

問卷調查 (Questionnaire)

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.效度與信度
- 2.Rcmdr- Cronbach's α 內部一致性信度
- 3.Rmcdr-值標籤設定
- 4.Rmcdr-相關係數
- 5.Rcmdr-Levene檢定(同質性檢定)
- 6.Rmcdr-獨立樣本t檢定(變異數同質性與不同質性)
- 7.課程回顧

1.效度與信度

效度 (Validity)

- 效度：是研究方法中關於測量方法（衡量工具）能否如實反映所欲測量對象的能力，即測量值與實際值相符合的程度。一項效度高的衡量較一項效度低的衡量更為準確。
- 常用效度的指標：
 1. 建構效度 (Construct validity)：測量工具是否真的量度所測之物，使用因素分析法 (Factor analysis)。
 2. 內容效度 (Content validity)：測驗內容是否能代表樣本行為 (代表性、適切性)。
 3. 效標效度 (Criterion validity)：測量工具與其他量度工具的關係。

參考: [https://en.wikipedia.org/wiki/Validity_\(statistics\)](https://en.wikipedia.org/wiki/Validity_(statistics))

信度 (Reliability)

- 信度表示研究者對同一現象重複測量時，兩次結果會一致的高低程度。
- 信度的同義詞包括：
 - 可靠性、重複性、重現性
 - Reliability，Repeatability，Reproducibility
- 使用相同的衡量工具與相同衡量時間點，其結果愈相同，表示內部一致性信度（Internal Consistency Reliability）越高，代表各個衡量項目能夠測量到相同的構念，亦即測驗項目之間的同質性很高。
- 信度包括四種常用指標：**內部一致性信度**、再測信度、複本信度、評分者間信度。

參考: [https://en.wikipedia.org/wiki/Reliability_\(statistics\)](https://en.wikipedia.org/wiki/Reliability_(statistics))

Cronbach's α 係數

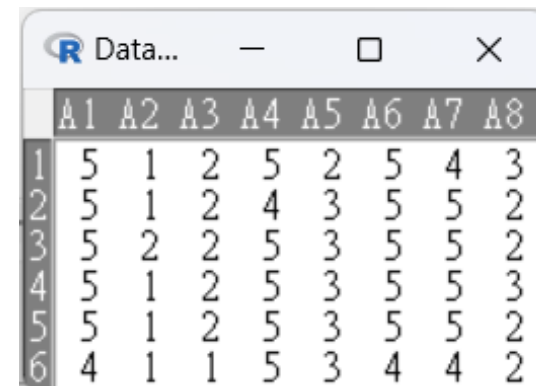
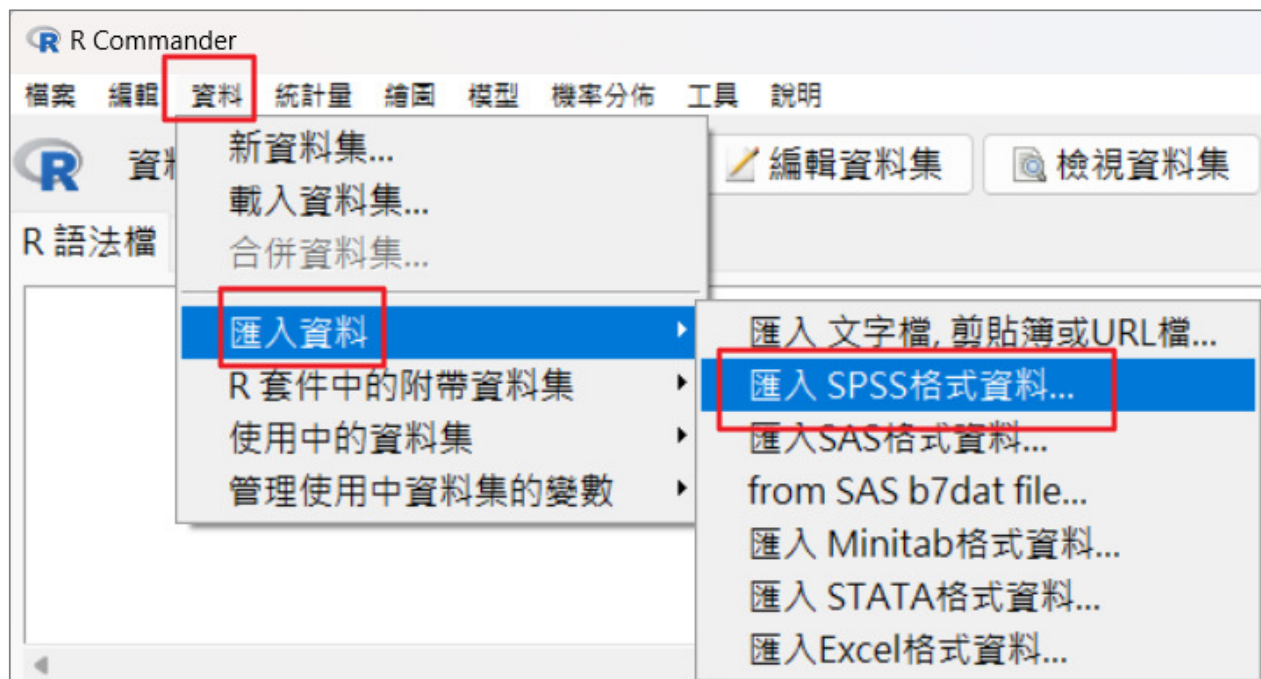
- 信度可以使用「**Cronbach's α 係數**」， α 係數是由美國教育心理學家—李·約瑟夫·克隆巴赫（[Lee Joseph Cronbach](#)）於1951所提出，是目前社會科學研究最常使用的信度分析方法。
- 一般研究報告中的 α 值包括以下結果：
 1. 大於或等於0.9 代表內部一致性信度很高（Excellent）
 2. 0.8–0.9 算好（Good）
 3. 0.7–0.8 可接受（Acceptable）：一般 α 係數 ≥ 0.7 為佳。
 4. 0.6–0.7 可疑的（Questionable）
 5. 0.5–0.6 較差（Poor）
 6. 小於0.5 不可接受（Unacceptable）

2.Rcmdr- Cronbach's α 内部一致性信度

- Rcmdr demo
- SPSS demo
- 人工計算

α 係數-Rcmdr

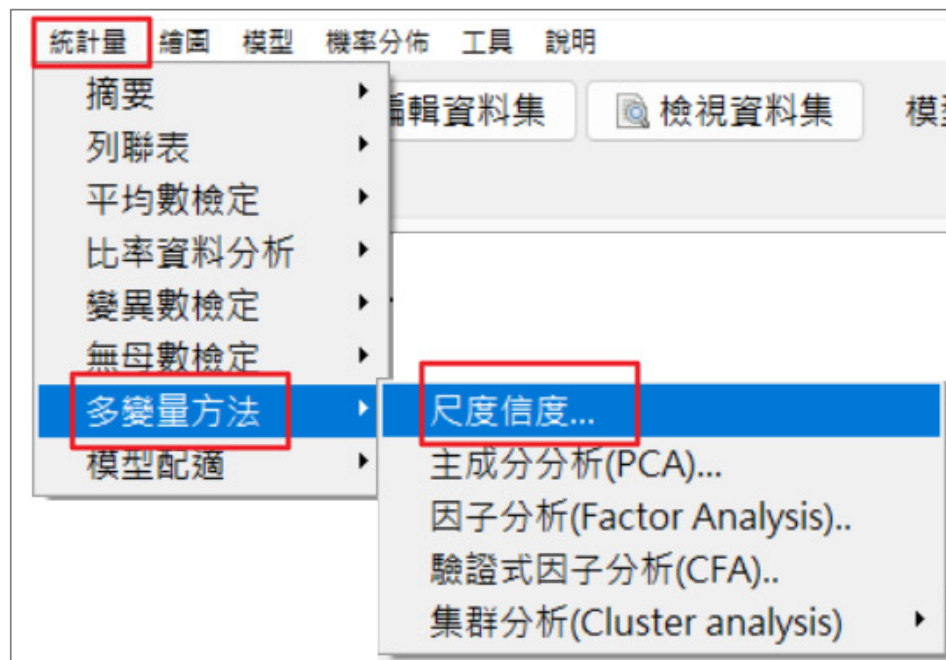
- https://github.com/rwepa/teaching-Rcmdr/blob/main/Rcmdr_reliability.sav
- 問卷採用李克特五點量表，題項數目8，樣本數6。
- 本章資料來源：吳明隆，SPSS操作與應用-問卷統計分析實務，五南，2009。



	A1	A2	A3	A4	A5	A6	A7	A8
1	5	1	2	5	2	5	4	3
2	5	1	2	4	3	5	5	2
3	5	2	2	5	3	5	5	2
4	5	1	2	5	3	5	5	3
5	5	1	2	5	3	5	5	2
6	4	1	1	5	3	4	4	2

尺度信度

- 統計量 \ 多變量方法 \ 尺度信度



Cronbach's α 係數

- Alpha reliability = 0.6374

```
> reliability(cov(Dataset[,c("A1","A2","A3","A4","A5","A6","A7","A8")], use="complete.obs"))  
Alpha reliability = 0.6374  
Standardized alpha = 0.6457  
  
Reliability deleting each item in turn:  
      Alpha Std.Alpha r(item, total)  
A1 0.4734    0.4801    0.8076  
A2 0.6292    0.6361    0.2370  
A3 0.4734    0.4801    0.8076  
A4 0.7111    0.7274   -0.1309  
A5 0.7111    0.7274   -0.1309  
A6 0.4734    0.4801    0.8076  
A7 0.5147    0.5303    0.6002  
A8 0.7049    0.6935    0.0000
```

α 係數-SPSS

- SPSS22: 分析 \ 尺度 \ 可靠度分析



Alpha 值
折半信度
Guttman 值
平行模式檢定
嚴密平行模式檢定

可靠性統計資料

Cronbach 的 Alpha	項目個數
.637	8

α 係數-SPSS (續)

- SPSS26: 分析 \ 比例 \ 信度分析



α
折半
Guttman
平行
嚴格平行

可靠性統計量	
Cronbach's Alpha	項目數
.637	8

α 係數-人工計算

- Cronbach's α 公式：
- $\alpha = \frac{k}{k-1} \times \left(1 - \frac{\sum_{i=1}^k S_i^2}{S^2} \right)$ ， k 題項個數， $S_i^2 = \frac{\sum_{i=1}^{n_i} (x_i - \bar{x}_i)^2}{n_i - 1}$ ， S^2 為題項加總後之變異數。
- 開啟 Rcmdr_reliability.sav，Ctrl+A → Ctrl+C → Excel → Ctrl+V → 新增變數名稱 A1...A8

	A1	A2	A3	A4	A5	A6	A7	A8
1	5	1	2	5	2	5	4	3
2	5	1	2	4	3	5	5	2
3	5	2	2	5	3	5	5	2
4	5	1	2	5	3	5	5	3
5	5	1	2	5	3	5	5	2
6	4	1	1	5	3	4	4	2



	A	B	C	D	E	F	G	H	I
1	樣本	A1	A2	A3	A4	A5	A6	A7	A8
2	Sample1	5	1	2	5	2	5	4	3
3	Sample2	5	1	2	4	3	5	5	2
4	Sample3	5	2	2	5	3	5	5	2
5	Sample4	5	1	2	5	3	5	5	3
6	Sample5	5	1	2	5	3	5	5	2
7	Sample6	4	1	1	5	3	4	4	2

α 係數-人工計算 – 完成版

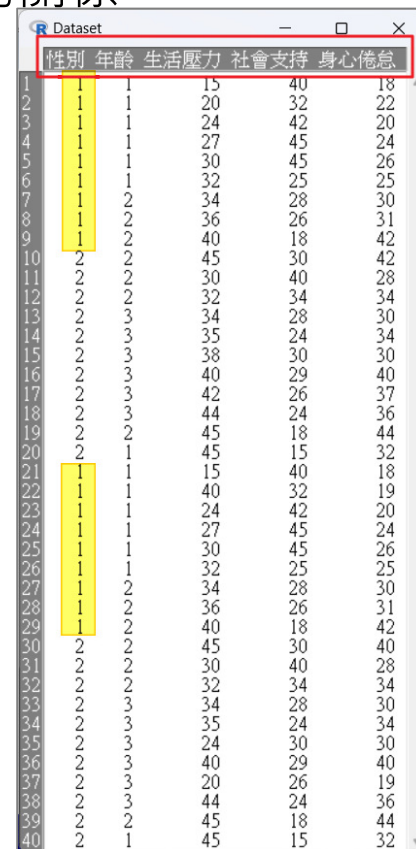
J12	=(k值/(k值-1))*(1-變異數加總/整體變異數)									
	A	B	C	D	E	F	G	H	I	J
1	樣本	A1	A2	A3	A4	A5	A6	A7	A8	加總
2	Sample1	5	1	2	5	2	5	4	3	27
3	Sample2	5	1	2	4	3	5	5	2	27
4	Sample3	5	2	2	5	3	5	5	2	29
5	Sample4	5	1	2	5	3	5	5	3	29
6	Sample5	5	1	2	5	3	5	5	2	28
7	Sample6	4	1	1	5	3	4	4	2	24
8	平均值	4.833333	1.166667	1.833333	4.833333	2.833333	4.833333	4.666667	2.333333	
9	變異數加總	0.166667	0.166667	0.166667	0.166667	0.166667	0.166667	0.266667	0.266667	1.533333
10	整體變異數									3.466667
11	k值									8
12	Cronbach's α									0.637363
13										
14	$\alpha = \frac{k}{k-1} \times \left(1 - \frac{\sum_{i=1}^k S_i^2}{S^2} \right), k \text{ 題項個數}, S_i^2 = \frac{\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}{n_i - 1},$									
15	S^2 為題項加總後之變異數。									
16										
17										
18										
19										
20	<pre>> reliability(cov(Dataset[,c("A1","A2","A3","A4","A5","A6","A7","A8")], use="complete.obs"))</pre>									
21	Alpha reliability = 0.6374									
22										
23	Powered by: alan9956@gmail.com									
24	Date: 2024.6.8									

3.Rmcdrr-值標籤設定

- 1: 男生, 2: 女生...

運動員問卷調查

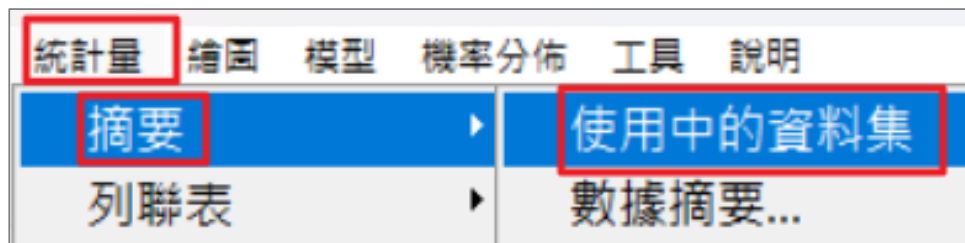
- https://github.com/rwepa/teaching-Rcmdr/blob/main/Rcmdr_athlete_pressure.csv
- 某體育教練想研究運動員之生活壓力、社會支持與身心倦怠的關係。
- 隨機抽取40名運動員。
- 背景變項：
 - 性別：1：男性，2：女性。
 - 年齡：1：25歲（含）以下，2：26~30歲，3：31歲（含）以上。
- 每位運動員填寫三張量表：
 - 生活壓力量表：分數高，生活壓力愈大。
 - 社會支持量表：分數高，社會支持愈多。
 - 身心倦怠量表：分數高：身心倦怠感愈大。
- 問題：
 1. 生活壓力、社會支持與身心倦怠的相關性為何？
 2. 不同性別對生活壓力、社會支持與身心倦怠是否有顯著差異？



	性別	年齡	生活壓力	社會支持	身心倦怠
1	1	1	15	40	18
2	1	1	20	32	22
3	1	1	24	42	20
4	1	1	27	45	24
5	1	1	30	45	26
6	1	1	32	25	25
7	1	2	34	28	30
8	1	2	36	26	31
9	1	2	40	18	42
10	2	2	45	30	42
11	2	2	30	40	28
12	2	2	32	34	34
13	2	3	34	28	30
14	2	3	35	24	34
15	2	3	38	30	30
16	2	3	40	29	40
17	2	3	42	26	37
18	2	3	44	24	36
19	2	2	45	18	44
20	2	1	45	15	32
21	1	1	15	40	18
22	1	1	40	32	19
23	1	1	24	42	20
24	1	1	27	45	24
25	1	1	30	45	26
26	1	1	32	25	25
27	1	2	34	28	30
28	1	2	36	26	31
29	1	2	40	18	42
30	2	2	45	30	40
31	2	2	30	40	28
32	2	2	32	34	34
33	2	3	34	28	30
34	2	3	35	24	34
35	2	3	24	30	30
36	2	3	40	29	40
37	2	3	20	26	19
38	2	3	44	24	36
39	2	2	45	18	44
40	2	1	45	15	32

摘要 summary 函數

- 統計量 \ 摘要 \ 使用中的資料集



```
> summary(Dataset)
```

性別	年齡	生活壓力	社會支持	身心倦怠
Min. :1.00	Min. :1.00	Min. :15	Min. :15.00	Min. :18.00
1st Qu.:1.00	1st Qu.:1.00	1st Qu.:30	1st Qu.:24.75	1st Qu.:25.00
Median :2.00	Median :2.00	Median :34	Median :28.50	Median :30.00
Mean :1.55	Mean :1.95	Mean :34	Mean :29.95	Mean :30.68
3rd Qu.:2.00	3rd Qu.:3.00	3rd Qu.:40	3rd Qu.:35.50	3rd Qu.:36.00
Max. :2.00	Max. :3.00	Max. :45	Max. :45.00	Max. :44.00

結構 str 函數

- str(Dataset)

- 須將背景變項的整數 (int) 改為因子 (factor)

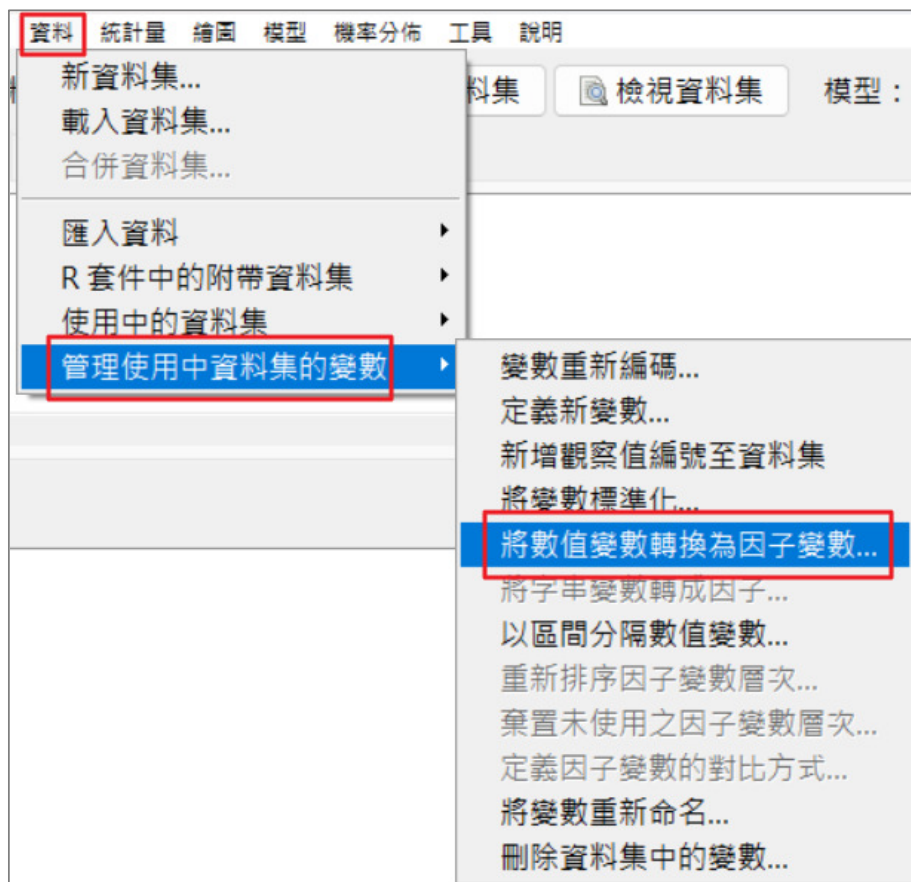
```
str(Dataset)
```

Output

```
> str(Dataset)
'data.frame': 40 obs. of 5 variables:
 $ 性別      : int  1 1 1 1 1 1 1 1 1 2 ...
 $ 年齡      : int  1 1 1 1 1 1 2 2 2 2 ...
 $ 生活壓力: int  15 20 24 27 30 32 34 36 40 45 ...
 $ 社會支持: int  40 32 42 45 45 25 28 26 18 30 ...
 $ 身心倦怠: int  18 22 20 24 26 25 30 31 42 42 ...
```

將數值變數轉換為因子變數

- 資料 \ 管理使用中資料集的變數 \ 將數值變數轉換為因子變數

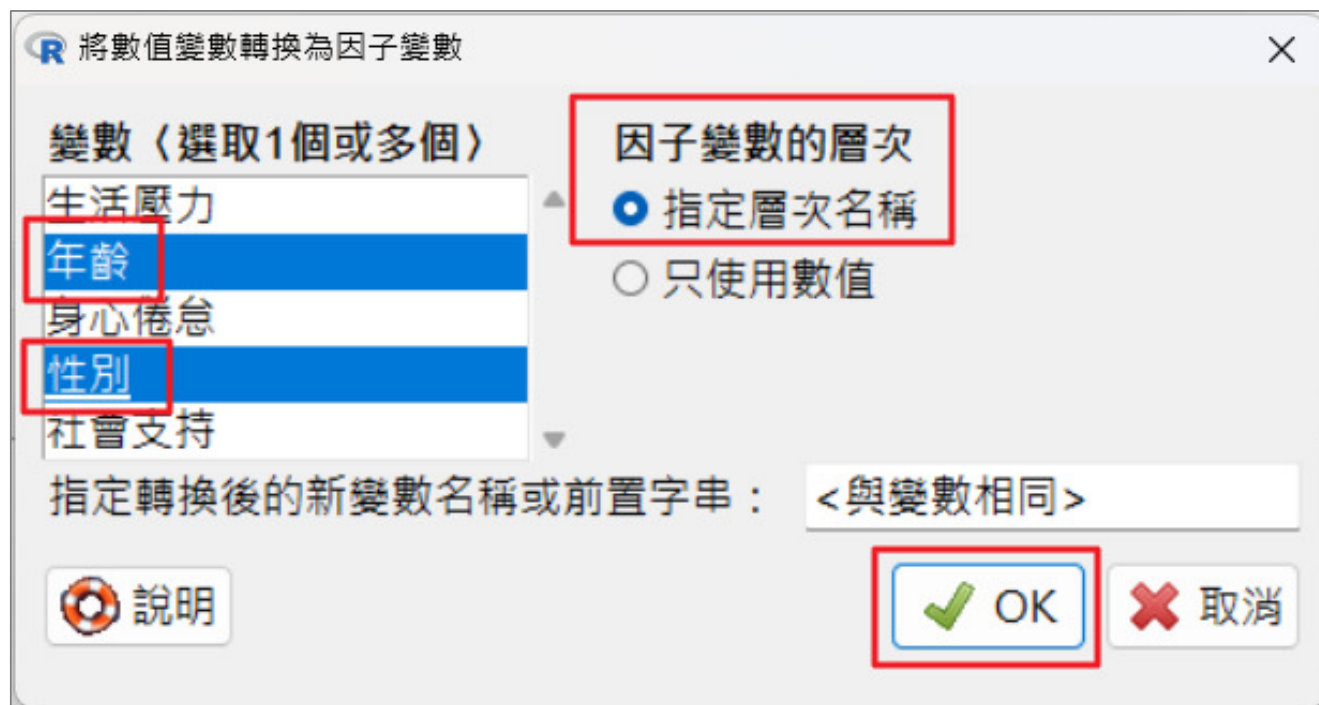


類似 SPSS 值標籤設定

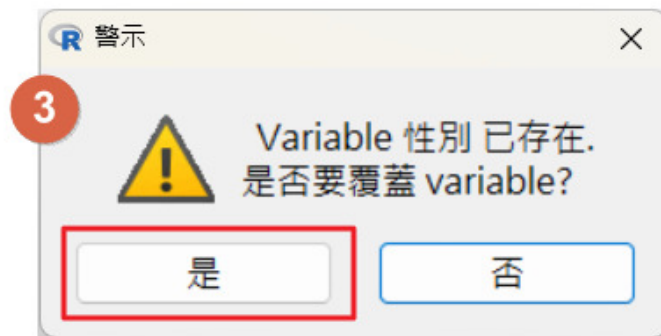
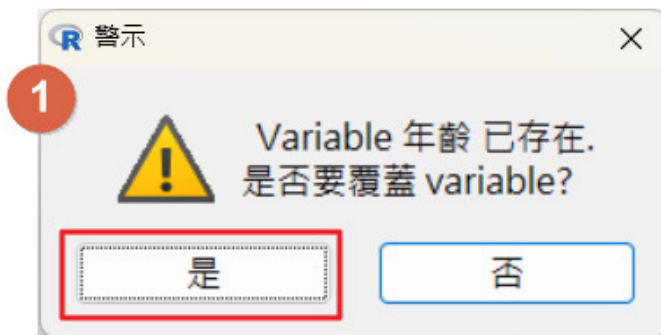
值
{1, 男生}...
{1, 25歲以下}...

將數值變數轉換為因子變數 (續)

- 使用 Ctrl + 左鍵 選取年齡、性別



將數值變數轉換為因子變數 (續)



將數值變數轉換為因子變數 (續)

```
Dataset <- within(Dataset, {  
  年齡 <- factor(年齡, labels=c('25歲以下', '26-30歲', '31歲以上'))  
  性別 <- factor(性別, labels=c('男生', '女生'))  
})
```

將數值變數轉換為因子變數-完成版

Output

```
> str(Dataset)
'data.frame': 40 obs. of 5 variables:
 $ 性別      : Factor w/ 2 levels "男生","女生": 1 1 1 1 1 1 1 1 1 2 ...
 $ 年齡      : Factor w/ 3 levels "25歲以下","26-30歲",...: 1 1 1 1 1 1 2 2 2 2 ...
 $ 生活壓力: int  15 20 24 27 30 32 34 36 40 45 ...
 $ 社會支持: int  40 32 42 45 45 25 28 26 18 30 ...
 $ 身心倦怠: int  18 22 20 24 26 25 30 31 42 42 ...
```

```
> summary(Dataset)
```

性別	年齡	生活壓力	社會支持	身心倦怠
男生:18	25歲以下:14	Min. :15	Min. :15.00	Min. :18.00
女生:22	26-30歲 :14	1st Qu.:30	1st Qu.:24.75	1st Qu.:25.00
	31歲以上:12	Median :34	Median :28.50	Median :30.00
		Mean :34	Mean :29.95	Mean :30.68
		3rd Qu.:40	3rd Qu.:35.50	3rd Qu.:36.00
		Max. :45	Max. :45.00	Max. :44.00

4.Rmcdr-相關係數

相關矩陣

- 統計量 \ 摘要 \ 相關矩陣



相關矩陣 – 完成版

Output

Pearson correlations:

	生活壓力	身心倦怠	社會支持
生活壓力	1.0000	0.8125	-0.6704
身心倦怠	0.8125	1.0000	-0.6186
社會支持	-0.6704	-0.6186	1.0000

Number of observations: 40

Pairwise two-sided p-values:

	生活壓力	身心倦怠	社會支持
生活壓力		<.0001	<.0001
身心倦怠	<.0001		<.0001
社會支持	<.0001	<.0001	

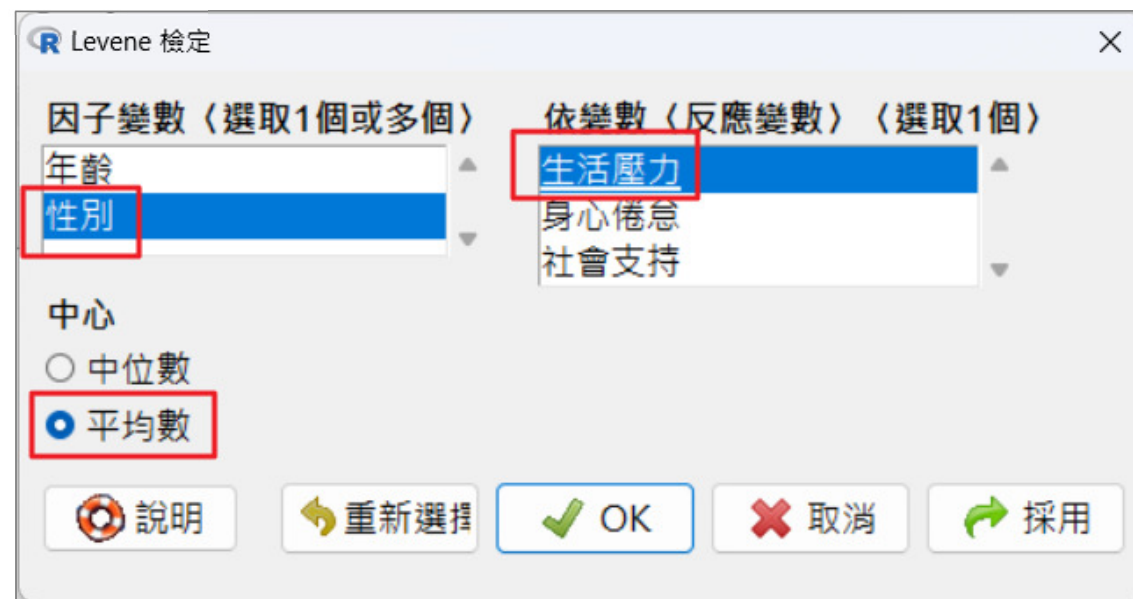
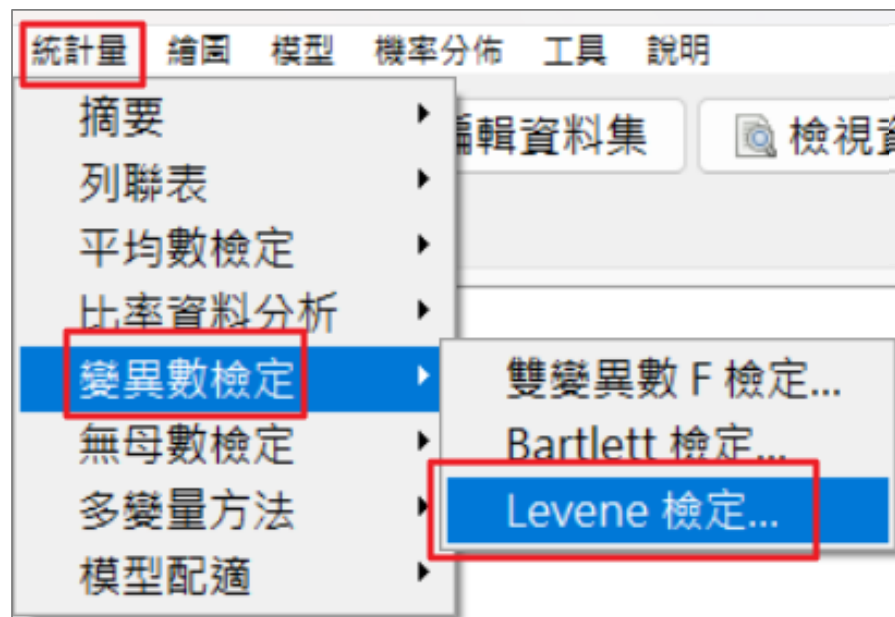
Adjusted p-values (Holm's method)

	生活壓力	身心倦怠	社會支持
生活壓力		<.0001	<.0001
身心倦怠	<.0001		<.0001
社會支持	<.0001	<.0001	

5.Rcmdr-Levene検定(同質性検定)

Levene檢定

- 統計量 \ 變異數分析 \ Levene檢定




```
> Tapply(生活壓力 ~ 性別, var, na.action=na.omit, data=Dataset) # variances by group
      男生      女生
61.83007 54.92641
```

```
> leveneTest(生活壓力 ~ 性別, data=Dataset, center="mean")
Levene's Test for Homogeneity of Variance (center = "mean")
      Df F value Pr(>F)
group 1  0.0009 0.9756
      38
```

變異數同質性

```
> Tapply(社會支持 ~ 性別, var, na.action=na.omit, data=Dataset) # variances by group
      男生      女生
91.79085 46.18182
```

```
> leveneTest(社會支持 ~ 性別, data=Dataset, center="mean")
Levene's Test for Homogeneity of Variance (center = "mean")
      Df F value Pr(>F)
group 1  6.4519 0.01529 *
      38
```

變異數不同質性

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> Tapply(身心倦怠 ~ 性別, var, na.action=na.omit, data=Dataset) # variances by group
      男生      女生
51.03595 36.01732
```

```
> leveneTest(身心倦怠 ~ 性別, data=Dataset, center="mean")
Levene's Test for Homogeneity of Variance (center = "mean")
      Df F value Pr(>F)
group 1  0.3532 0.5558
      38
```

變異數同質性

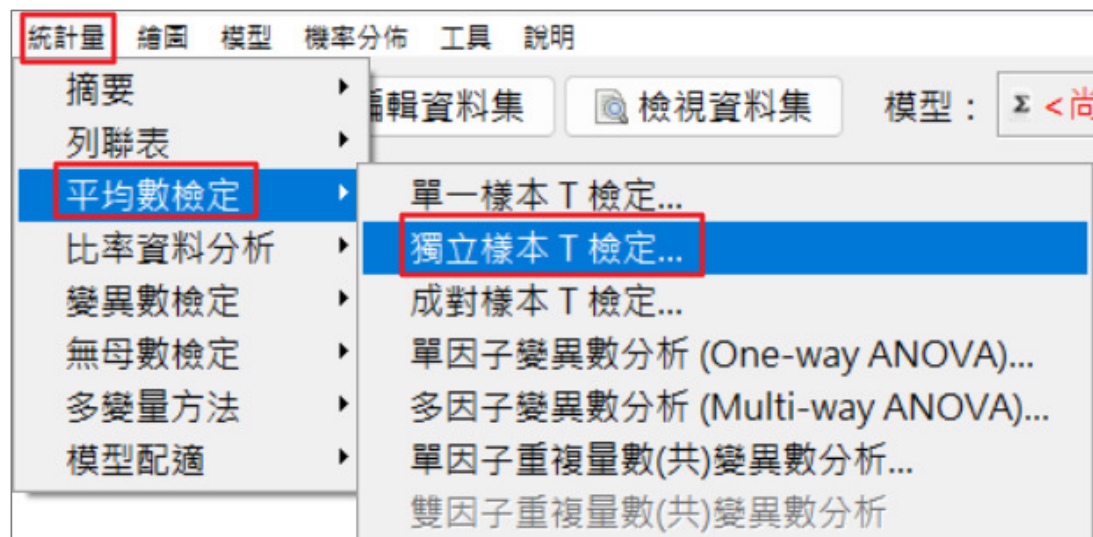
$H_0: \sigma_{男}^2 = \sigma_{女}^2$ 變異數相等;符合變異數同質性

$H_1: \sigma_{男}^2 \neq \sigma_{女}^2$ 變異數不相等;不符合變異數同質性 (p值<0.05)

6.Rmcdrr-獨立樣本t檢定 (變異數同質性與不同質性)

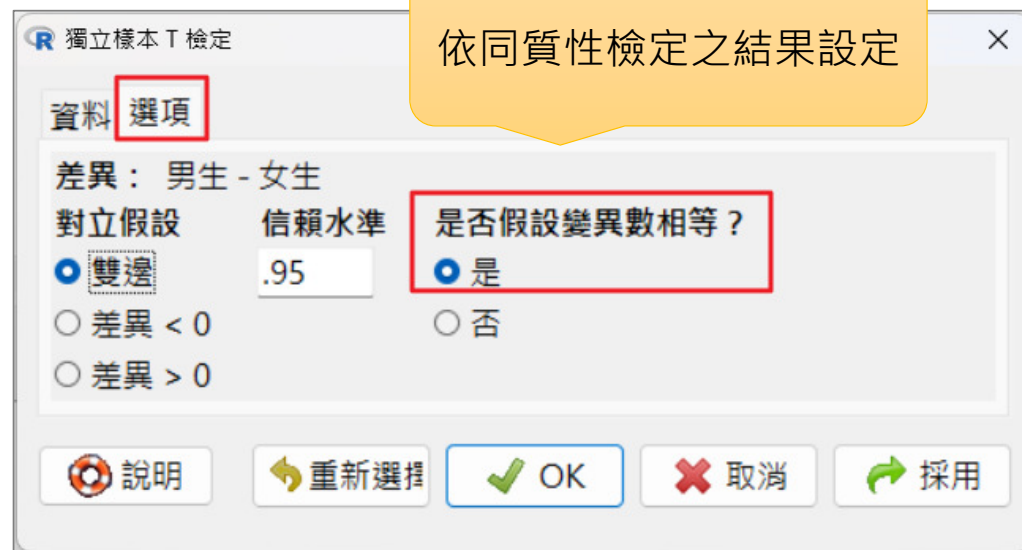
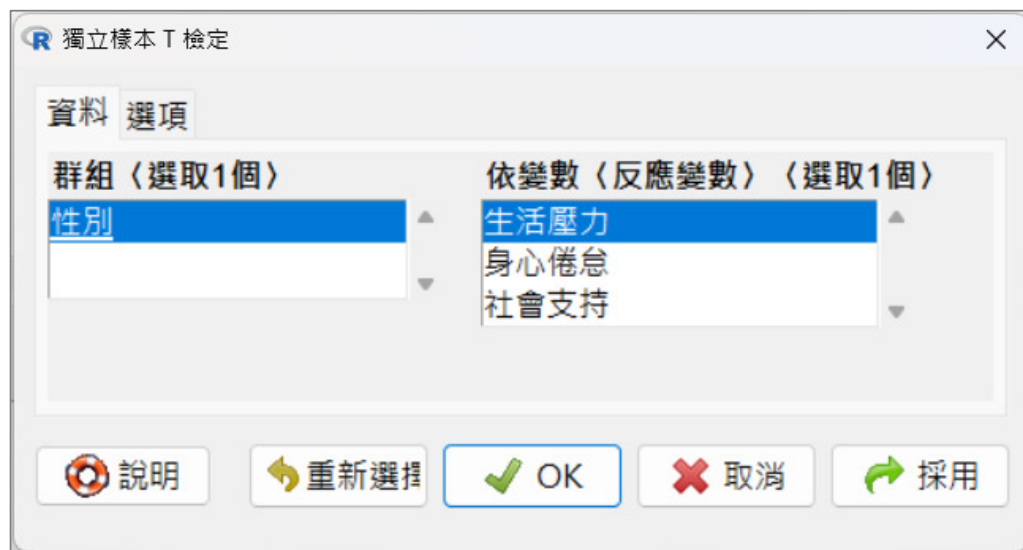
獨立樣本 T 檢定

- 生活壓力：變異數同質性
- 社會支持：變異數不同質性
- 身心倦怠：變異數同質性
- 統計量 \ 平均數檢定 \ 獨立樣本 T 檢定



獨立樣本 T 檢定-遺項視窗

- 群組：性別；依變數：生活壓力；選項：是假設變異數相等



- p 值 <0.05 ，達到顯著水準，表示男女運動員的生活壓力感受有顯著差異，女生平均生活壓力高於男生平均生活壓力。
- 同理，社會支持與身心倦怠亦可採用獨立樣本 t 檢定。

```
> t.test(生活壓力 ~ 性別, alternative = "two.sided", conf.level = .95, var.equal = TRUE, data = Dataset)

Two Sample t-test

data: 生活壓力 by 性別
t = -3.1712, df = 38, p-value = 0.002999
alternative hypothesis: true difference in means between group 男生 and group 女生 is not equal to 0
95 percent confidence interval:
 -12.577344 -2.776191
sample estimates:
mean in group 男生 mean in group 女生
      29.77778      37.45455
```


7.課程回顧

Reviews

- R, RStudio
- Rcmdr
- Import Tab File (匯入Tab檔案)
- 宗教社會服務概況資料分析
- 研究方法與開放資料
- ggplot2 簡介
- 套件 package
- t檢定
- 變異數分析(ANOVA)
- 迴歸 (Regression)
- 問卷調查

謝謝您的聆聽

Q & A



李明昌

alan9956@gmail.com

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