

R-web 資料分析應用：分析方法 – 平均數檢定

溫怡然 副統計分析師

前兩期的生統 eNews 向大家介紹了許多實用的圖表繪製功能，相信大家對於資料特性已有初步的瞭解。本期開始，我們要利用【雲端資料分析暨導引系統】(R-web, <http://www.r-web.com.tw>) 來做進一步的分析，首先，就由『平均數檢定』揭開序幕。

平均數檢定會依據研究的問題與感興趣的群體而有所不同，其檢定包含：單一樣本平均數t檢定、獨立雙樣本平均數差異t檢定、成對雙樣本平均數差異t檢定、獨立多樣本平均數差異檢定（或稱變異數分析），我們會利用例子逐一介紹，並搭配雙尾、單尾檢定，介紹如下表；而在使用t檢定的時候，要注意資料需具有或近似”常態分配”的這個條件。本系列分析將統一使用源自基隆社區為基礎的整合篩檢計畫（Keelung Community-based Integrated Screen Program, KCIS）的心血管疾病資料作為範例資料檔，有關此資料的詳細資訊及變數定義請參閱[首期生統eNews](#)。

表一 雙尾、單尾檢定介紹

雙尾檢定	左尾檢定	右尾檢定
$H_0: \mu = \mu_0$	$H_0: \mu \geq \mu_0$	$H_0: \mu \leq \mu_0$
關心兩個方向的結果	關心特定方向的結果	關心特定方向的結果
α 平均分配在兩邊	α 全部放在左邊	α 全部放在右邊

➤ 單一樣本平均數 t 檢定

單一樣本平均數 t 檢定的使用時機為：當資料中僅討論單一樣本且樣本數較大時（通常以樣本筆數 ≥ 30 為區分標準），檢定母體平均數是否大於、小於或等於某一特定數值。這一部分我們搭配左尾檢定來舉例：腰圍是判斷罹患心血管疾病的一項重要指標，衛生署國民健康局 2002 年及 2007 年的調查發現，臺灣各年齡層男性腰圍明顯增加，如男性腰圍超過 90 公分、女性超過 80 公分就要有所警惕¹。現在，我們用範例資料來做檢定，檢定資料檔中男性的平均腰圍是否在安全值範圍內（< 90 公分）。

在 R-web 主選單中依序點選【分析方法】→【平均數檢定】→【單一樣本】即可進入分析頁面。

操作畫面如上圖所示。首先選擇要進行分析的資料檔，點選”使用者個人資料檔”後選擇”cvd_m”(男性)的檔案，系統將自動帶出參數設定畫面。在步驟二中選擇要進行分析的變數”Waist”(腰圍)，設定檢定平均數的地方填入”90”。最後點選【進階選項】如下圖，選擇檢定方向設定”單尾(左)”，如需要顯示樣本敘述統計量，可勾選後儲存設定即可【開始分析】。



下圖為分析結果，左側可以看到檢定的變數及相關設定，檢查沒問題後即可看分析結果，第一個表格為敘述統計量；在第二個表格中顯示虛無假設為母體平均數 ≥ 90 ，而 p-值 $< 2.22e-16^{***}$ 表顯著，拒絕虛無假設，代表資料中男性平均腰圍沒有超過標準值。

單一樣本平均數t檢定 - 分析結果

- 分析方法：單一樣本平均數t檢定
- 資料名稱：cvd_m
- 變數名稱：Waist
- 顯著水準：0.05
- 檢定平均數：90
- 檢定方向：左尾檢定
- 計算時間：0.305秒
- 樣本敘述統計量^I：

變數名稱	樣本數	平均數	中位數	最小值	最大值	標準差
Variable	Count	Mean	Median	Minimum	Maximum	Std. dev.
Waist	23417	84.3726	84	40	179	9.4415

I：樣本敘述統計量皆不包含遺失值

- 單一樣本平均數t檢定：

虛無假設：母體平均數 ≥ 90 $H_0: \mu \geq 90$							
變數名稱	t檢定統計量	自由度	臨界值	p-值 ^I	樣本平均數與母體平均數的差異	母體平均數的 95% 信賴區間	
						95% C.I. for μ	
Variable	t-statistics	d.f.	t(d.f., α)	p-value	Difference between sample mean and null	下界 Lower	上界 Upper
Waist	-91.2087	23416	-1.6449	$< 2.22e-16^{***}$	-5.6274	-Inf	84.4741

I：顯著性代碼：'****' : < 0.001, '***' : < 0.01, '**' : < 0.05, '#' : < 0.1

- 分析結果建議：由於檢定結果P-值($< 2.22e-16$) < 顯著水準0.05，因此可拒絕虛無假設。

➤ 獨立雙樣本平均數差異 t 檢定

獨立雙樣本平均數差異 t 檢定的使用時機為：當資料中兩組獨立樣本之樣本數皆較大時（通常以樣本筆數 ≥ 30 為區分標準），檢定兩組樣本間母體平均數的差異是否大於、小於或等於某一特定數值。這一部分我們搭配雙尾檢定來舉例：世界衛生組織指出，吸菸與二手菸暴露是心血管疾病的主要病因²，而在上一段落中，我們知道腰圍是判斷心血管疾病的一項指標¹，進而我們懷疑，有吸菸者與沒有吸菸者的腰圍是否會有統計上的顯著差異呢？讓我們來做個檢定吧！

在 R-web 主選單中依序點選【分析方法】→【平均數檢定】→【獨立雙樣本】即可進入分析頁面。

The screenshot displays the R-web analysis interface, divided into three steps:

- 步驟一：資料匯入** (Step 1: Data Import): A dropdown menu for '使用者個人資料檔' (User Personal Data File) is set to 'CVD'. Below it, a list of data files is shown, with 'CVD' selected. The text '您所選擇的資料檔為：CVD' (The data file you selected is: CVD) is displayed.
- 步驟二：資料型態設定** (Step 2: Data Type Setting): A dropdown menu for '資料型態為一檢定變數及一分組變數' (Data type as a test variable and a group variable) is highlighted. A link for '說明' (Help) is visible.
- 步驟三：參數設定** (Step 3: Parameter Setting): A list of variables is shown on the left, with 'Waist' selected in the '檢定變數' (Test Variable) field and 'Tobacco' selected in the '分組變數' (Grouping Variable) field. Below, a text input field for '設定檢定平均數差異 (預設為0)' (Set test mean difference (default 0)) contains the value '0'.

At the bottom, there are three buttons: '開始分析' (Start Analysis), '進階選項' (Advanced Options), and '重新設定' (Reset).

操作畫面如上圖所示。首先選擇要進行分析的資料檔，點選”使用者個人資料檔”後選擇”CVD”的檔案（此為原始檔），系統將自動帶出資料型態

設定，選擇欲進行檢定的資料型態中”資料型態為一檢定變數及一分組變數”。接著，系統會再帶出參數設定畫面，選擇要進行分析的變數之檢定變數”Waist”（腰圍）、分組變數”Tobacco”（抽菸與否），設定檢定平均數差異的地方填入”0”。

最後點選【進階選項】如下圖，選擇要進行分析的類別，系統會自動幫您分類好：1.選擇檢定方向設定”雙尾”；2.依雙樣本變異數檢定結果進行 t 檢定，此功能會先幫您檢定兩群體變異數是否相等，再依結果執行 t 檢定，如您想假設兩母體具有（或不具有）相同變異數，亦可自行點選；3.顯示樣本敘述統計量可依需求勾選，儲存設定後即可【開始分析】。

下圖為分析結果，左側可以看到檢定的變數及相關設定，檢查沒問題後即可看分析結果，第一個表格中顯示虛無假設為兩組資料的變異數相等，此為進階選項中，我們請系統幫忙執行的變異數檢定，而 p-值為 0.027363* 表顯著，拒絕虛無假設，表示兩組資料變異數不相等；第二個表格中的虛無假設則是母體平均數差異=0，而 p-值 < 2.22e-16*** 為顯著，拒絕虛無

假設，代表有吸菸者與沒有吸菸者的腰圍是有所差異的。

(獨立)雙樣本平均數差異t檢定 - 分析結果

- 分析方法：(獨立)雙樣本平均數差異t檢定
- 資料名稱：CVD
- 檢定變數：Waist
- 分組變數：Tobacco(0, 1)
- 顯著水準：0.05
- 檢定平均數差異：0
- 檢定方向：雙尾檢定
- 計算時間：0.134秒

- 雙樣本變異數(標準差)差異檢定^I：

虛無假設：兩組資料的變異數相等						
$H_0: \sigma_1^2/\sigma_2^2 = 1$						
變數名稱 variable	F檢定統計量 F-statistics	分子自由度 d.f. of numerator	分母自由度 d.f. of denominator	p-值 ^{II} p-value	母體變異數比值的 95% 信賴區間	
					95% C.I. for ratio	
					下界 lower	上界 upper
Waist	0.9726	45226	17279	0.027363 *	0.9487	0.9969

I：分組變數為Tobacco
II：顯著性代碼：'****'：< 0.001, '***'：< 0.01, '**'：< 0.05, '#'：< 0.1

- 雙樣本平均數差異t檢定(獨立樣本)^{II}：

虛無假設：母體平均數差異 = 0							
$H_0: \mu_1 - \mu_2 = 0$							
變數名稱 Variable	t檢定統計量 t-statistics	自由度 d.f.	臨界值 t(d.f., $\alpha/2$)	p-值 ^{III} p-value	樣本平均數與母體 平均數的差異 Difference between sample and null means	母體平均數差異 的 95% 信賴區間	
						95% C.I. for difference	
						下界 Lower	上界 Upper
Waist	-60.1386	30877.3	-1.96	< 2.22e-16 ***	-5.6122	-5.7951	-5.4293

I：分組變數為Tobacco
II：根據雙樣本變異數檢定結果，假設兩母體不具有相同變異數進行雙樣本平均數差異t檢定
III：顯著性代碼：'****'：< 0.001, '***'：< 0.01, '**'：< 0.05, '#'：< 0.1

- 分析結果建議：由於檢定結果P-值(< 2.22e-16) < 顯著水準0.05，因此可拒絕虛無假設。

➤ 成對雙樣本平均數差異 t 檢定

成對雙樣本平均數差異 t 檢定的使用時機為：比較資料中兩兩成對樣本的差異（如：減肥前體重與減肥後體重），且資料之樣本數較大時（通常以樣本筆數 ≥ 30 為區分標準）；分析時，即計算兩兩成對樣本的差異值，並檢定此差異值資料的母體平均數是否大於、小於或等於某一特定數值。

這部分我們搭配右尾檢定來舉例：國民健康署提醒，高血壓如果沒有即時發現，往往會導致後續心臟病、中風等嚴重的心血管疾病³；但別以為血壓正常就算安全，如果收縮壓及舒張壓相差超過 60 mmHg，表示血管彈性變差、血管硬化變嚴重，進而增加罹病風險⁴。我們用範例資料來檢定研究對象的平均脈壓差是否超過標準值 ($> 60 \text{ mmHg}$) ？

在 R-web 主選單中依序點選【分析方法】→【平均數檢定】→【成對雙樣本】即可進入分析頁面。

The screenshot shows the R-web analysis interface with two main steps:

- 步驟一：資料匯入 (Step 1: Data Input):** A dropdown menu for '使用者個人資料檔' (User Personal Data File) is open, showing a list of files: CVD, CVD_100, CVD_BP, cvd_f, and cvd_m. The 'CVD' file is selected and highlighted with a red box. Below the list, it says '您所選擇的資料檔為：CVD' (The data file you selected is: CVD).
- 步驟二：參數設定 (Step 2: Parameter Setting):** A list of variables (AC, HDL, TG, Betelnut, Alc_Drink, FamilyHx) is shown. Two variables are selected for comparison: 'SysBP' (labeled '檢定樣本一' - Test Sample 1) and 'DiaBP' (labeled '檢定樣本二' - Test Sample 2). Below this, a text box for '設定檢定平均數差異 (預設為0)' (Set test mean difference (default 0)) contains the value '60'.

At the bottom of the interface, there are three buttons: '開始分析' (Start Analysis), '進階選項' (Advanced Options), and '重新設定' (Reset). The '進階選項' button is highlighted with a red box.

操作畫面如上圖所示。首先選擇要進行分析的資料檔，點選”使用者個人資料檔”後選擇”CVD”的檔案（此為原始檔），系統將自動帶出參數設定畫面，選擇要進行分析的變數之檢定樣本一”SysBP”（收縮壓）、檢定樣本二”DiaBP”（舒張壓），設定檢定平均數差異的地方填入”60”。最後點選【進階選項】如下圖，選擇檢定方向設定”單尾（右）”，如需要顯示樣本敘述統計量，可勾選後儲存設定即可【開始分析】。

進階選項設定：

設定顯著水準

α :

選擇檢定方向： **單尾(右)** ▼

顯示樣本敘述統計量

下圖為分析結果，左側可以看到檢定的變數及相關設定，檢查沒問題後即可看分析結果，第一個表格為敘述統計量；第二個表格中顯示虛無假設為母體平均數差異 ≤ 60 ，而 p-值為 1 表不顯著，無法拒絕虛無假設，代表資料中研究對象之平均脈壓差在標準值範圍內。

(成對)雙樣本平均數差異t檢定 - 分析結果

- 分析方法：(成對)雙樣本平均數差異t檢定
- 資料名稱：CVD
- 變數名稱：SysBP - DiaBP
- 顯著水準：0.05
- 檢定平均數差異：60
- 檢定方向：右尾檢定
- 計算時間：0.389秒
- 樣本敘述統計量¹：

變數名稱 Variable	樣本數 Count	平均數 Mean	中位數 Median	最小值 Minimum	最大值 Maximum	標準差 Std. dev.
SysBP	63256	123.2666	120.5	70	276	20.8228
DiaBP	63245	78.0701	77	40	140	11.9757

I：樣本敘述統計量皆不包含遺失值

- 雙樣本平均數差異t檢定(成對樣本)：

虛無假設：母體平均數差異 ≤ 60 $H_0: \mu_1 - \mu_2 \leq 60$							
變數名稱 Variable	t檢定統計量 t-statistics	自由度 d.f.	臨界值 t(d.f., 1- α)	p-值 ¹ p-value	樣本平均數與母體平均數的差異 Difference between sample mean and null	母體平均數差異的 95% 信賴區間 95% C.I. for difference	
						下界 Lower	上界 Upper
SysBP - DiaBP	-260.4796	63204	1.6449	1	-14.8404	45.0659	Inf

I：顯著性代碼： '****' : < 0.001, '***' : < 0.01, '**' : < 0.05, '#' : < 0.1

- 分析結果建議：由於檢定結果P-值(1) > 顯著水準0.05，因此無法拒絕虛無假設。

➤ 獨立多樣本平均數差異檢定

獨立多樣本平均數差異檢定（或稱變異數分析）的使用時機為：當資料中包含多組樣本（三組以上之樣本）且樣本數皆較大時（通常以樣本筆數 ≥ 30 為區分標準），檢定多組樣本間母體平均數是否有差異。此部分接續獨立雙樣本平均數差異 t 檢定的例子來延伸：我們已知有吸菸者與沒有吸菸者的腰圍是有所差異的，接著就進一步將吸菸者分為三個等級，每日一包、每日兩包及每日三包以上，檢定這三組吸菸者的平均腰圍是否有差異？

在 R-web 主選單中依序點選【分析方法】→【平均數檢定】→【獨立多樣本（或稱變異數分析）】即可進入分析頁面。

The screenshot displays the R-web analysis interface, divided into three steps:

- 步驟一：資料匯入** (Step 1: Data Import): A dropdown menu for '使用者個人資料檔' (User Personal Data File) is open, showing a list of files including 'CVD_100', 'CVD_BP', 'cvd_f', 'cvd_m', and 'cvd_tobacco'. The 'cvd_tobacco' file is highlighted with a red box. Below the list, it says '您所選擇的資料檔為: cvd_tobacco'.
- 步驟二：資料型態設定** (Step 2: Data Type Setting): A dropdown menu for '資料型態' (Data Type) is set to '資料型態為一檢定變數及一分組變數' (Data type is a test variable and a group variable), which is highlighted with a red box. A '說明' (Help) link is next to it.
- 步驟三：參數設定** (Step 3: Parameter Setting): A list of variables is shown on the left, including 'AC', 'HDL', 'TG', 'Betelnut', 'Alc_Drink', 'FamilyHx', and 'Tobacco'. On the right, there are two input fields: '檢定變數' (Test Variable) with 'Waist' entered, and '分組變數' (Group Variable) with 'Tobacco_Consumpt' entered.

At the bottom of the interface, there are three buttons: '開始分析' (Start Analysis), '進階選項' (Advanced Options), and '重新設定' (Reset). The '進階選項' button is highlighted with a red box.

操作畫面如上圖所示。首先選擇要進行分析的資料檔，點選”使用者個人資料檔”後選擇”cvd_tobacco”的檔案（有吸菸者），系統將自動帶出資料

型態設定，選擇欲進行檢定的資料型態中”資料型態為一檢定變數及一分組變數”。接著，系統會再帶出參數設定畫面，選擇要進行分析的變數之檢定變數”Waist”（腰圍）、分組變數”Tobacco_Consumption”（菸草消費量）。

最後點選【進階選項】如下圖，模式設定中，選擇要進行分析的類別（此例為菸草消費量的三組類別），系統會自動幫您分類好；顯示樣本敘述統計量及顯示多樣本變異數（標準差）差異檢定可依需求勾選。多重比較、殘差及影響點也可自行選擇，儲存設定後即可【開始分析】。

下圖為分析結果，左側可以看到檢定的變數及相關設定，檢查沒問題後即可看分析結果，第一個表格為敘述統計量；第二個表格為模式摘要；第三個表格顯示虛無假設為各母體的變異數相等，p-值為 0.023663* 表顯著，拒絕虛無假設，表示各組資料變異數不完全相等。

(獨立)多樣本平均數差異檢定(或稱變異數分析) - 分析結果

- 分析方法：(獨立)多樣本平均數差異檢定(或稱變異數分析)
- 資料名稱：cvd_tobacco
- 檢定變數：Waist
- 分組變數：Tobacco_Consumption(1, 2, 3)
- 顯著水準：0.05
- 計算時間：0.837秒

- 樣本敘述統計量^I：

變數名稱 Variable	分組變數(Tobacco_Consumption) grouping variable(Tobacco_Consumption)	樣本數 Count	平均數 Mean	中位數 Median	最小值 Minimum	最大值 Maximum	標準差 Std. dev.
Waist	1	14208	82.1411	82	40	175	10.4262
	2	1573	85.6348	86	48	125	9.9479
	3	164	87.4878	89	41	109	10.305
	不分組(Total)	15945	82.5407	83	40	175	10.4424

I：樣本敘述統計量皆不包含遺失值

- 模式摘要：

判定係數 R ²	1.23 %
調整判定係數 Adjusted R ²	1.22 %
檢定變數平均數 Mean of responses	82.5407
殘差標準誤差 Residual std. err.	10.3788

- 多樣本變異數(標準差)差異檢定^I：

虛無假設：各母體的變異數相等 H ₀ : $\sigma_1^2 = \dots = \sigma_3^2$					
變數名稱 variable	F檢定統計量 F-statistics	分子自由度 d.f. of numerator	分母自由度 d.f. of denominator	臨界值 F(d.f.1,d.f.2,1- α)	p-值 ^{II} p-value
Waist	3.7447	2	15942	2.9963	0.023663 *

I：分組變數為Tobacco_Consumption
II：顯著性代碼： '****' : < 0.001, '***' : < 0.01, '**' : < 0.05, '#' : < 0.1

第四個表格中的虛無假設則是各母體的平均數相等，而 p -值 $< 2.22e-16^{***}$ 為顯著，拒絕虛無假設，表示各組資料平均數不完全相等。在最後的表格五中，我們可以看到兩兩比較的結果：第一組與第二組、第一組與第三組的 P -值為 $< 2e-16^{***}$ 為顯著，表示組與組之間的平均腰圍是有差異的；而第二組與第三組的 P -值為 $0.093787\#$ 為不顯著，代表兩組間的平均腰圍沒有統計上的差異。

• 變異數分析^I：

虛無假設：各母體的平均數相等						
$H_0: \mu_1 = \dots = \mu_3$						
來源	平方和	自由度	均方和	F檢定統計量	臨界值	p-值 ^{II}
source	sum of squares	d.f.	mean square	F-statistics	F(d.f.1,d.f.2,1- α)	p-value
處理	21341.52	2	10670.76	99.0605	2.9963	$< 2.22e-16^{***}$
誤差	1717267	15942	107.7197			
總和	1738608	15944				

I：分組變數為Tobacco_Consumption

II：顯著性代碼：'***' : < 0.001 , '**' : < 0.01 , '*' : < 0.05 , '#' : < 0.1

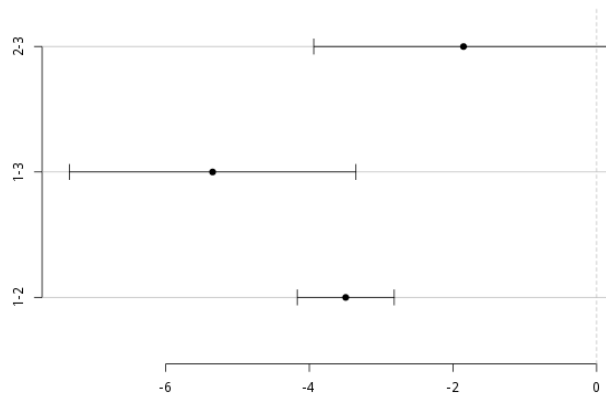
- 分析結果建議：由於檢定結果 P -值($< 2.22e-16$) $<$ 顯著水準 0.05 ，因此可拒絕虛無假設。
- 兩兩(pairwise)多重比較(分組變數為Tobacco_Consumption)：

◦ Scheffe兩兩比較^I：

	差異	95% 信賴區間		修正P-值
		下界 Lower	上界 Upper	Adj. p-value
1-2	-3.4937	-4.16883	-2.81856	$< 2e-16^{***}$
1-3	-5.34673	-7.34211	-3.35135	$< 2e-16^{***}$
2-3	-1.85303	-3.93785	0.23179	$0.093787\#$

I：顯著性代碼：'***' : < 0.001 , '**' : < 0.01 , '*' : < 0.05 , '#' : < 0.1

95% Scheffe confidence intervals



本期生統 eNews 介紹到此，這次分別介紹了平均數檢定的四大項目：單一樣本平均數 t 檢定、獨立雙樣本平均數差異 t 檢定、成對雙樣本平均數差異 t 檢定、獨立多樣本平均數差異檢定，相信大家對於檢定方法與操作方式有更深入的瞭解。下一期的生統 eNews 將為大家介紹『中位數檢定』，敬請期待！

參考資料

1. 台灣男人腰圍 5 年增加 3 公分

<https://tw.news.yahoo.com/台灣男人腰圍-5年增加3公分-213000643.html>

2. 天冷、須嚴防心血管疾病發作猝死

<http://health99.hpa.gov.tw/txt/HealthyHeadlineZone/HealthyHeadlineDetail.aspx?TopicNo=6728>

3. 保護心血管 從定期量血壓開始

http://www.mohw.gov.tw/cht/ministry/DM2_P.aspx?f_list_no=7&fod_list_no=4557&doc_no=44955

4. 脈壓差逾 60 心血管硬化高危險群

<http://www.cvmetrics.com.tw/super%20monopoly/news/news2.htm>