R	✓	✓	✓
Priority Support	✓	✓	✓
Access via Web Browser			

A red arrow points from the "DOWNLOAD" button of the RStudio Desktop section to a yellow callout bubble containing the text "免費版". Another red arrow points from the "DOWNLOAD" button of the RStudio Server section to a yellow callout bubble containing the text "伺服器版本".

R/RStudio環境的基礎觀念

The screenshot displays the RStudio interface with four main components highlighted:

- Source 程式編輯區 (1):** Shows an R script file named "R語言資料視覺化應用(入門)-2021.R" containing comments about the file, author, date, email, and GitHub repository.
- 控制台(console) (2):** Shows the R console output for version 4.1.1, including copyright information and a welcome message.
- Environment (工作空間, xxx.Rdata) (3):** Shows the Global Environment pane stating "Environment is empty".
- Packages (4):** Shows the Packages pane listing various R packages available in the User Library, such as abind, ade4, AER, affy, affydata, affyio, agricolae, airGR, airGTeaching, airports, airway, AlgDesign, ALLMLL, and albrechtzager.

平均數檢定 - R

```
> library(foreign)
> myfile <- "c:/rdata/單一母體平均數檢定.sav"
> score <- read.spss(myfile)
> class(score) # list
[1] "list"
> score$成績
[1] 75 85 25 60 80 85 88 85 80 88 78 85 78 78 78 75 70 83 78 82 90 82 78 90 47 60 49 83
[29] 75 82 70 85 80 80 54 78 80 85 56 82
> mytest <- t.test(x = score$成績, alternative = c("two.sided"), mu = 70)
> mytest
1 2 3
One Sample t-test

data: score$成績
t = 2.5924, df = 39, p-value = 0.01335
alternative hypothesis: true mean is not equal to 70
95 percent confidence interval:
 71.21971 79.88029
sample estimates:
mean of x
 75.55

> mytest$p.value # 0.01334634
[1] 0.01334634
>
```

結果與 P.44 SPSS - 平均值檢定相同

補充篇：Python平均數檢定

Python

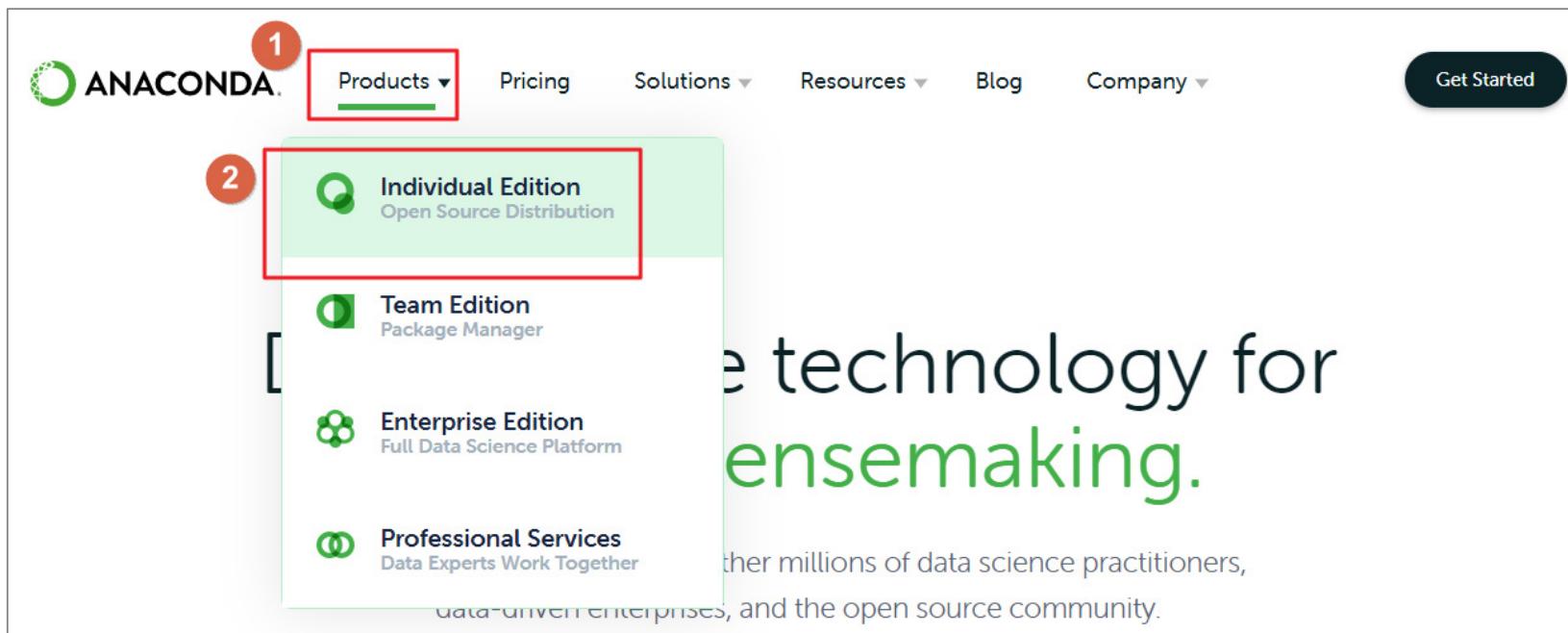
The screenshot shows the Python.org homepage. The top navigation bar includes links for Python, PSF, Docs, PyPI, Jobs, and Community. The main content area features the Python logo and navigation links for About, Downloads, Documentation, Community, Success Stories, News, and Events. A sidebar on the left displays a Python code snippet for calculating Fibonacci numbers:

```
# Python 3: Fib
>>> def fib(n):
>>>     a, b =
>>>     while a
>>>         pri
>>>         a,
>>>         print()
>>>     fib(1000)
0 1 1 2 3 5 8 13
```

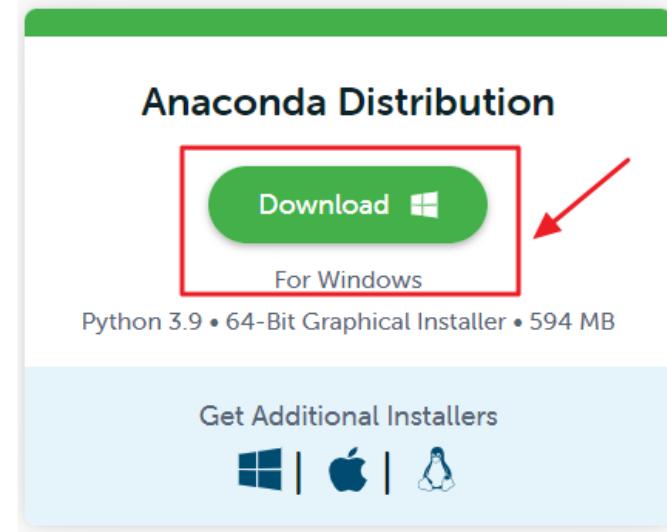
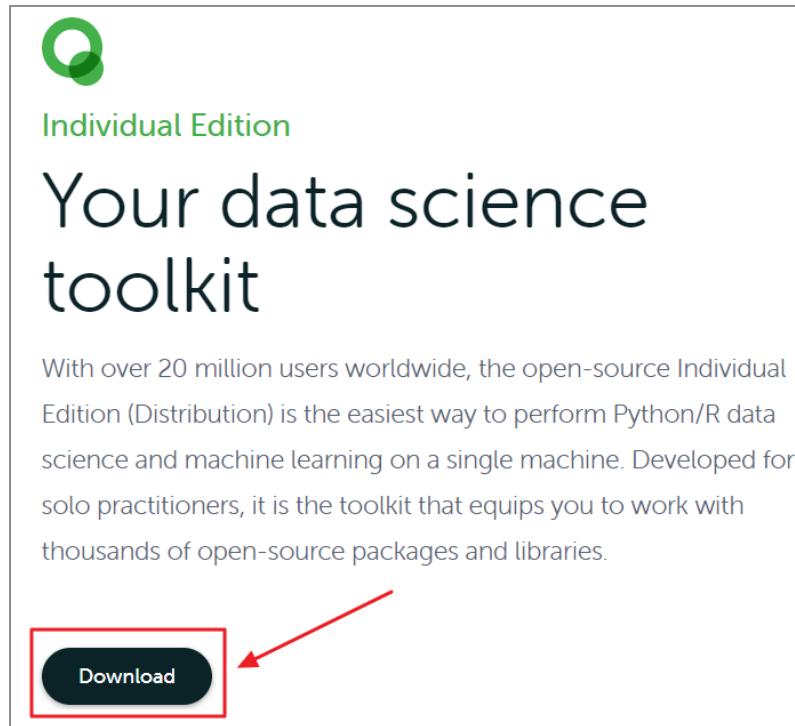
A dropdown menu is open over the 'Downloads' link, listing options: All releases, Source code, Windows, macOS, Other Platforms, License, and Alternative Implementations. A callout box at the bottom states: "Python is a programming language that lets you work quickly and integrate systems more effectively. [» Learn More](#)". An orange arrow points from this callout towards the 'Alternative Implementations' link in the dropdown menu.

Anaconda 下載

- <https://www.anaconda.com/>



Anaconda 下載 (續)



Anaconda3-2022.05-Windows-x86_64.exe

下載 64位元,
594 MB

統計套件



- **scipy.stats:** <https://scipy.org/>
 - Scientific computing tools for Python (科學計算)
 - 模組包含：統計、最佳化、線性代數、積分、插值、特殊函數、快速傅立葉變換、訊號處理和圖像處理、常微分方程式。
- **Statsmodels:** <https://www.statsmodels.org/>
 - 提供敘述性統計、統計檢定、統計模型，繪圖功能。
 - Statsmodels 加強 SciPy的統計功能。



參考資料

- RWEPA: <http://rwepa.blogspot.com/>
- 楊世瑩, SPSS 26統計分析嚴選教材(適用SPSS 26~22), 碩峰, 2020.
- 線上說明:
<https://www.ibm.com/docs/zh-tw/db2woc?topic=spss-tutorial-analyzing-statistics>
- R入門資料分析與視覺化應用教學(付費)
 - <https://mastertalks.tw/products/r?ref=MCLEE>
- R商業預測與應用(付費)
 - <https://mastertalks.tw/products/r-2?ref=MCLEE>
- Python 程式設計-李明昌 <免費電子書>
 - <http://rwepa.blogspot.com/2020/02/pythonprogramminglee.html>

謝謝您的聆聽

Q & A



李明昌

alan9956@gmail.com

<http://rwepa.blogspot.tw/>