平均數檢定 (Hypothesis Test for Mean)

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

- 1.平均數檢定
- 2.SPSS-單一樣T檢
- 3.SPSS-獨立樣本T檢定
- 4.SPSS-成對樣本T檢定

1.平均數檢定

假設檢定之基本概念

- (一)假設檢定(Hypothesis test)
  - 對母體參數作出一適當的假設，然後根據隨機抽樣之樣本，利用樣本統計量之抽樣分配來決定接受或拒絕假設的過程。
  - 假設檢定的第一個步驟便是建立假設，因此以下須定義統計假設。
- (二)統計假設(Statistical hypothesis)
  - 指對一個或多個母體參數的一個推測。

統計假設

- 統計假設的範例：
  1. 年平均地面氣溫為24度：單一母體參數之推估。
  2. 台灣電腦公司所生產的電腦之不良率小於0.1：單一母體參數之推估。
  3. 台北市與新北市的年平均AQI是否相同：兩母體參數比較之推測。
- 統計假設依其性質可分為以下兩種 ：
  1. 虛無假設(null hypothesis)：通常為研究者欲推翻之統計假設，即假設檢定中之主要假設，一般以 H_0 表示。
  2. 對立假設(alternative hypothesis)： 假設虛無假設不成立，即虛無假設之互補假設，一般以 H_1 表示。

範例：單一母體參數之推估

- 某手機業者宣稱其手機之平均待機時間為96小時，請問消費者欲檢定此手機業者之宣稱是否為真，該如何進行假設？
- 【解】
- 虛無假設 H_0：平均待機時間大於或等於96小時，即$H_0: \mu \geq 96$。
- 對立假設 H_1：平均待機時間小於96小時。即 $H_0: \mu < 96$，小於表示左尾檢定。
- 說明：
  1. 消費者懷疑業者所宣稱，最後消費者希望 拒絕 H_0，而接受 H_1。
  2. 虛無假設與對立假設不可以有交集。

假設檢定的二大類別

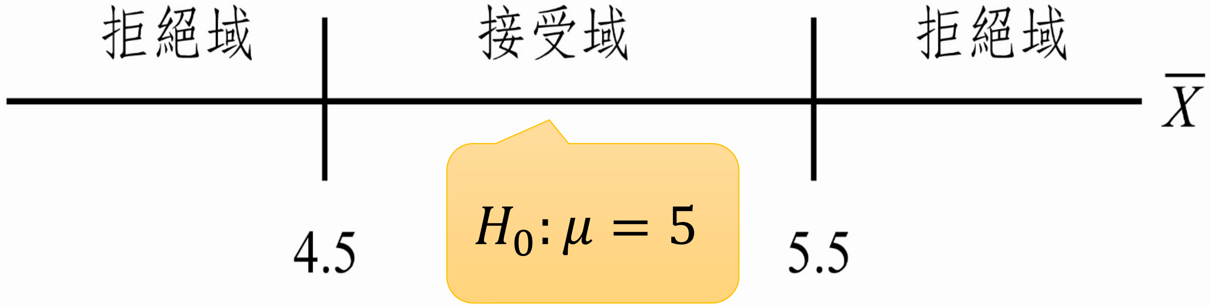
- 依據樣本統計量所定訂拒絕 H_0 範圍的不同，可將假設檢定的形式分成以下兩種：
- 單尾檢定(one-tailed test)：
  - 當樣本統計量僅在大於某個數值或小於某個數值之其中一種情形之下拒絕 H_0 之檢定。
  - 若拒絕 H_0 為樣本統計量大於某個數值時，則此單尾檢定又稱右尾檢定（right-tailed test）。
  - 若拒絕 H_0 為小於某個數值時，則稱左尾檢定（left-tailed test）。
- 雙尾檢定(two-tailed test)：
  - 當樣本統計量大於某個數值或小於某個數值均可能拒絕 H_0 之檢定。

假設檢定之判斷

- 判斷以下檢定為哪一種類?
- (1) $H_0: \mu \geq 250$，$H_1: \mu < 250$
- (2) $H_0: \mu = 250$，$H_1: \mu \neq 250$
- (3) $H_0: p \leq 0.1$，$H_1: p > 0.1$
- 【解】
- (1)及(3)為單尾檢定之假設且(1)為左尾檢定之假設，H_1是小於。
- (2)為雙尾檢定之假設。
- (3)為右尾檢定之假設，H_1是大於。

拒絕域 vs.接受域

- 依據誤差程度可以決定一個拒絕虛無假設的範圍，而此範圍稱之為臨界域或拒絕域 (critical region)。
- 決定了拒絕域，相對地可以產生接受域 (acceptance region) 。
- 考慮 $H_0: \mu = 5$ & $H_1: \mu \neq 5$ 為例。



假設檢定-型I錯誤、型II錯誤

- 型 I 錯誤（Type I error）
  - 當虛無假設 H_0 為真而拒絕 H_0，稱之為型 I 錯誤。
  - 造成型 I 錯誤的機率以 μ 表示，定義如下：
  - $\alpha = P(\text{型I錯誤}) = P(\text{拒絕}H_0|H_0\text{為真})$
- 型 II 錯誤（Type I error）
  - 當虛無假設 H_0 為假（H_1 為真）而接受 H_0，稱之為型 II 錯誤。
  - 造成型 II 錯誤的機率以 β 表示，定義如下：
  - $\beta = P(\text{型II錯誤}) = P(\text{拒絕}H_0|H_1\text{為真})$

範例：α值的推論

- 考慮常態母體之變異數為5，今對此母體平均數作以下之假設，$H_0: \mu = 5 \& H_1: \mu \neq 5$，並決定其拒絕域為 $\{\bar{x} > 5.5$ 或 $\bar{x} < 4.5\}$ 。請問以樣本個數為20之樣本所得之樣本平均數來檢定母體平均數所造成之型 I 錯誤之機率，其 α 值為何？
- 【解】

參考「平均數估計(Mean Estimation)」 P.54

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \rightarrow N(0,1)$$

$$\begin{aligned}
\alpha &= P(\text{拒絕}H_0|H_0\text{為真}) = P(\bar{X} > 5.5\text{或}\bar{X} < 4.5 \mid \mu = 5) \\
&= 1 - P(4.5 \leq \bar{X} \leq 5.5|\mu = 5) = 1 - P\left(\frac{4.5-5}{\sqrt{\frac{5}{20}}} \leq \frac{\bar{X}-5}{\sqrt{\frac{5}{20}}} \leq \frac{5.5-5}{\sqrt{\frac{5}{20}}}\right) \\
&= 1 - P(-1 \leq Z \leq 1)\text{，其中 } R.V.Z \sim N(0,1) \\
&= 0.318
\end{aligned}$$

範例：β值的推論

- 承上題範例，若 $\mu = 6$ 時，求造成型II 錯誤之機率β 值。
- 【解】

$$\beta = P(4.5 \le \bar{X} \le 5.5 | \mu = 6)$$

$$= P\left(\frac{4.5-6}{\sqrt{\frac{5}{20}}} \le Z \le \frac{5.5-6}{\sqrt{\frac{5}{20}}}\right)$$ ，$R.V.\ \bar{X} \sim \left(6,\ \frac{5}{20}\right), \mu = 6, \sigma^2 = \frac{5}{20}$

$$= P(-3 \le Z \le -1)$$

$$= P(Z \le -1) - P(Z < -3)$$

$$= 0.158$$

α $vs.$ β

事實 / 決策	H_0 為真	H_0 非真
拒絕 H_0	型 I 錯誤 (α)	√
接受 H_0	√	型 II 錯誤 (β)

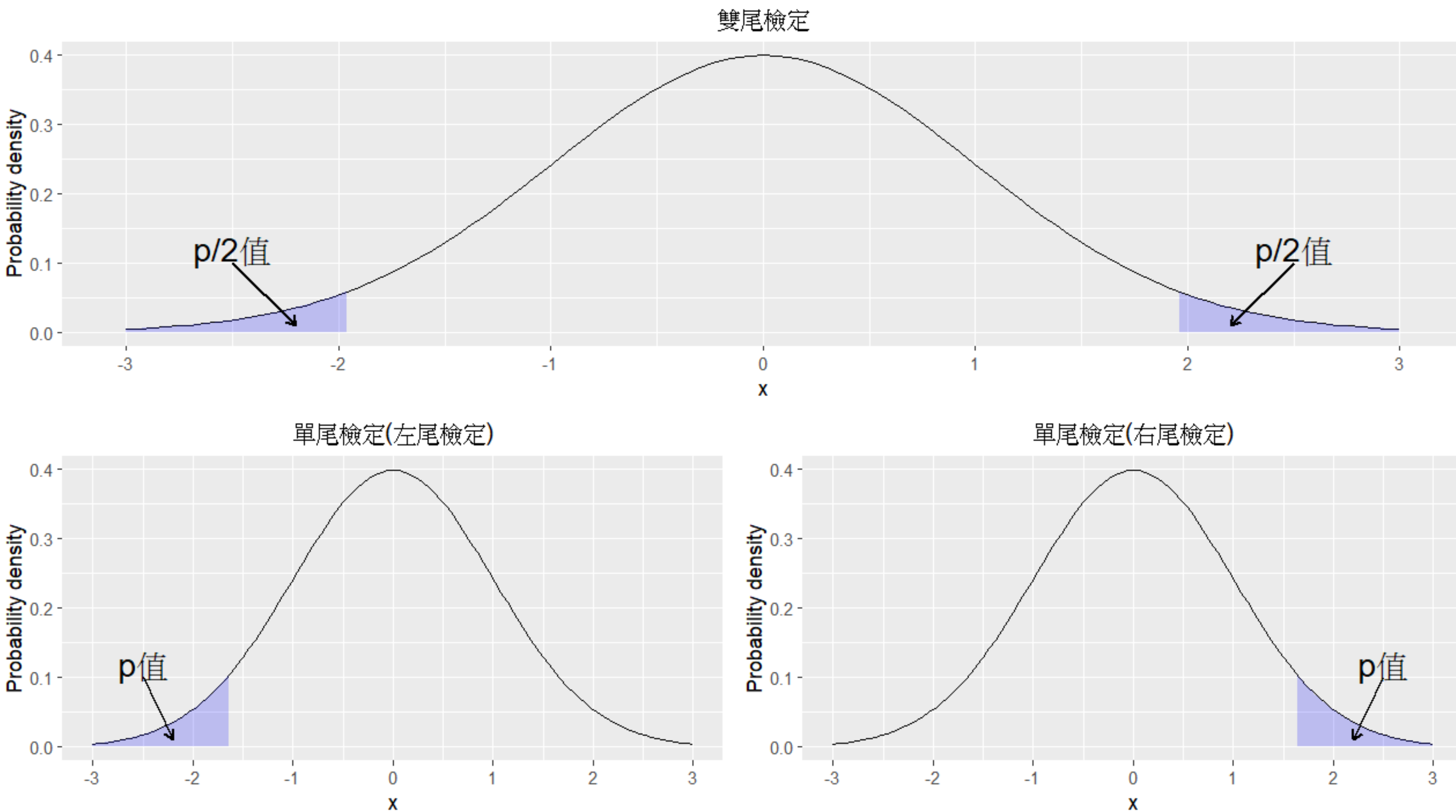
假設檢定二大方法

- 在假設檢定中，通常事先並不控制 β 值，而先決定容許犯型 I 錯誤的最大機率 α 值，即顯著水準（significant level）。相對地，在給予對立假設成立之參數後，1 - β 則稱為此檢定之「檢定力」(power of the test)， 此時β值越小，其檢定力（$1-\beta$）越大。

統計量 vs. 統計值

- 檢定統計假設的兩種方法：
  1. 臨界值法(critical value method)：給定顯著水準 α 值，然後決定拒絕域後，再依據所得之樣本，計算其樣本之統計量 (即為檢定值或稱為統計值)，最後再判定上述檢定值是否落在拒絕域中。
  2. P 值法(P-value method)：在 H_0 為真的條件下，計算由樣本統計值而拒絕 H_0 的最大機率。不論是單尾或雙尾檢定，若 P 值小於 α 值，則拒絕虛無假設 H_0，否則便勉強接受 H_0，一般 SPSS等統計軟體採用此方法較簡單。

P值法


雙尾檢定
Probability density
0.4
0.3
0.2
0.1
0.0
p/2值
p/2值
-3
-2
-1
0
1
2
3
x
單尾檢定(左尾檢定)
p值
單尾檢定(右尾檢定)
p值

假設檢定程序 [五大步驟]

- (1)建立假設(虛無假設與對立假設)。
- (2)選擇檢定之統計量，並給定顯著水準 α 值。
- (3)決定檢定方法(臨界值法或P 值法)，若選擇臨界值法，則決定拒絕域。 P 值法可以使用 $\alpha = 0.05$。
- (4)蒐集樣本並計算檢定值。
- (5)結論：
  - 臨界值法：若檢定值落在拒絕域，則拒絕 H_0 ，否則便接受 H_0 。
  - P 值法：計算P 值，若P 值小於α，則拒絕 H_0，否則便接受 H_0。

單一母體平均數之假設檢定

- 常態母體且 σ^2 已知，採用常態分配。
- 常態母體且 σ^2 未知，採用 $t(n-1)$ 分配。
- 大樣本（$n > 30$），採用常態分配。

case 1. 常態母體且 σ^2 已知

- 在常態母體且 σ^2 已知之條件下，若 $\bar{x}$ 為隨機樣本 $x_1, x_2, \ldots, x_n$ 之平均數，且顯著水準為 α ，則以 $z_0 = \frac{\bar{x}_0 - \mu_0}{\sigma/\sqrt{n}}$ 為檢定值，其拒絕域與 P 值如下：
- 左尾檢定：$H_0: \mu \geq \mu_0$，$H_1: \mu < \mu_0$
  - 拒絕域為 $\{z_0 < -z_\alpha\}$，P 值$= P(Z < z_0)$ 。
- 右尾檢定：$H_0: \mu \leq \mu_0$，$H_1: \mu > \mu_0$
  - 拒絕域為 $\{z_0 > z_\alpha\}$ ，P 值$= P(Z > z_0)$ 。
- 雙尾檢定： $H_0: \mu = \mu_0$，$H_1: \mu \neq \mu_0$
  - 拒絕域為 $\{|z_0| < z_{\frac{\alpha}{2}}\}$ ，P 值$= 2P(Z > |z_0|)$ 。

範例：常態母體且 σ^2 已知

- 某一廠商產品重量之標準差為5公克。今此廠商宣稱其產品的平均重量恰為250公克，若隨機由該公司抽取16件產品秤其重量，得其平均數為246公克，請以顯著水準 $\alpha = 0.05$ 檢定此廠商宣稱是否為真？(假設母體具常態分配)
- 【解】
- 方法一：臨界值法
- 步驟 1：$H_0: \mu = 250$，$H_1: \mu \neq 250$，即雙尾檢定。
- 步驟 2：$X \sim N(\mu, 5^2)$，$\bar{x} = 246$，$n = 16$ 以 $Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$ 為檢定之統計量。
- 步驟 3：$\{|z_0| > z_{\alpha/2}\} = \{|z_0| > z_{0.025}\} = \{|z_0| > 1.96\}$
- 步驟 4：$|z_0| = \left|\frac{\bar{x}_0 - \mu_0}{\sigma/\sqrt{n}}\right| = \left|\frac{246 - 250}{\sqrt{25/16}}\right| = 3.2 > 1.96$
- 步驟 5：拒絕 H_0 ，即此廠商宣稱非真。

範例：常態母體且 σ^2 已知 (續)

- 方法二：P 值法
- 步驟 1：參考方法1。
- 步驟 2：參考方法1。
- 步驟 3：P值 $= 2P\left(Z < \frac{246-250}{5/\sqrt{16}}\right)$

  $= 2P(Z < -3.2) = 2 \times 0.0007 = 0.0014$
- 步驟 4：P值 $= 0.0014 < 0.05$
- 步驟 5：拒絕 H_0，即此廠商宣稱非真。

case 2. 常態母體且 σ^2 未知

- 若 $\bar{x}$、s^2 分別為隨機樣本 $x_1, x_2, \cdots, x_n$ 之平均數與變異數，顯著水準為 α，以 $t_0 = \frac{\bar{x}_0 - \mu_0}{s/\sqrt{n}}$ 為檢定值，則其拒絕域與P 值如下：
- 左尾檢定：$H_0: \mu \geq \mu_0$，$H_1: \mu < \mu_0$
    - 拒絕域為 $\{t_0 < -t_\alpha(n-1)\}$， P 值$= P(T < t_0)$ 。
- 右尾檢定：$H_0: \mu \leq \mu_0$，$H_1: \mu > \mu_0$
    - 拒絕域為 $\{t_0 > t_\alpha(n-1)\}$， P 值$= P(T > t_0)$ 。
- 雙尾檢定： $H_0: \mu = \mu_0$，$H_1: \mu \neq \mu_0$
    - 拒絕域為 $\{|t_0| > t_{\frac{\alpha}{2}}(n-1)\}$， P 值$= 2P(T > |t_0|)$ 。

範例：常態母體且 σ^2 未知

- 某一廠牌行動電話宣稱其平均重量不超過78公克，今隨機抽取此廠牌行動電話10支，得其平均重80公克，標準差4公克。請以臨界值法且顯著水準$\alpha = 0.05$ 來檢定此廠商宣稱是否為真？(假設母體具有常態分配)
- 【解】
- 步驟 1：$H_0: \mu \le 78$，$H_1: \mu > 78$，即右尾檢定。
- 步驟 2：以$t_0 = \frac{\bar{x}_0 - \mu_0}{s/\sqrt{n}}$ 為檢定之統計量。
- 步驟 3：$\{t_0 > t_{0.05}(10-1)\} = \{t_0 > t_{0.05}(9)\} = \{t_0 > 1.833\}$
- 步驟 4：$t_0 = \frac{\bar{x}_0 - \mu_0}{s/\sqrt{n}} = \frac{80-78}{4/\sqrt{10}} = 1.581$
- 步驟 5：因為統計值 $1.581 < 1.833$，勉強接受此廠商之宣稱平均重量不超過78公克 。

練習 t 分配查表法

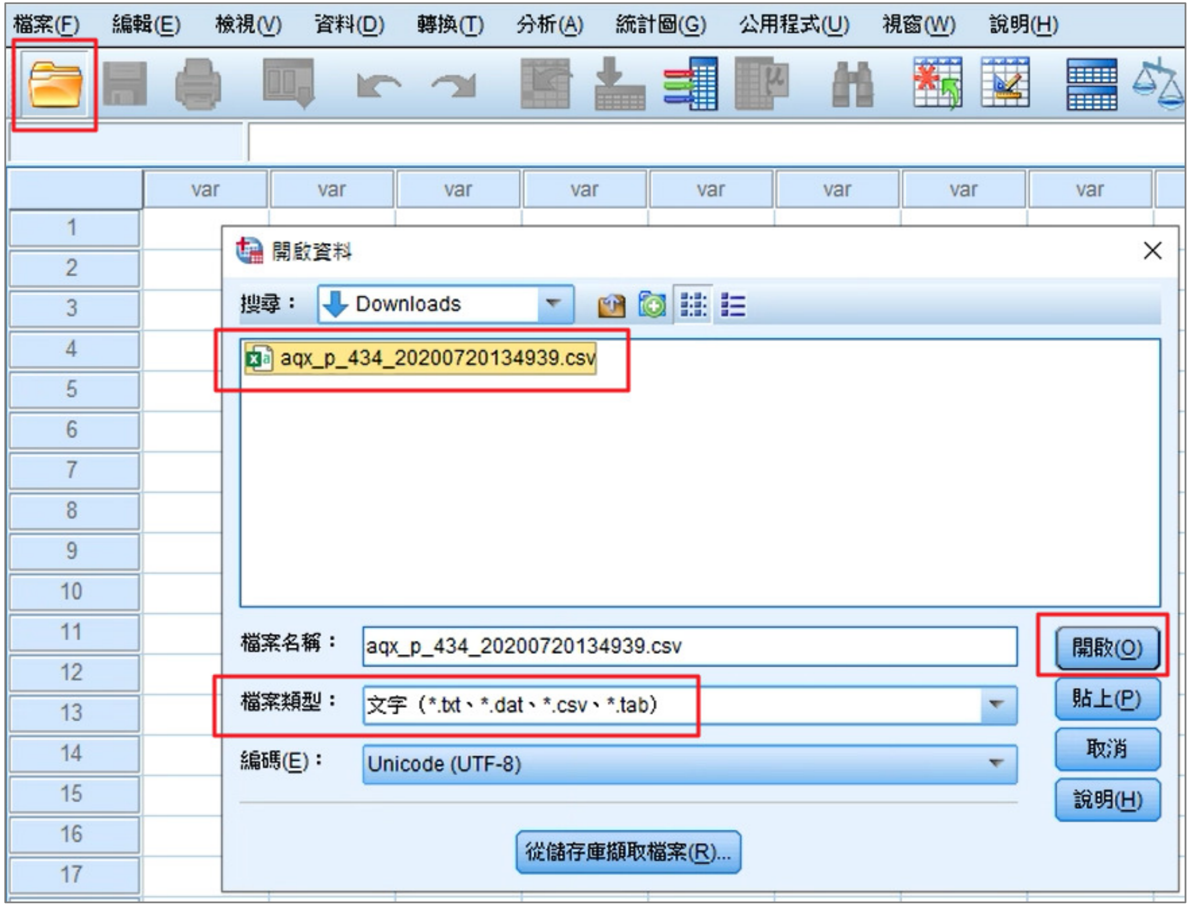
2.SPSS-單一樣T檢定

單一樣本T檢定

- 分析 \ 比較平均數法 \ 單一樣本T檢定
- 使用時機
  1. 樣本數小於30
  2. 母體標準差未知
  3. 檢定一組資料的平均值與指定的目標值之間是否存在統計上差異。
  4. 例：淡水每日平均AQI是否與全台平均AQI相同。

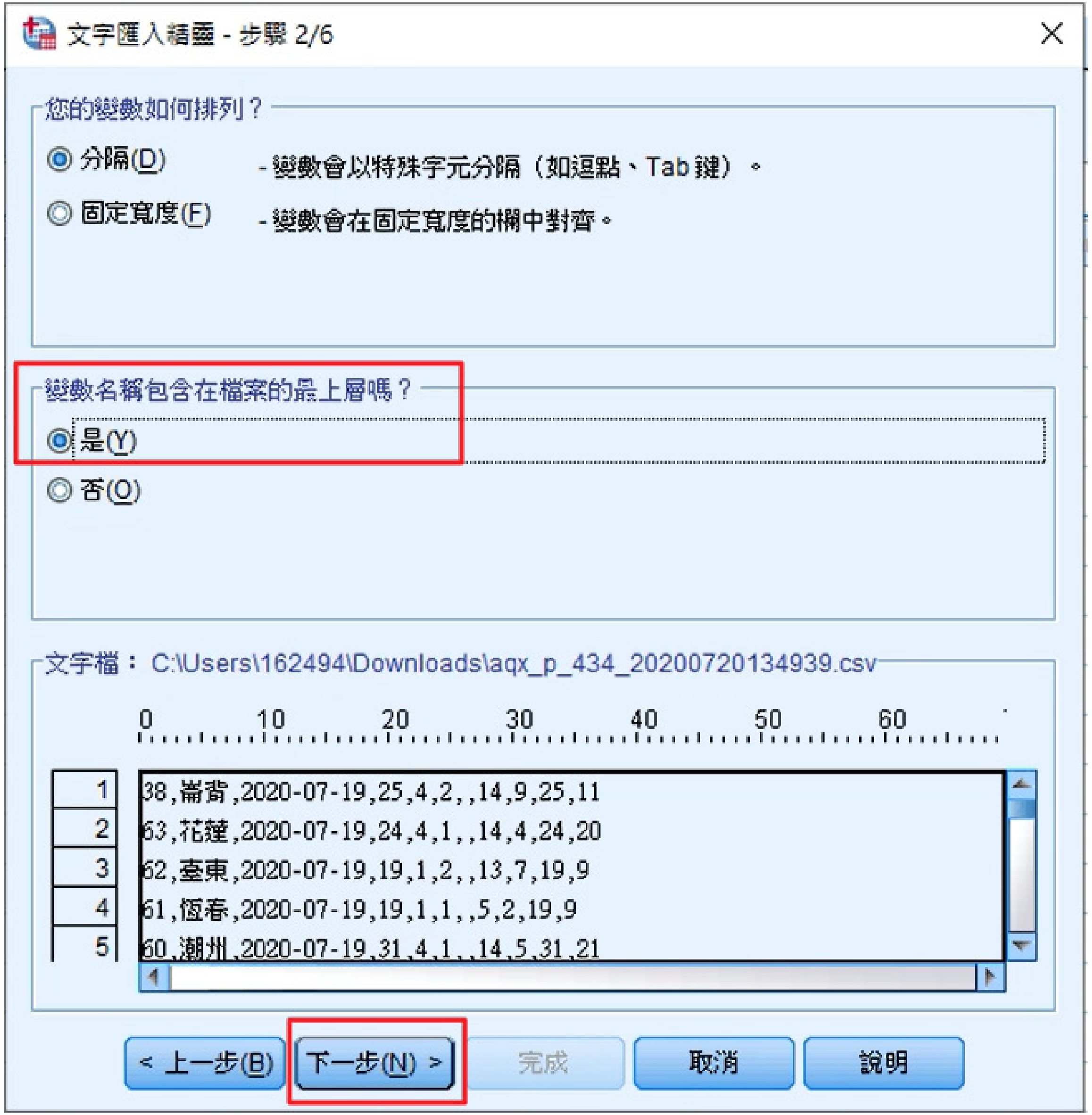
範例：全台平均AQI

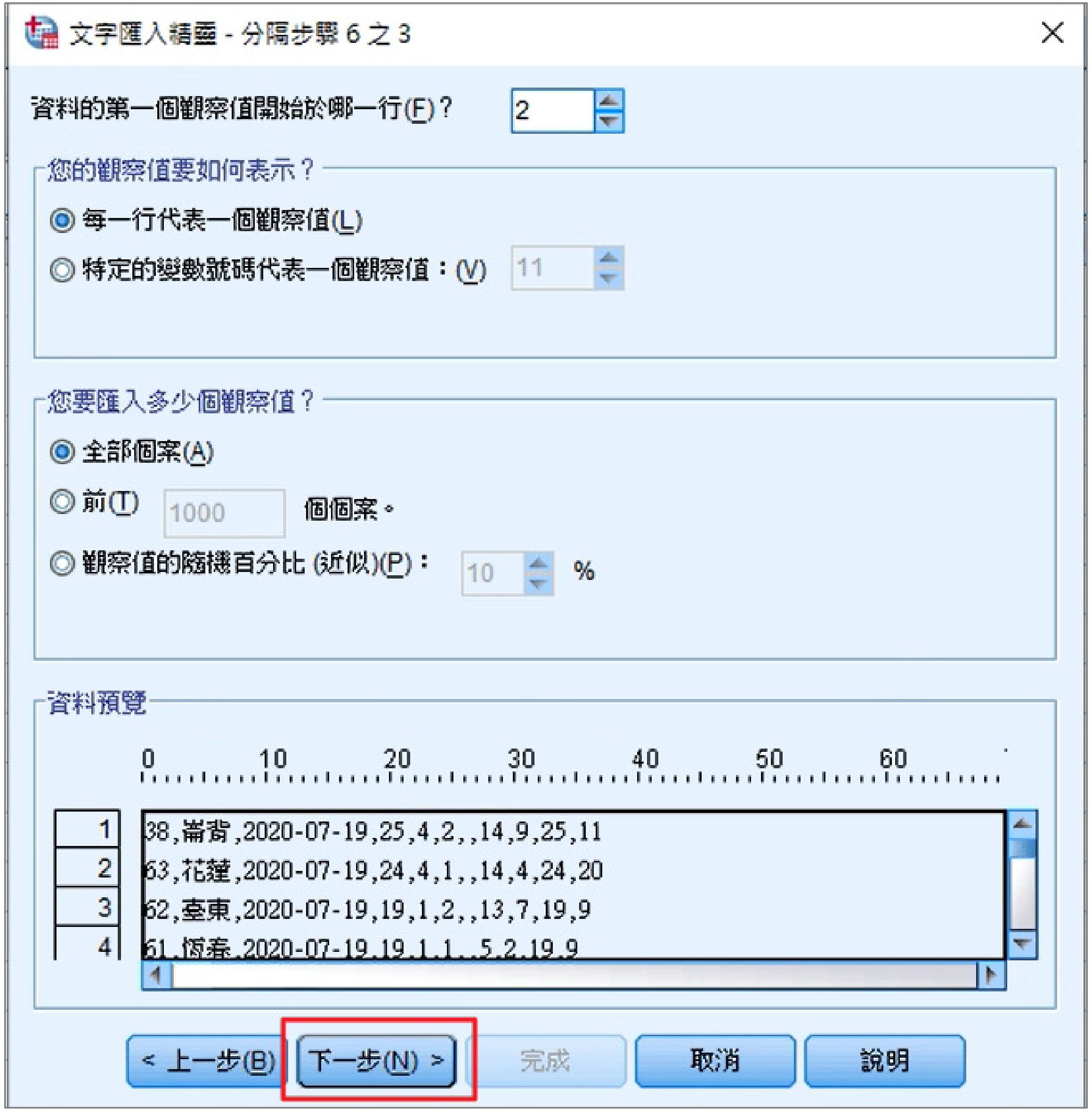
- 下載：https://github.com/rwepa/DataDemo/blob/master/aqx_p_434_20200720134939.csv
- 開啟資料文件 \ 檔案類型： 文字...




文字匯入精靈 - 步驟 6 之 1
628 840 1 81 28.5
630 2400 0 73 40.33
632 10200 0 83 31.08
633 870 0 93 31.17
var1
var2
var3
歡迎進入文字匯入精靈！
此精靈將協助您讀取文字檔的資料，並指定變數的相關資訊。
您的文字檔符合預先定義的格式嗎？
是(Y)
瀏覽(B)...
否(O)
文字檔： C:\Users\162494\Downloads\aqx_p_434_20200720134939.csv
0 10 20 30 40 50 60
1 SiteId,SiteName,MonitorDate,AQI,SO2SubIndex,COSubIndex,O3SubIndex,PM1
2 38,崙背,2020-07-19,25,4,2,,14,9,25,11
3 63,花蓮,2020-07-19,24,4,1,,14,4,24,20
4 62,臺東,2020-07-19,19,1,2,,13,7,19,9
5 61,恆春,2020-07-19,19,1,1,,5,2,19,9
6 60,潮州,2020-07-19,31,4,1,,14,5,31,21
7 59,屏東,2020-07-19,37,3,3,,16,13,37,22
< 上一步(B)
下一步(N) >
完成
取消
說明

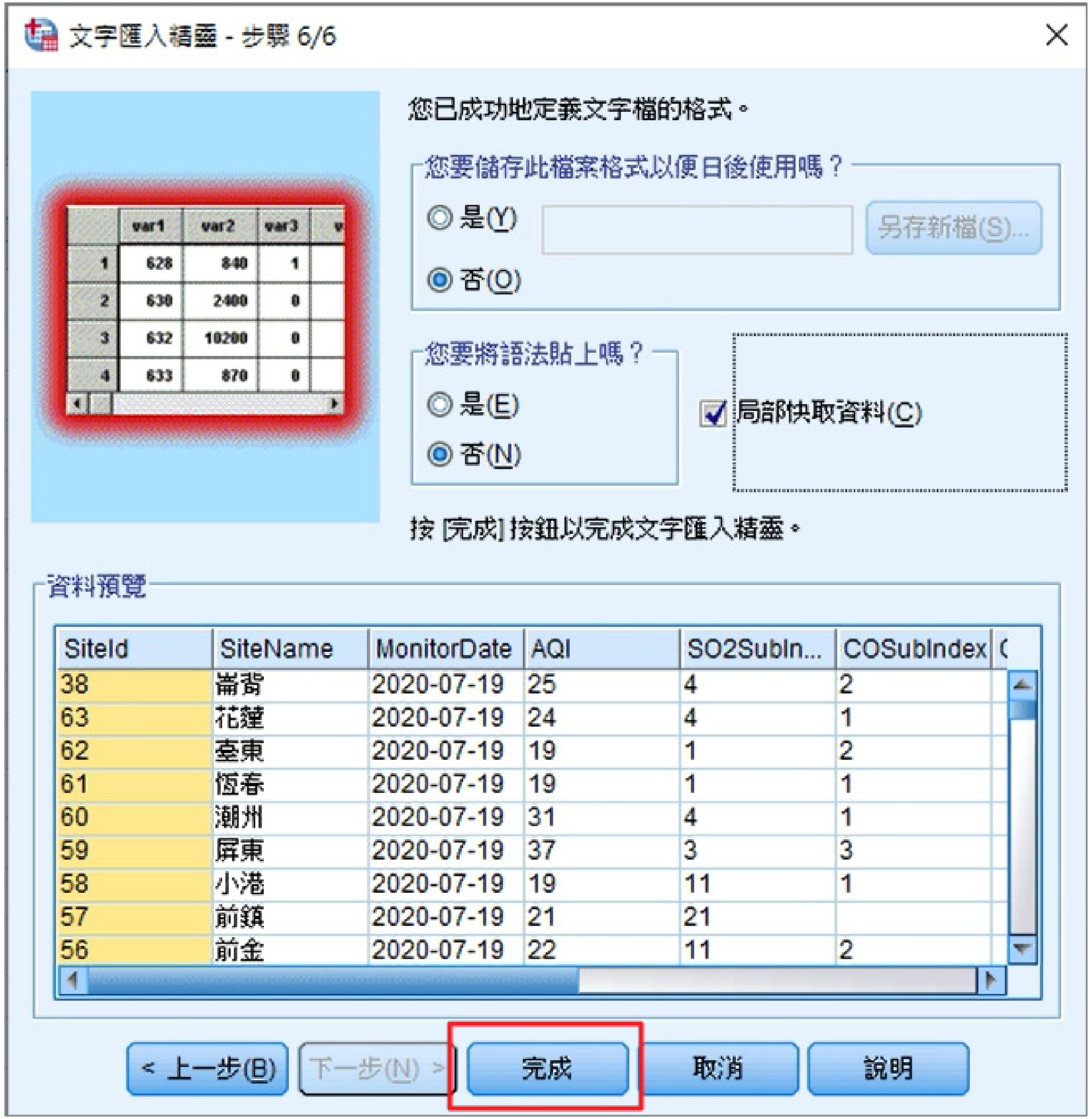
變數名稱位置


文字匯入精靈 - 步驟 2/6
您的變數如何排列？
分隔(D)
- 變數會以特殊字元分隔（如逗點、Tab 鍵）。
固定寬度(F)
- 變數會在固定寬度的欄中對齊。
變數名稱包含在檔案的最上層嗎？
是(Y)
否(O)
文字檔： C:\Users\162494\Downloads\aqx_p_434_20200720134939.csv
0 10 20 30 40 50 60
1 38,崙背,2020-07-19,25,4,2,,14,9,25,11
2 63,花蓮,2020-07-19,24,4,1,,14,4,24,20
3 62,臺東,2020-07-19,19,1,2,,13,7,19,9
4 61,恆春,2020-07-19,19,1,1,,5,2,19,9
5 60,潮州,2020-07-19,31,4,1,,14,5,31,21
< 上一步(B)
下一步(N) >
完成
取消
說明


文字匯入精靈 - 分隔步驟 6 之 3
資料的第一個觀察值開始於哪一行(F)？
2
您的觀察值要如何表示？
每一行代表一個觀察值(L)
特定的變數號碼代表一個觀察值：(V)
11
您要匯入多少個觀察值？
全部個案(A)
前(T)
1000
個個案。
觀察值的隨機百分比 (近似)(P)：
10
%
資料預覽
0
10
20
30
40
50
60
1
2
3
4
38,崙背,2020-07-19,25,4,2,,14,9,25,11
63,花蓮,2020-07-19,24,4,1,,14,4,24,20
62,臺東,2020-07-19,19,1,2,,13,7,19,9
61,恆春,2020-07-19,19,1,1,,5,2,19,9
< 上一步(B)
下一步(N) >
完成
取消
說明

文字匯入精靈 - 分隔步驟 4/6
變數間要顯示何種定界字元？
Tab
空格(S)
逗點(C)
分號(E)
其他(R)：
文字修飾詞為何？
無(O)
單一報價(Q)
雙重報價(D)
其他(H)：
資料預覽
SiteId SiteName MonitorDate AQI SO2SubIn... COSubIndex
38 崙背 2020-07-19 25 4 2
63 花蓮 2020-07-19 24 4 1
62 臺東 2020-07-19 19 1 2
61 恆春 2020-07-19 19 1 1
60 潮州 2020-07-19 31 4 1
59 屏東 2020-07-19 37 3 3
58 小港 2020-07-19 19 11 1
57 前鎮 2020-07-19 21 21
56 前金 2020-07-19 22 11 2
< 上一步(B)
下一步(N) >
完成
取消
說明

文字匯入精靈 - 步驟 5/6
資料預覽中所選的變數規格
透過前 200 個記錄中呈現的值決定資料格式。
如果欄包含前 200 個記錄中的多種資料類型，則變數類型設定為字串。
字串變數的長度（字元數）由前 200 個記錄中呈現的最長的值決定。如果後續的記錄具有更長的值，則會被截斷。
變數名稱(V)：
MonitorDate
原始名稱：
MonitorDate
資料格式(D)：
日期/時間
yyyyddd
yy/mm/dd
yyyy/mm/dd
q Q yy
資料預覽
SiteId
SiteName
MonitorDate
AQI
SO2SubIn...
COSubIndex
38
崙背
2020-07-19
25
4
2
63
花蓮
2020-07-19
24
4
1
62
臺東
2020-07-19
19
1
2
< 上一步(B)
下一步(N) >
完成
取消
說明


文字匯入精靈 - 步驟 6/6
您已成功地定義文字檔的格式。
您要儲存此檔案格式以便日後使用嗎？
是(Y)
另存新檔(S)...
否(O)
您要將語法貼上嗎？
是(E)
否(N)
局部快取資料(C)
按 [完成] 按鈕以完成文字匯入精靈。
資料預覽
SiteId SiteName MonitorDate AQI SO2SubIn... COSubIndex
38 崙背 2020-07-19 25 4 2
63 花蓮 2020-07-19 24 4 1
62 臺東 2020-07-19 19 1 2
61 恆春 2020-07-19 19 1 1
60 潮州 2020-07-19 31 4 1
59 屏東 2020-07-19 37 3 3
58 小港 2020-07-19 19 11 1
57 前鎮 2020-07-19 21 21
56 前金 2020-07-19 22 11 2
< 上一步(B)
下一步(N) >
完成
取消
說明

資料視圖

	SiteId	SiteName	MonitorDate	AQI	SO2SubInd...	COSubIndex	O3SubIndex	PM10SubIndex	NO2SubInd...	O38SubIndex	PM25SubIndex
1	38	崙背	2020/07/19	25	4	2	.	14	9	25	11
2	63	花蓮	2020/07/19	24	4	1	.	14	4	24	20
3	62	臺東	2020/07/19	19	1	2	.	13	7	19	9
4	61	恆春	2020/07/19	19	1	1	.	5	2	19	9
5	60	潮州	2020/07/19	31	4	1	.	14	5	31	21
6	59	屏東	2020/07/19	37	3	3	.	16	13	37	22
7	58	小港	2020/07/19	19	11	1	.	9	10	19	17
8	57	前鎮	2020/07/19	21	21	.	.	5	12	.	11
9	56	前金	2020/07/19	22	11	2	.	5	15	22	12
10	54	左營	2020/07/19	23	9	5	.	9	22	23	15
11	53	楠梓	2020/07/19	27	9	3	.	14	23	27	23
12	52	林園	2020/07/19	25	4	1	.	11	5	25	19
13	51	大寮	2020/07/19	29	9	2	.	12	8	29	22
14	50	鳳山	2020/07/19	22	9	3	.	13	11	.	22
15	49	仁武	2020/07/19	27	11	2	.	14	14	27	27
16	48	橋頭	2020/07/19	27	7	3	.	10	19	27	22
17	47	美濃	2020/07/19	31	3	1	.	10	4	31	21
18	32	西屯	2020/07/19	33	3	2	.	13	11	33	24
19	33	彰化	2020/07/19	30	9	2	.	12	8	30	22

資料視圖 變數視圖

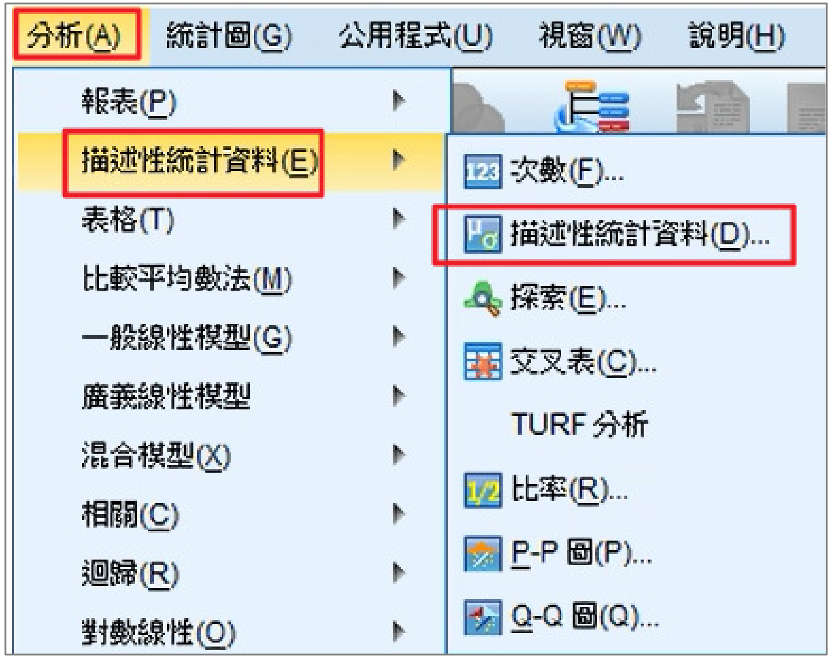
變數視圖

	名稱	類型	寬度	小數	標籤	數值	遺漏	直欄	對齊	測量	角色
1	SiteId	數值型	2	0		無	無	8	右	尺度	輸入
2	SiteName	字串	9	0		無	無	9	左	名義(N)	輸入
3	MonitorDate	日期	10	0		無	無	10	右	尺度	輸入
4	AQI	數值型	2	0		無	無	8	右	尺度	輸入
5	SO2SubIndex	數值型	2	0		無	無	13	右	名義(N)	輸入
6	COSubIndex	數值型	2	0		無	無	8	右	尺度	輸入
7	O3SubIndex	數值型	1	0		無	無	8	右	名義(N)	輸入
8	PM10SubIndex	數值型	2	0		無	無	8	右	尺度	輸入
9	NO2SubIndex	數值型	2	0		無	無	8	右	尺度	輸入
10	O38SubIndex	數值型	2	0		無	無	8	右	尺度	輸入
11	PM25SubIndex	數值型	2	0		無	無	8	右	尺度	輸入
12											
13											

資料視圖 變數視圖

範例：全台平均AQI假設檢定

- 下載：https://github.com/rwepa/DataDemo/blob/master/aqx_p_434_20200720134939.csv
- 分析 \ 描述性統計資料 \ 描述性統計資料
- 部分變數有遺漏值，AQI平均數 33.19



描述性統計資料

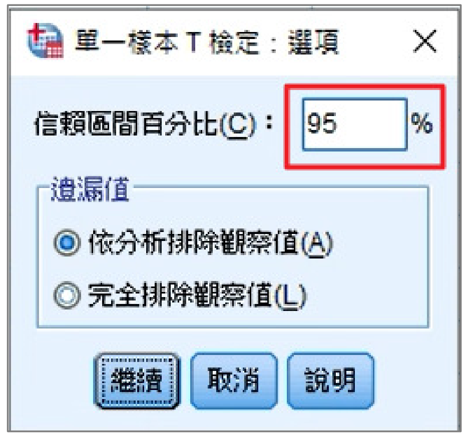
[資料集1]

描述性統計資料

	N	最小值	最大值	平均數	標準偏差
SiteId	1000	1	84	39.65	23.113
MonitorDate	1000	2020/07/07	2020/07/19	2020/07/13	3 17:46:26.3...
AQI	1000	8	159	33.19	15.949
SO2SubIndex	1000	1	43	5.65	3.444
COSubIndex	921	0	28	3.30	3.420
O3SubIndex	0				
PM10SubIndex	999	5	40	15.92	5.355
NO2SubIndex	1000	1	54	14.10	7.484
O38SubIndex	857	12	159	33.88	16.327
PM25SubIndex	987	3	80	23.81	10.136
有效的 N (listwise)	0				

範例：全台平均AQI假設檢定(續)

- H_0: $\mu_{AQI} = 33$
- H_1: $\mu_{AQI} \neq 33$



範例：全台平均AQI假設檢定(續)

- p值=0.708, $0.708 > 0.05$, 因此示拒絕 H_0, 即全台平均AQI大將等於33。
- 結果是雙尾檢定，如果是單尾檢定，須計算顯著性除以 2 ，再與 α值比較。

T 檢定

單一樣本統計資料

	N	平均數	標準偏差	標準錯誤平均值
AQI	1000	33.19	15.949	.504

單一樣本檢定

	檢定值 = 33					
	T	df	顯著性（雙尾）	平均差異	95% 差異數的信賴區間 下限	95% 差異數的信賴區間 上限
AQI	.375	999	.708	.189	-.80	1.18

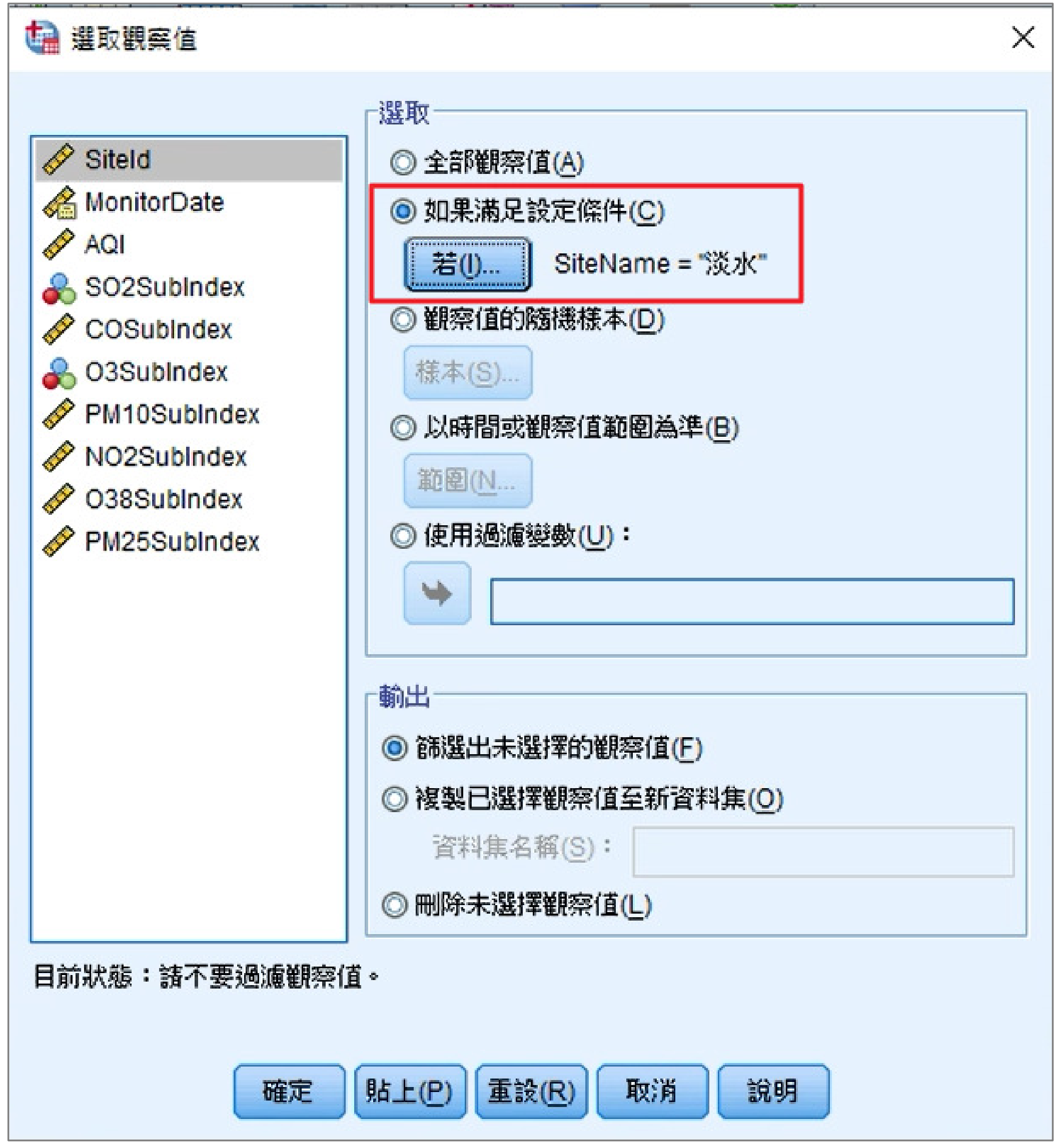
TRY:
檢定值改為35結果?

範例：淡水平均AQI假設檢定

- H_0: $\mu_{淡水_{AQI}} = 33$
- H_1: $\mu_{淡水_{AQI}} \neq 33$
- 思考題：如何篩選出「淡水」資料並進行單一樣本T檢定？

選擇觀測值

- 資料 \ 選擇觀測值
- 如果滿足設定條件 \ 若



選擇觀測值：If

- 輸入：
  SiteName = "淡水"

選擇觀測值輸出結果

- 最右側新增變數 filter_$。
- 不是淡水資料最左側加上「斜線」表示隱藏，不是刪除。

	SiteId	SiteName	MonitorDate	AQI	SO2SubIndex	COSubIndex	O3SubIndex	PM10SubIndex	NO2SubInd...	O38SubIndex	PM25SubIndex	filter_$
36	4	新店	2020/07/19	54	6	3	.	19	16	54	36	0
37	5	土城	2020/07/19	58	6	5	.	19	20	58	36	0
38	6	板橋	2020/07/19	44	7	7	.	19	30	44	35	0
39	7	新莊	2020/07/19	48	40	6	.	24	25	48	43	0
40	8	菜寮	2020/07/19	40	3	7	.	22	32	38	40	0
41	9	林口	2020/07/19	44	6	3	.	19	29	44	34	0
42	10	淡水	2020/07/19	36	3	5	.	18	17	36	35	1
43	16	大同	2020/07/19	45	6	18	.	24	44	.	45	0
44	11	士林	2020/07/19	48	4	6	.	23	23	48	39	0
45	12	中山	2020/07/19	47	4	8	.	21	42	47	34	0
46	13	萬華	2020/07/19	49	6	8	.	20	37	49	36	0
47	14	古亭	2020/07/19	61	6	6	.	19	30	61	36	0
48	15	松山	2020/07/19	48	6	7	.	20	37	48	39	0
49	20	平鎮	2020/07/19	38	4	2	.	19	12	38	32	0
50	21	龍潭	2020/07/19	41	6	2	.	19	12	41	31	0
51	22	湖口	2020/07/19	32	6	2	.	11	21	32	29	0
52	23	竹東	2020/07/19	43	43	2	.	17	7	37	31	0
53	24	新竹	2020/07/19	31	7	3	.	17	13	29	31	0
54	25	頭份	2020/07/19	24	7	.	.	14	12	.	24	0

資料視圖 變數視圖

3.SPSS-獨立樣本T檢定

獨立樣本T檢定

- 獨立樣本T檢定，適用於對兩樣本平均數的檢定，目的是比較變異數相同的兩個母體之間平均數的差異，或比較來自同一母體之兩個樣本之平均數的差異。其檢定方式與變異數是否相同，可分為兩種：
  - 兩獨立小樣本均數檢定（變異數相同）
  - 兩獨立小樣本均數檢定（變異數不同）
- 雙尾檢定
- $H_0: \mu_1 - \mu_2 = 0$，$H_1: \mu_1 - \mu_2 \neq 0$
- 上式可以化簡為:
  $H_0: \mu_1 = \mu_2$，$H_1: \mu_1 \neq \mu_2$

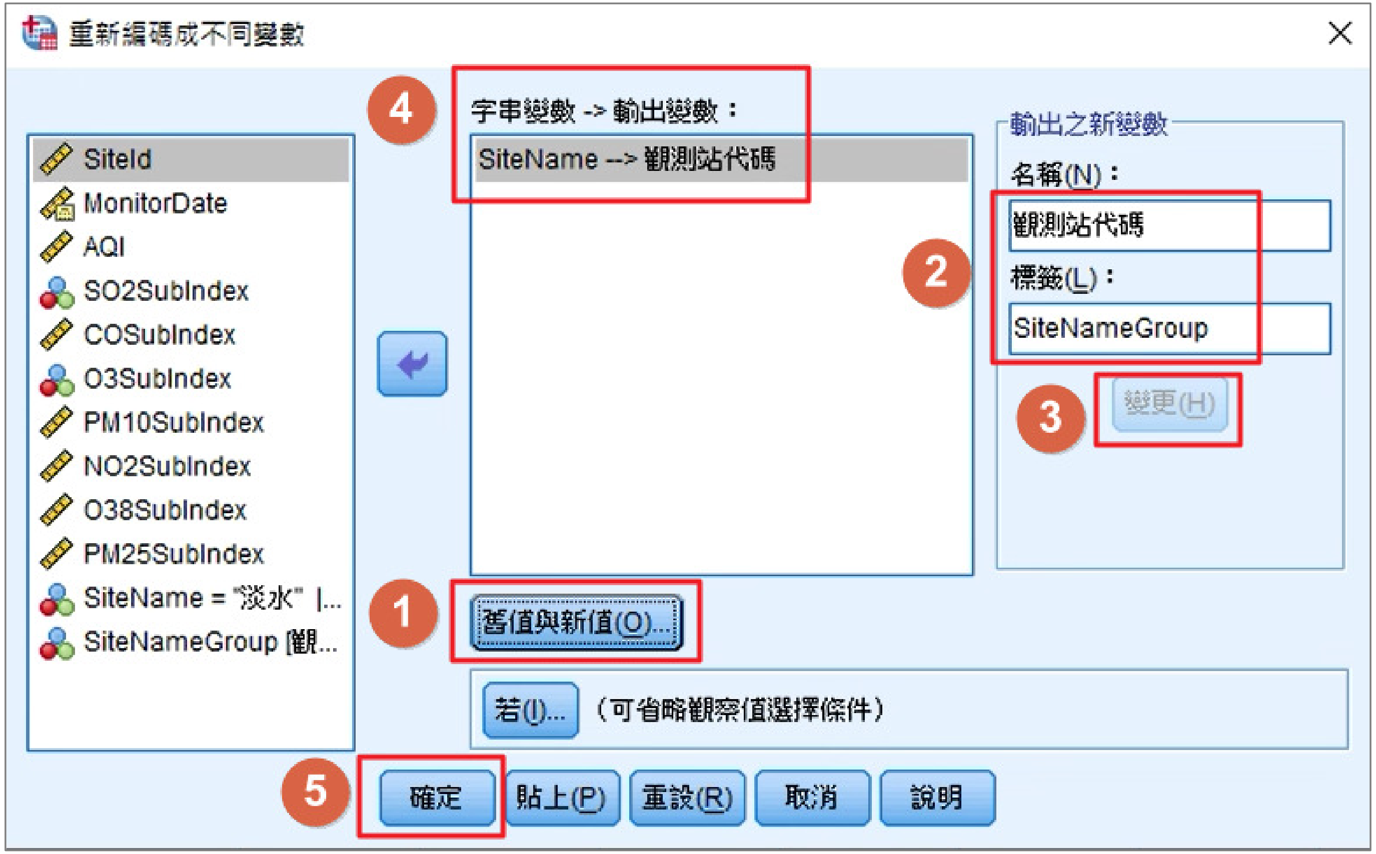
選擇觀測值

- 資料 \ 選擇觀測值 \ 選取 \ 如果滿足設定條件 \ 若
- 輸入：SiteName = "淡水" | SiteName = "士林“
- 完成後，資料集顯示 SiteName 為淡水與士林，其餘表示隱藏。

	SiteId	SiteName	MonitorDate	AQI	SO2SubIndex	COSubIndex	O3SubIndex	PM10SubIndex	NO2SubInd...	O38SubIndex	PM25SubIndex	filter_$
19	10	淡水	2020/07/10	30	3	2	.	16	14	30	24	1
20	11	士林	2020/07/10	38	6	3	.	24	16	38	29	1
21	10	淡水	2020/07/09	35	4	2	.	17	18	35	21	1
22	11	士林	2020/07/09	40	4	3	.	24	18	40	29	1
23	11	士林	2020/07/08	39	4	5	.	23	22	39	28	1
24	10	淡水	2020/07/08	27	4	5	.	17	19	27	21	1
25	11	士林	2020/07/07	35	6	5	.	20	22	35	32	1
26	10	淡水	2020/07/07	29	4	5	.	19	23	29	26	1
27	38	崙背	2020/07/19	25	4	2	.	14	9	25	11	0
28	63	花蓮	2020/07/19	24	4	1	.	14	4	24	20	0
29	62	臺東	2020/07/19	19	1	2	.	13	7	19	9	0
30	61	恆春	2020/07/19	19	1	1	.	5	2	19	9	0
31	60	潮州	2020/07/19	31	4	1	.	14	5	31	21	0
32	59	屏東	2020/07/19	37	3	3	.	16	13	37	22	0

重新編碼成不同變數

- 轉換 \ 重新編碼成不同變數【參考:es_2024.02.27_descriptive_statistics.pdf, P.60】



舊值與新值

- 依序新增：
  - '淡水' --> 1
  - '士林' --> 2

獨立樣本T檢定


獨立樣本 T 檢定
檢定變數(T)：
選項(O)...
SiteId
SiteName
MonitorDate
SO2SubIndex
COSubIndex
O3SubIndex
PM10SubIndex
NO2SubIndex
O38SubIndex
PM25SubIndex
AQI
分組變數(G)：
觀測站代碼(1 2)
定義組別(D)...
確定
貼上(P)
重設(R)
取消
說明

獨立樣本T檢定輸出結果

T 檢定

1

群組統計資料

	SiteNameGroup	N	平均數	標準偏差	標準錯誤平均值
AQI	1	13	35.23	11.684	3.241
	2	13	49.69	21.823	6.053

2

3

獨立樣本檢定

		Levene 的變異數相等測試		針對平均值是否相等的 t 測試						
									95% 差異數的信賴區間	
		F	顯著性	T	df	顯著性（雙尾）	平均差異	標準誤差	下限	上限
AQI	採用相等變異數	4.550	.043	-2.106	24	.046	-14.462	6.866	-28.631	-.292
	不採用相等變異數			-2.106	18.358	.049	-14.462	6.866	-28.865	-.058

步驟1 變異數檢定：

H_0: σ^2 淡水 $=\sigma^2$ 士林

H_1: σ^2 淡水 $\neq \sigma^2$ 士林

因為$0.043 < 0.05$，接受H_1，變異數不相等。

步驟2 t檢定：

H_0: μ 淡水 $= \mu$ 士林

H_1: μ 淡水 $\neq \mu$ 士林

採用p值=0.049, $0.049 < 0.05$，

因此淡水與士林平均AQI不相等。

4.SPSS-成對樣本T檢定

成對樣本T檢定

- 前面「獨立樣本T檢定」，其兩組受測樣本間為獨立，並無任何關聯。如：甲乙班、男女生、兩不同年度、不同地區等。
- 如果兩組受測樣本為相依：
    - 若同組人員，受訓後的打字速度是否高於受訓前。
    - 同一部車，左右使用不同廠牌輪胎，經過一段時間後檢查其磨損程度。
    - 同一地區，使用不同自然淨化工法，研究水質淨化之差異。
    - 兩組受測樣本間為相依（同一個人、同一部車與同一地區），此時可以使用成對樣本T檢定。

SPSS-成對樣本T檢定

- 分析 \ 比較平均數法 \ 成對樣本T檢定
- 輸入資料時，考量輸入 2 行不同變數較方便。

謝謝您的聆聽

Q & A

李明昌
alan9956@gmail.com
http://rwepa.blogspot.tw/