



R 「survey」 套件簡介-- 以 104 年網路沉迷研究為例

王俞才

一、R 是什麼？

R 軟體是基於 1976 年由 John Chambers 在 AT&T 的貝爾實驗室開發的 S 語言所建構而成的。在 1992 年，現職紐西蘭奧克蘭大學的兩位統計系教授 Ross Ihaka 和 Robert Gentleman(兩位教授的名字字首都是 R, 因此 R 也成為了這個軟體的名稱) 為了教授初階的統計課程，開始發展 R 軟體，而現階段則是由 R Development Core Team 負責開發。此外，S 語言的發明者 John Chambers 也是 R Development Core Team 的成員之一，因此不難了解 S 語言與 R 軟體之間有著密不可分的關係。

目前的 R 軟體為開放原始碼的自由軟體(Free and Open-Source Software)，提供了包含資料處理、統計計算及繪圖等完整且強大的功能，所有人都可以免費取得 R 軟體的原始碼並加以修改或擴充其功能。此外，R 軟體也提供了多種執行檔(包含 Windows、MacOS 和 Linux 等)方便使用者在不同的平台上執行。

二、如何學習 R？

對於 R 軟體的初學者而言，最大的問題不外乎就是「我該使用哪個函數或套件？」。其實 R 是相當容易學習且使用上也非常簡單的。如圖 1 所示，R 程式撰寫流程基本上依序為確認目標、資料準備、資料前處理，接著尋找合適的套件、函數或自行撰寫程式，最後再將結果彙整。使用者可以藉由 R 軟體網站中 CRAN(Comprehensive R Archive Network)Task Views 尋找合適的套件，也可以使用網站中的 Google 站內搜尋尋找需要的套件或函數。如同本篇文章要介紹的「survey」套件，就可以在 CRAN Task Views 底下的 Social Sciences 分類中找到。此外，若在 R 軟體中使用「??關鍵字」也可以幫助使用者找到相關的函數，而輸入「?函數名稱」或「help(函數名稱)」則可搜尋該函數的定義或範例。

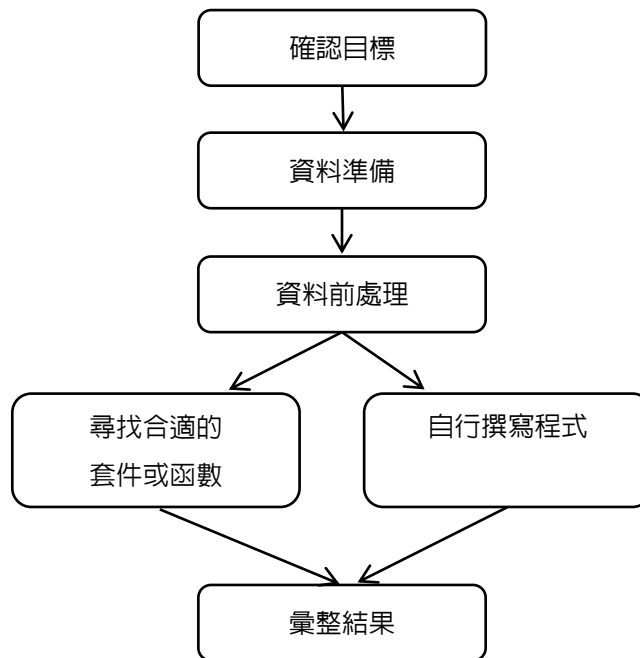


圖 1. 程式撰寫流程圖

三、範例資料

有鑑於國人上網越來越容易、上網時間越來越長，網路使用風險似乎也有日益嚴重趨勢，國家發展委員會特於「104 年數位機會調查」委外服務案中規畫「網路沉迷研究」子案，主要是為了評估民衆因上網所造成的人際、健康與時間管理等等問題，期望透過深度討論國內網路沉迷的現象，作為政府日後相關政策的參考向。該調查是以臺灣及金馬地區的 22 個縣市為調查範圍，並以年滿 12 歲且有上網經驗的民衆為研究對象進行電話訪問。該調查已於 105 年 2 月 15 日釋出，有興趣的讀者可至 SRDA 學術調查研究資料庫網站申請下載。

本文做為範例的資料檔取自「104 年網路沉迷研究」，經過簡單的資料前處理(如：刪除遺失值、重新編碼...等)後，並假設其為真實「母體」資料。隨後再分別進行三種抽樣設計：簡單隨機抽樣、分層隨機抽樣及群集抽樣作為本篇文章的三種範例資料¹。首先我們就利用一些基本的函數來對母體資料集(pop)做簡單的探索性分析。

¹ 若讀者需要該範例資料檔或程式碼，歡迎來信索取：wuysi@gate.sinica.edu.tw

```

> head(pop) #顯示前 6 筆資料
  gender    age unlimited cluster minute  area
1   女 40-49 歲      是      12    120 南部地區
2   女 40-49 歲     不是       1    300 北部地區
3   女 12-19 歲     不是      12    120 南部地區
4   女 40-49 歲     不是       1     50 北部地區
5   女 40-49 歲      是       8     60 北部地區
6   女 40-49 歲     不是       8    480 北部地區

> tail(pop) #顯示尾 6 筆資料
  gender    age unlimited cluster minute  area
1390  男 60 歲以上      是       1     30 北部地區
1391  男 60 歲以上      是       3    180 北部地區
1392  男 60 歲以上      是      12    120 東部地區
1393  男 60 歲以上     不是       1    180 中部地區
1394  男 60 歲以上     不是       3     90 中部地區
1395  男 60 歲以上     不是       3     30 北部地區

> dim(pop) #取得資料集維度
[1] 1395    6

> names(pop) #取得變數名稱
[1] "gender"  "age"    "unlimited"  "cluster"  "minute"  "area"
    
```

從執行結果可以發現，母體資料集 pop 共有 1395 個觀測值及 6 個變數，且透過 names() 函數可以得知 6 個變數的變數名稱：

- gender：請問您的性別是？
- age：請問您大約幾歲？
- unlimited：請問您採用的是吃到飽方案嗎？
- cluster：群集
- minute：請問您平均每天上網多久(分鐘)？
- area：請問您目前住在哪一個地區？

接著我們可以利用 `summary()` 函數來查看各個變數分佈的情形。對於資料集有初步的了解後，我們就可以接續進行三種不同的抽樣設計。

> summary(pop)			#查看各變數分佈的情形			
gender	age	unlimited	cluster	minute	area	
男:678	12-19 歲 :219	是 :744	Min. :1.000	Min. :2.0	北部地區:647	
女:717	20-29 歲 :275	不是:651	1st Qu.: 1.000	1st Qu. : 90.0	中部地區:350	
	30-39 歲 :337		Median :5.000	Median : 180.0	南部地區:351	
	40-49 歲 :349		Mean :5.351	Mean : 209.7	東部地區: 35	
	50-59 歲 :174		3rd Qu.: 9.000	3rd Qu. : 300.0	離島地區: 12	
	60 歲以上: 41		Max. :14.000	Max. :1200.0		

1. 簡單隨機抽樣：

這部分我們將採用 `sample()` 函數來進行簡單隨機抽樣，該函數也是在 R 軟體中最基本的抽樣函數，其基本格式為：

`sample(x, size, replace=FALSE, prob=NULL)`

其中 `x` 為一正整數或含有一個或多個元素的向量；`size` 為一正整數，代表想要被抽出的樣本數；`replace` 詢問是否為抽出可放回的抽樣；`prob` 用於設定各樣本被抽出的機率。

接著我們就利用 `sample` 函數來執行簡單隨機抽樣，並設定抽出的樣本數為 200。同時為了讓讀者可以跟我得到相同的抽樣結果，在執行任何抽樣過程之前，我先利用了 `set.seed()` 函數來建立隨機起始點。

> set.seed(1)			#建立隨機起始點			
> srs<-pop[sample(1:nrow(pop), 200),]			#執行簡單隨機抽樣			
> head(srs)			#顯示前 6 筆資料			
gender	age	unlimited	cluster	minute	area	
371	男 40-49 歲	是	9	390	北部地區	
519	男 12-19 歲	不是	5	240	北部地區	
798	女 20-29 歲	是	12	300	北部地區	
1265	男 20-29 歲	是	9	60	南部地區	
281	男 40-49 歲	是	1	120	北部地區	
1249	男 20-29 歲	不是	1	240	南部地區	

2. 分層隨機抽樣：

分層隨機抽樣是將母體依照某衡量標準，區分成若干個不重複的子母體，我們稱之為『層』，且層與層之間有很大的變異性，而層內的變異性較小。在區分不同層後，再從每一層中利用簡單隨機抽樣抽出所須比例的樣本數，最後將所得各層樣本合起來即為樣本。利用分層隨機抽樣可保持樣本資料與母體分佈的一致性，在分析資料時也可以減少資料不平衡的問題。

分層抽樣我們可以透過 `sampling` 套件中的 `strata()` 函數來實現，其基本格式為：

```
strata( data , stratanames=NULL , size , method=c("srswor" , "srswr" ,  
"poisson" , "systematic") , pik , description=FALSE )
```

其中 `data` 為待被抽樣的資料集；`stratanames` 為將被作為分層依據的變數名稱；`size` 用於設定各分層中將要被抽出的樣本數，該值的順序必須與該變數中各水準出現的順序一致，且必須將資料集按照該變數的水準進行升冪排列；`method` 用於選擇 4 種抽樣方法，分別為隨機抽出不放回(`srswor`)、隨機抽出放回(`srswr`)、卜松(`poisson`)及系統抽樣(`systematic`)，預設為隨機抽出不放回；`pik` 用於設定各分層被抽出的機率；`description` 參數則是詢問是否輸出包含各分層基本資訊的抽樣結果。

這邊我們將利用資料集中的變數 `area` 作為分層的依據，並設定樣本數為 200，而各層應抽出的人數則依「比例配置法」(Proportional allocation)決定。其執行過程如下：

```
> round(prop.table(table(pop$area))*200)                                #查看各層應抽出的人數
北部地區  中部地區  南部地區  東部地區  離島地區
      93      50      50      5      2

> poporder<-pop[order(pop$area) , ]                                    #將資料集依變數 area 的水準排序
> install.packages("sampling")                                         #安裝 sampling 套件
> library("sampling")                                                  #載入 sampling 套件
> set.seed(1)                                                           #建立隨機起始點
> stratID<-strata(poporder ,                                           #執行分層隨機抽樣
                 stratanames="area" ,
                 size=c(93 , 50 , 50 , 5 , 2) ,
                 method="srswor" ,
                 description= TRUE)
```

```

Stratum 1
Population total and number of selected units: 647 93
Stratum 2
Population total and number of selected units: 350 50
Stratum 3
Population total and number of selected units: 351 50
Stratum 4
Population total and number of selected units: 35 5
Stratum 5
Population total and number of selected units: 12 2
Number of strata 5
Total number of selected units 200
> head(stratID) #顯示前 6 筆資料
      area  ID_unit  Prob  Stratum
9  北部地區      9  0.1437403      1
15 北部地區     15  0.1437403      1
33 北部地區     33  0.1437403      1
40 北部地區     40  0.1437403      1
42 北部地區     42  0.1437403      1
49 北部地區     49  0.1437403      1
> strat<-getdata(poporder , stratID$ID_unit) #取得分層抽樣所得資料
> head(strat) #顯示前 6 筆資料
      gender  age  unlimited  cluster  minute  area
11  女 40-49 歲      是         3     180 北部地區
17  女 40-49 歲     不是        14     480 北部地區
35  女 40-49 歲      是         3     840 北部地區
42  女 40-49 歲      是         4     120 北部地區
44  女 40-49 歲      是         8         60 北部地區
119 女 40-49 歲      是        12     240 北部地區

```

3. 群集抽樣：

群集抽樣的方法就是將母體分成幾個群集(或部落、區域)，再從這幾個群集中抽出數個群集進行抽樣或普查。有時群集抽樣又稱部落抽樣或叢聚抽樣。在考慮使用群集抽樣時，一般會要求各群集對資料整體有較好的代表性，即群集間的變異小，而群集內的變異大。因此當群與群之間差距較大時，群集抽樣常常會出現分佈不廣或樣本代表性較差等缺點。

我們可以透過 `sampling` 套件中的 `cluster()` 函數來執行群集抽樣，其基本格式為：

```
cluster( data, clustername, size, method=c("srswor", "srswr", "poisson",  
"systematic"), pik, description=FALSE)
```

該函數的參數除了 `clustername` 與 `size` 略有差異外，其餘參數的涵義都跟 `strata()` 函數相同。`clustername`，顧名思義，指用來劃分群組的變數名稱。而 `size` 為一正整數，代表欲被抽出的群集數。

以下我們將利用資料集中的變數 `cluster2` 作為分群的依據，隨機抽出 3 個群集。其抽樣過程如下：

```
> set.seed(1) #建立隨機起始點
> cluID<-cluster(pop, #執行群集抽樣
  clustername="cluster",
  size=3,
  method="srswor",
  description=TRUE)
Number of selected clusters: 3
Population total and number of selected units 1395 230
> head(cluID) #顯示前 6 筆資料
```

	Cluster	ID_unit	Prob
1	4	529	0.2142857
2	4	690	0.2142857
3	4	78	0.2142857
4	4	936	0.2142857
5	4	596	0.2142857
6	4	799	0.2142857

² 變數 `cluster` 為作者以隨機亂數生成，僅供範例說明使用，並未考量其分佈是否恰當或樣本代表的好壞。

```
> cluster<-getdata(pop,cluID$ID_unit) #取得群集抽樣所得資料
> head(cluster) #顯示前 6 筆資料
```

	gender	age	unlimited	cluster	minute	area
529	女	30-39 歲	是	4	480	北部地區
690	女	30-39 歲	是	4	360	南部地區
78	女	40-49 歲	不是	4	60	南部地區
936	女	30-39 歲	不是	4	120	南部地區
596	女	30-39 歲	不是	4	420	中部地區
799	女	20-29 歲	不是	4	120	北部地區

四、「survey」套件

survey 套件是由現職紐西蘭奧克蘭大學生物統計學教授 Thomas Lumley 所撰寫及負責維護更新，主要的功能在處理複雜的調查資料分析。survey 套件的功能相當強大，包含了最基本的求取平均數、總和、比值到迴歸模型的建立及主成分分析，甚至是資料視覺化...等功能，對於從事有關調查資料分析工作的研究人員或學生而言有相當大的幫助。此外，survey 套件有一個專屬的網站(圖 2)，其網址為 <http://r-survey.r-forge.r-project.org/survey/>，該網站不只包含了套件內容的簡介，還有許多有關 R 軟體的基礎教學。

Survey analysis in R

This is the homepage for the `survey` package, which provides facilities in R for analyzing data from complex surveys. The current version is 3.29. A much earlier version (2.2) was published in *Journal of Statistical Software*.

An experimental package for very large surveys such as the American Community Survey can be found [here](#).

A port of a much older version of the survey package (version 3.6-8) to S-PLUS 8.0 is available from CSAN (thanks to Patrick Aboiyom at laushtefit).

Features:

- Means, totals, ratios, quantiles, contingency tables, regression models, loglinear models, survival curves, snk tests, for the whole sample and for domains.
- Variances by Taylor linearization or by replicate weights (BRR, jackknife, bootstrap, multistage bootstrap, or user-supplied).
- Multistage sampling with or without replacement.
- PPS sampling with or without replacement. Horvitz-Thompson and Yates-Grundy estimates and a range of approximations.
- Post-stratification, generalized raking, raking, GREG estimators, trimming of weights.
- Two-phase designs. Estimated weights for augmented IPW estimators.
- Graphics.
- Support for using multiply imputed data.
- Database-backed design objects for large data sets (now with replicate weights, too).
- Some support for parallel processing on multicore computers.
- Multivariate analysis: principal components, factor analysis (experimental).
- Likelihood ratio (Rao-Scott) tests for glm, Cox models, loglinear models.

The `NEWS` file gives a history of features and bug fixes.

Comparison shopping
Alex Zaslavsky keeps a comprehensive list of [survey analysis software](#) for the ASA Section on Survey Research Methods.

User-generated ratings and reviews of this package (and others) at [cran.r-project.org](#).

Using the survey package:

- Specifying a survey design
- Creating replicate weights
- Sample summary statistics
- Using supplied replicate weights
- Domain (subpopulation) estimation
- Tables of summary statistics
- Post-stratification and calibration
- Linearly PSUs
- Regression models
- Tests of association
- Stratification within PSUs
- Graphics
- Multiple imputation and ordinal logistic regression
- Database-backed survey objects
- Programming with survey objects

Technical notes and comparisons with other software
Some examples (in PDF) translated from Stata and SUDAAN examples at [UCLA Academic Technology Services](#).



圖 2. SURVEY 套件網站

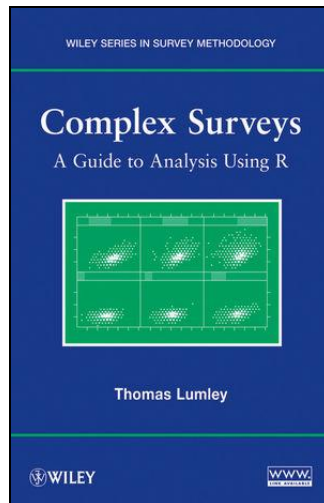


圖 3. Complex Surveys: A Guide to Analysis Using R

另外，Thomas Lumley 也撰寫了一本書叫「Complex Surveys: A Guide to Analysis Using R」(圖 3)，這本書也是以 survey 套件為基礎，讀者都可以透過網站或書籍對 survey 套件有更進一步的認識。

在使用 survey 套件時最主要有兩個步驟要執行：第一個是我們必須提供資料集名稱及抽樣設計(含抽樣方法、母體大小、權數……)的資訊以建立 survey.design 物件；第二個步驟則是以先前建立的 survey.design 物件，再利用指定的函數來執行統計分析、建立模型或繪圖。而這些指定的函數名稱都是以 svy*為開頭，例如：svymean()、svytotal()、svyratio()、svychisq()、svyglm()等。

在 survey 套件中 svydesign()即是用來建立 survey.design 物件的函數，它包含了許多可以設定的參數，在此我們僅說明最重要的幾個參數：

ids：為資料集中的一個變數名稱。當採用群集抽樣時，需給定每個樣本所屬的群集。~1 或~0 表示沒有分群。

strata：作為分層依據的變數名稱。NULL 表示沒有分層。

fpc：為一向量，給定每個樣本所屬層別的樣本數大小。

data：為一包含原始數據的資料集。

下面我們將以先前利用簡單隨機抽樣、分層隨機抽樣及群集抽樣等三種抽樣方式所抽出的樣本對 survey 套件及其函數做簡單的介紹，推估國人每天平均的上網時間。

1. 簡單隨機抽樣：

在說明 survey 套件的使用方法之前，我們先以簡單隨機抽樣樣本中的變數 minute，並用土法煉鋼的方式來計算國人每天上網時間的平均數及其變異數的估計量，其公式如下：

$$\hat{\mu}_x = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{200} (390 + 240 + 300 + \dots + 180) = 199.75 \quad \dots\dots\dots(a)$$

$$\widehat{Var}[\hat{\mu}_x] = \frac{N-n}{N} \times \frac{\widehat{Var}[x]}{n} = \frac{1395-200}{1395} \times \frac{30980.59}{200} = 132.6946 \quad \dots\dots\dots(b)$$

其中，

$$\widehat{Var}[x] = \sum_{i=1}^n \frac{(X_i - \bar{x})^2}{n-1} = \frac{[(390 - 199.75)^2 + \dots + (180 - 199.75)^2]}{200 - 1} = 30980.59$$

由上面的計算結果可以得知國人每天上網的平均時間為 199.75 分鐘，且變異數的估計量為 132.69 分鐘。接著我們就使用 survey 套件來實現此一過程，第一個步驟為建立 survey.design 物件：

```
> install.packages("survey") #安裝 survey 套件
> library("survey") #載入 survey 套件
> des.srs<-svydesign( ids=~1 , #建立 survey.design 物件
                    strsta=NULL ,
                    fpc=rep(1395 , 200) ,
                    data=srs)
> summary(des.srs) #查看 des.srs 的摘要
Independent Sampling design
svydesign(ids = ~1 , strsta = NULL , fpc = rep(1395 , 200) , data = srs)
Probabilities:
  Min. 1st Qu. Median Mean 3rd Qu. Max.
0.1434 0.1434 0.1434 0.1434 0.1434 0.1434
Population size (PSUs): 1395
Data variables:
[1] "gender" "age" "unlimited" "cluster" "minute" "area"
```

透過 `summary()` 函數我們可以得知 `survey.design` 物件(`des.srs`)為 Independent Sampling design，代表這個抽樣設計為簡單隨機抽樣，而且當我們透過 `fpc` 參數指定母體大小時，這個簡單隨機抽樣會被認定是抽出不放回的。一旦我們完成建立 `survey.design` 物件之後，我們就可以透過 `survey` 套件中許多 `svy*` 的函數來做資料的統計分析。

```
> svymean(~minute, design=des.srs) #求取上網時間的平均數
```

	mean	SE
minute	199.75	11.519

透過 `svymean()` 函數，我們可以得到國人每天上網時間的平均數、以及平均數的標準差分別為 199.75 分鐘和 11.519 分鐘。這個結果與先前公式(a)與公式(b) ($\sqrt{132.694} = 11.519$) 所得到結果一樣。`survey` 套件中大部分的 `svy*` 函數都跟 `svymean` 一樣，以方程式(`~variable name`)的寫法去選擇需要被分析的變數。他跟你熟悉的 `lm()` 函數的寫法有點類似，但他的變數通常只放在「~」的一邊。接著我們可以以中央極限定理(CLT)為基礎，用手或是利用 `confint()` 函數來計算平均數的信賴區間。

```
> confint(svymean(~minute, design=des.srs)) #求取平均數 95%的信賴區間
```

	2.5 %	97.5 %
minute	177.1726	222.3274

假設現在我們想了解在不同性別下國人每天的平均上網時間，則我們可以利用 `svyby()` 函數來幫我們實現：

```
> svyby(~minute, by=~gender, design = des.srs, FUN=svymean) #求取不同性別下的平均上網時間
```

	gender	minute	se
男	男	181.4216	13.54838
女	女	218.8265	18.64260

當我們只想知道男生(或特定族群)的特性時，那麼我們可以利用 `subset()` 函數來實現：

```
> svymean(~minute, design=subset(des.srs, gender=="男")) #求取男生的平均上網時間
```

	mean	SE
minute	181.42	13.548

2. 分層隨機抽樣：

相較於簡單隨機抽樣，建立分層隨機抽樣的 survey.design 物件是需要多一些程序的。首先我們必須從母體資料集(pop)找出每個分層的母體樣本數，並將它建立成一個新的物件：

```
> class.table<-as.data.frame(table(pop$area))           #找出每個分層的樣本數，並建立資料集
> class.table
      Var1 Freq
1  北部地區  647
2  中部地區  350
3  南部地區  351
4  東部地區   35
5  離島地區   12
```

接著我們必須按照分層隨機抽樣樣本(strat)中每個樣本所屬的分層將 class.table 中的各分層人數匯入，匯入的方法可以利用 merge()函數來執行。merge()函數可以簡單且快速地将兩個擁有相同變數名稱的資料作合併。但需注意的是，由於變數 area 已套用中文的選項說明，因此資料合併後的排列順序會依照各水準字首的筆劃多寡，按遞增方式呈現(「中」部地區<「北」部地區<「東」部地區<「南」部地區<「離」島地區)。

```
> names(class.table)[1]<-"area"                       #將 class.table 的第一個變數名稱更改為 area
> strat<-merge(strat, class.table, by="area")         #將 strat 與 class.table 資料集做合併
> head(strat)                                       #顯示前 6 筆資料
      area gender  age  unlimited  cluster  minute  Freq
1  中部地區  女  40-49 歲    不是      5     240  350
2  中部地區  女  40-49 歲    不是      9      60  350
3  中部地區  女  40-49 歲     是       1      60  350
4  中部地區  女  40-49 歲    不是      1     180  350
5  中部地區  女  40-49 歲    不是      1     240  350
6  中部地區  女  12-19 歲     是       1      60  350
```

此處的分層隨機抽樣樣本多了一個變數 Freq，於是我們就可以開始建立屬於分層隨機抽樣的 survey.design 物件，他跟簡單隨機抽樣的 survey.design 物件只有兩個地方略有差異：一個是在參數 strata 需要指定分層依據的變數名稱，也就是變數 area；另一個則是參數 fpc 需利用變數 Freq 來給定每個樣本所屬層別的樣本數大小。

```

> des.strat<-svydesign(ids=~1, strata=~area, fpc=~Freq, data=strat) #建立 survey.design
物件
> summary(des.strat) #查看 des.strat 的摘要
Stratified Independent Sampling design
svydesign(ids = ~1, strata = ~area, fpc = ~Freq, data = strat)
Probabilities:
  Min. 1st Qu. Median Mean 3rd Qu. Max.
0.1425 0.1428 0.1429 0.1434 0.1437 0.1667
Stratum Sizes:
      北部地區 中部地區 南部地區 東部地區 離島地區
obs          93      50      50      5      2
design.PSU    93      50      50      5      2
actual.PSU   93      50      50      5      2
Population stratum sizes (PSUs):
中部地區 北部地區 東部地區 南部地區 離島地區
  350      647      35      351      12
Data variables:
[1] "area" "gender" "age" "unlimited" "cluster" "minute" "Freq"

```

接下來就可以跟簡單隨機抽樣一樣利用 `svy*` 函數做一連串的統計分析，這裡我們就不再贅述。

```

svymean(~minute, design=des.strat) #求取上網時間的平均數
  mean SE
minute 225 11.634

```

3. 群集抽樣：

群集抽樣樣本(cluster)是從 14 個群集當中採用一階段群集抽樣(One-Stage Cluster)方式隨機抽取 3 個群集而組成的樣本。群集抽樣的 `survey.design` 物件跟簡單隨機抽樣不同之處在於變數 `ids`，此處我們必須給定每個樣本所屬的群集，也就是變數 `cluster`。另外參數 `fpc` 需要給定母體的群集個數。

```

> des.clu<-svydesign(ids=~cluster, strata=NULL, fpc=rep(14, nrow(cluster)), data=cluster)
                                                    #建立 survey.design 物件
> summary(des.clu)
                                                    #查看 des.clu 的摘要
1 - level Cluster Sampling design
With (3) clusters.
svydesign(ids = ~cluster, strata = NULL, fpc = rep(14, nrow(cluster)),
          data = cluster)
Probabilities:
      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.2143  0.2143  0.2143  0.2143  0.2143  0.2143
Population size (PSUs): 14
Data variables:
[1] "gender"  "age"      "unlimited" "cluster"  "minute"   "area"
> svymean(~minute, design=des.clu)
                                                    #求取上網時間的平均數
      mean    SE
minute 251.65  5.1541

```

以上是針對 survey 套件所作粗淺的介紹，當然讀者也可以透過 CRAN Task Views 選擇其他合適的套件加以運用。但如同先前所言，survey 套件所提供的函數是相當完善的，且擁有專屬的網站提供解說及教學。不僅如此，假設你對於套件有任何疑問，可以隨時寫 E-mail 給套件的作者，若套件的作者發現有任何缺失，也會立即修正錯誤或擴充套件的功能。當然這就是其他統計軟體所無法取代的優點之一。

五、資料引用

國家發展委員會資訊管理處(2016)。104 年網路沈迷研究 (AE120001)【原始數據】。取自中央研究院人文社會科學研究中心調查研究專題中心學術調查研究資料庫 doi:10.6141/TW-SRDA-AE120001-1。

六、參考書籍

Thomas Lumley(2010). Complex Surveys A guide to Analysis Using R. WILEY(U.S.A)

陳景祥(2010)，”R 軟體：應用統計方法”，台北市：東華書局。

黃文、王正林(2016)，”利用 R 語言打通大數據的經脈(第二版)”，台北市：佳魁資訊。



會員申請程序說明

一、會員資格

(一) 一般會員(二年期)

1. 國內外公私立研究機構之專任研究人員；
2. 國內外公私立大專院校之專兼任教師；
3. 政府機構之專任研究人員；
4. 捐贈或授權資料予「學術調查研究資料庫」之個人或單位代表。

(二) 一般會員(一年期)

1. 國內外公私立大專院校大學部學生、碩博士班研究生；
2. 國內外公私立研究機構研究助理人員；
3. 國內外公私立大專院校研究助理人員；
4. 政府機構研究助理人員。

若不符合前述一般會員資格，但已取得博士學位者，可以個案申請方式處理。

(三) 院內會員

中央研究院專任研究人員

(四) 網路會員

不限資格之個人

二、加入方式

請至 SRDA 網頁(<https://srda.sinica.edu.tw>) 點選「加入會員」，依步驟點選同意會員約定條款同意書中各項規定，填寫申請表提出申請，至個人電子郵件信箱中收取確認信並完成電子郵件確認程序。如需加入成為一般會員或院內會員，則請將在職(在學)文件影本以郵寄、傳真、手機拍照或掃描成電子檔附加於 Email 等方式擲回本中心。

- 審核時間：收到文件後一至三個工作天。
- 諮詢專線：(02)2787-1829；傳真：(02)2787-1802；Skype：csr_209。

三、資格認證

(一)一般會員：依照申請人的在職或在學身分，期限分為一年及兩年。會員效期到期之前，本資料庫將主動與會員聯絡資料校正及資格延長事宜。會員須於校正期間內確認有無資料異動情形，否則暫停其一般會員權限。

(二)院內會員：申請人須於資料庫網站填具申請表，提供中央研究院電子郵件信箱後並經確認，再提供中央研究院服務證影本向本資料庫提出申請。本資料庫於每年 11 月以電子郵件校正資料，離職後會員權限失效。