Q02

R 語言問卷分析(2)

量表題項分析

吳漢銘

國立政治大學 統計學系

https://hmwu.idv.tw

項目分析: 大綱

- **項目(題項、題目)分析**主要目的: 檢核題項適切(可靠)程度。
  - 探究高低分的受試者在每題項的差異比較。
  - 進行題項間同質性檢核。
  - 作為個別題項篩選或修改依據。

- 信度考驗: 檢核整份量表，或層面(構念)可靠程度。

- 預試問卷施測完:
  - 預試問卷項目分析。
  - 效度考驗(分析)。
  - 信度檢定(分析)。

本講義部份內容掃瞄自以下所列之上課用書，並無它用。
吳明隆, SPSS操作與應用：問卷統計分析實務（附光碟） 五南出版社, 出版日期／2010/10/05(2版 5刷)。

- 二元計分試題題項分析 (多元計分試題題項分析、分量表多元計分題項分析)

基本概念: 難度

- 分析測驗的難度、鑑別度及誘答力。
  - 將測驗總得分，前25%~33% 設為高分組。後25%~33%設為低分組。
  - 算出高低二組在每個試題答對人數的百分比，再算出試題的難度與鑑別度。

- **難度公式** $P = (PH + PL) / 2$
  - PH (PL): 高(低)分組在某題項答對人數百分比 。
  - P 範圍: (0, 1)
  - P 愈大(愈接近1): 題目愈容易，愈多答對者。
  - $P = 0.5$: 答對，答錯各佔一半，難易適中。
  - 較佳的測驗: P介於0.2至0.8之間。

基本概念: 鑑別度

- **鑑別度公式** $D = PH - PL$
  - 高分組答對百分比，與低分組答對百分比之差異值。
  - 主要目的: 判別試題是否具有區別受試者能力高低的功能。
  - 鑑別度範圍: (-1, 1)
  - 具鑑別度的測驗(愈接近1): 高分組答對百分比高於低分組答對百分比。
  - 一般: D在0.3以上。P在0.5左右。

- 二元計分試題題項分析之**鑑別度公式:** 點二列相關係數 (Point-Biserial Correlation)。

項目分析之步驟

項目分析之步驟

1. 量表反向題項反向計分。
2. 計算量表總分。
3. 排序量表總分。
4. 求出高低分組上下27%處分數 (臨界點) 。
5. 將量表分出高低分組。
6. t檢定: 檢定高低分在每個題項的差異。
7. 將未達顯著的題項刪除。
8. 同質性檢驗: 信度檢核、共同性與因素負荷量。

注意事項

- 若是測驗分數呈常態分佈，以27%作為分組，得到之鑑別度的可靠性最大。
- 採用25%~33%分組法也可。
- 若預試樣本數較大(小)，可選取大(小)於27%的分組法。

範例: 社會參與量表

- 「社會參與量表」經**專家效度**審核後，保留19題，進行問卷預試。
- 反向題: 2, 15, 18
- 隨機取樣退休公教人員填答。
- 回收問卷，刪除無效問卷，得有效問卷200份。
- 進行「項目分析」，以檢核「社會參與量表」19題項的適切性。

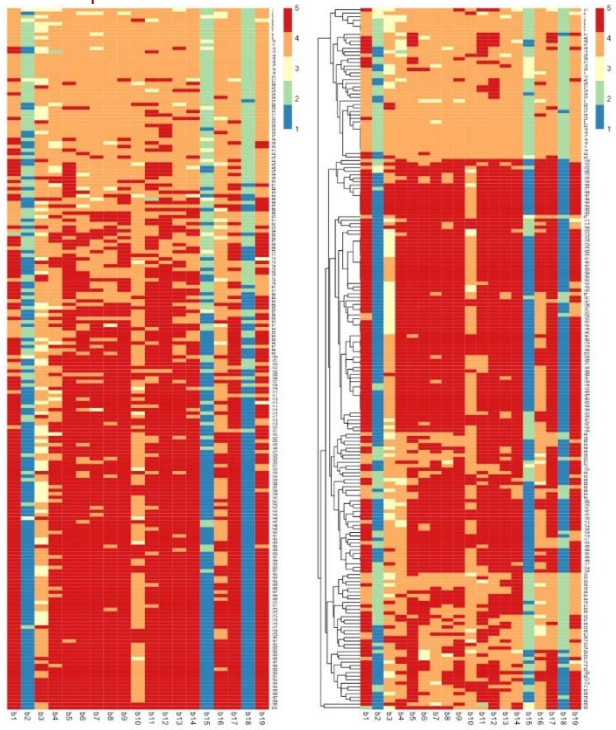
【社會參與量表】

	非常同意	大部分同意	一半同意	大部分不同意	非常不同意
01.我常會喜歡宗教活動中的一些儀式	□	□	□	□	□
02.我不覺得參加社會服務工作很有意義	□	□	□	□	□
03.我常會選擇自己喜歡的社團活動來參與	□	□	□	□	□
04.我常會參加各種進修學習活動	□	□	□	□	□
05.我常會主動地參加一些休閒娛樂活動	□	□	□	□	□
06.我常會參加一些宗教活動	□	□	□	□	□
07.我覺得參加社會服務工作之後使我的生活更加充實	□	□	□	□	□
08.我參與社團活動時我會積極地投入	□	□	□	□	□
09.我參加學習活動時，常會受到老師和同儕的肯定	□	□	□	□	□
10.我現在常會邀請家人或親友一起從事運動休閒	□	□	□	□	□
11.我常會鼓勵親朋好友一起參加宗教活動	□	□	□	□	□
12.我很願意奉獻自己的專長和經驗來服務別人	□	□	□	□	□
13.我常會做些適合我的運動來增進健康	□	□	□	□	□
14.我參加學習活動時心情都很愉快	□	□	□	□	□
15.我不覺得從事適當的休閒運動後能讓生活更充實	□	□	□	□	□
16.參加宗教活動之後讓我的心靈更為充實	□	□	□	□	□
17.我常會利用時間參加各種社會服務工作	□	□	□	□	□
18.我不覺得我能從社團活動的參與過程中獲得滿足感	□	□	□	□	□
19.我會主動和別人分享進修學習的心得	□	□	□	□	□

1. 檢查資料有無極端值或錯誤值

- 計算「次數分配表」及「描述性統計量」。
  - 觀察每個題項有無勾選錯誤值。李克特五點量表: 小於1，或大於5均為遺漏值或錯誤值。

```
> social_participation_data <- read.csv("data/參與量表-原始資料.csv")
> head(social_participation_data)
  b1 b2 b3 b4 b5 b6 b7 b8 b9 b10 b11 b12 b13 b14 b15 b16 b17 b18 b19
1  4  3  3  3  4  4  4  4  4   4   4   4   3   4   2   4   4   2   4
2  4  1  4  4  4  3  3  4  3   4   4   4   4   4   2   3   4   2   4
...
6  4  2  3  4  4  4  4  4  4   4   4   4   4   4   2   4   4   2   4
> library(RColorBrewer)
> library(pheatmap)
> pheatmap(social_participation_data,
+          color = rev(brewer.pal(5, "Spectral")),
+          cluster_rows = FALSE,
+          cluster_cols = FALSE,
+          fontsize_row = 5)
>
> check_summary <- function(x){
+   output <- c(min(x), max(x), length(x[is.na(x)]),
+                               length(x[is.nan(x)]))
+   names(output) <- c("min", "max", "NA's", "NaN's")
+   output
+ }
>
> apply(social_participation_data, 2, check_summary)
      b1 b2 b3 b4 b5 b6 b7 b8 b9 b10 b11 b12 b13 b14 b15 b16 b17 b18 b19
min    2  1  3  2  4  3  3  3  3   3   2   4   3   3   1   3   3   1   3
max    5  3  5  5  5  5  5  5  5   5   5   5   5   5   3   5   5   2   5
NA's   0  0  0  0  0  0  0  0  0   0   0   0   0   0   0   0   0   0   0
NaN's  0  0  0  0  0  0  0  0  0   0   0   0   0   0   0   0   0   0   0
```

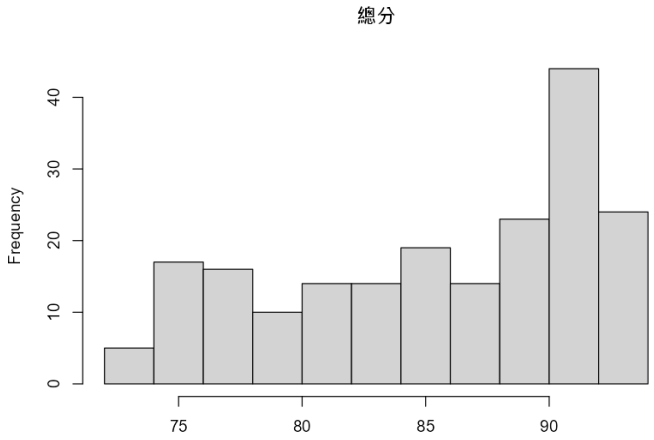


2. 反向題反向計分
3. 計算量表題項總分

- 量表中最好能編製一至三題反向題，以測知受試者填答的效度。

```
> # 反向題反向計分
> sp_data_used <- social_participation_data
> apply(sp_data_used[, c(2, 15, 18)], 2, table)
$b2
  1   2   3
124  71   5

$b15
  1   2   3
110  79  11

$b18
  1   2
118  82
> sp_data_used[, c(2, 15, 18)] <- 6 - sp_data_used[, c(2, 15, 18)]
> apply(sp_data_used[, c(2, 15, 18)], 2, table)
$b2
  3   4   5
  5  71 124

$b15
  3   4   5
 11  79 110

$b18
  4   5
 82 118
```

```
> # 計算量表題項總分
> sp_total_score <- apply(sp_data_used, 1, sum)
> sp_total_score
  [1] 72 73 73 74 74 75 75 75 75 75 76 76 76 76
...
[190] 93 93 94 94 94 94 94 94 94 94 94
> hist(sp_total_score, main = "總分")
```



4.求出高低分組的臨界分數
5. 進行高低分組

```
> bks <- quantile(sp_total_score, probs = c(0, 0.27, 1 - 0.27, 1))
> bks
   0%   27%   73%  100%
72.00 81.73 91.00 94.00
> # Levels: (72,81.7] (81.7,91] (91,94]
> LMH_group <- cut(sp_total_score, breaks = bks,
+                  labels = c("低分組(<= 81)", "不分組(82 - 91)", "高分組(> 91)"))
> table(LMH_group)
LMH_group
  低分組(<= 81) 不分組(82 - 91)    高分組(> 91)
             53              99              47
>
> # Levels: [72,81.7) [81.7,91) [91,94]
> LMH_group <- cut(sp_total_score, breaks = bks,
+                  labels = c("低分組(<= 81)", "不分組(82 - 91)", "高分組(> 91)"),
+                  include.lowest = T, right = F)
> table(LMH_group)
LMH_group
  低分組(<= 81) 不分組(82 - 91)    高分組(> 91)
             54              78              68
```

6. 求決斷值-臨界比, T-檢定

- 項目分析中以量表總分前27%及後27%的差異比較，稱為「二個極端組比較」。
- 比較結果的差異值稱為「決斷值」，或「臨界比」(critical Ratio，CR)。
- 決斷值未達顯著的題項，最好刪除。(t-檢定)

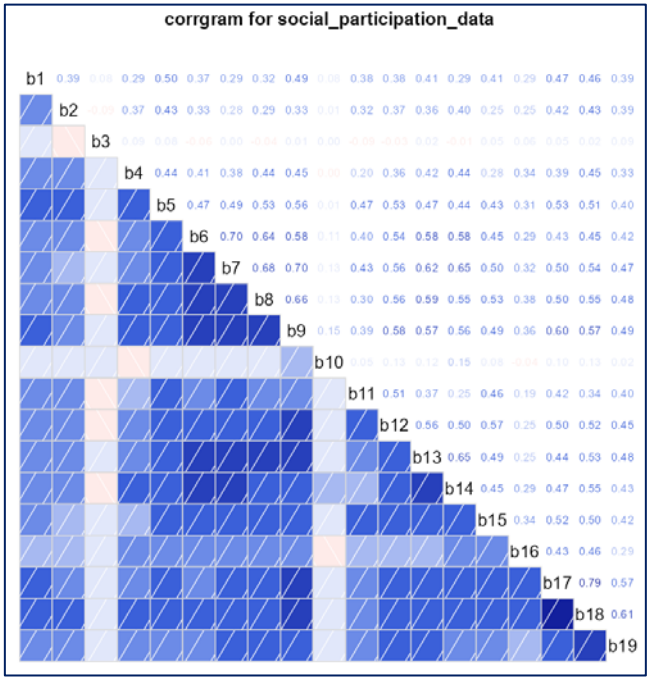
```
> low_id <- LMH_group == "低分組(<= 81)"
> high_id <- LMH_group == "高分組(> 91)"
> LH_group <- factor(c(rep("L", length(which(low_id))),
+                      rep("H", length(which(high_id)))))
> sp_data_LH <- rbind(sp_data_used[low_id, ], sp_data_used[high_id, ])
>
> my_t_test <- function(x){
+   is.var.equal <- var.test(x ~ LH_group)$p.value > 0.05
+   t.test(x ~ LH_group, var.equal = is.var.equal)$p.value
+ }
> significance_table <- sapply(sp_data_LH, my_t_test)
> significance_table
          b1           b2           b3           b4           b5           b6           b7
6.527598e-15 3.600876e-13 1.671834e-01 2.908671e-22 4.473941e-17 3.637639e-29 3.396719e-39
...
> var_id <- which(sapply(sp_data_LH, my_t_test) < 0.05)
> names(significance_table)[var_id] <- paste0(names(significance_table)[var_id], "*")
> significance_table
         b1*          b2*           b3          b4*          b5*          b6*          b7*
6.527598e-15 3.600876e-13 1.671834e-01 2.908671e-22 4.473941e-17 3.637639e-29 3.396719e-39
         b8*          b9*         b10*         b11*         b12*         b13*         b14*
5.167221e-48 9.519845e-24 8.170996e-03 5.308229e-13 3.665693e-17 4.746592e-30 4.116232e-23
        b15*         b16*         b17*         b18*         b19*
6.208575e-22 8.390497e-13 2.236296e-22 6.095099e-35 8.909022e-37
```

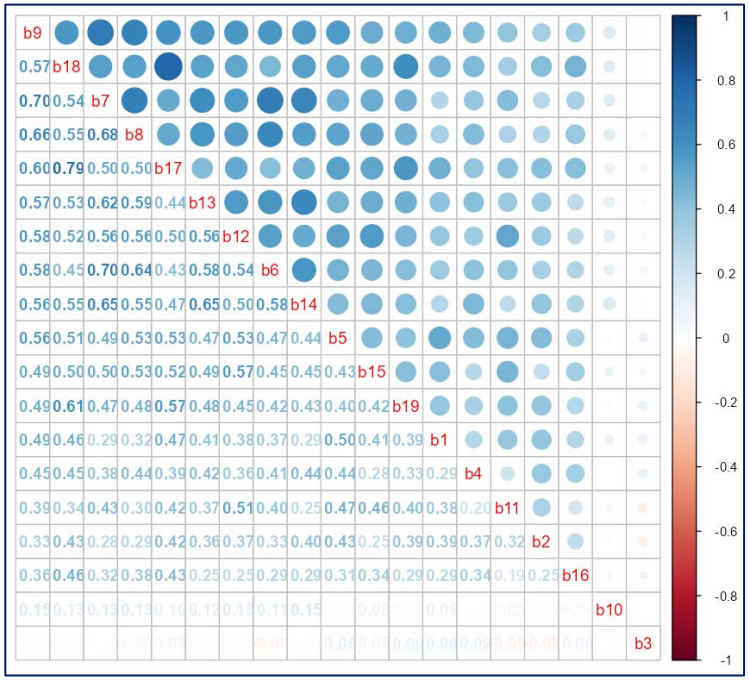
(獨立)雙樣本中位數差異檢定 (without assuming the normal distribution): Wilcoxon Rank-Sum Test (Mann-Whitney U Test)

```
wilcox.test(x ~ group, data)
```

7. 求題項與總分之相關

- 採用「同質性考驗」作為題項篩選的另一個指標。
- 個別題項與總分相關愈高(低)，表示題項與整體量表的同質性愈高(低)。
- 同質性不高的題項，最好刪除。
- 同質性考驗在求出個別題項與總分的相關係數。





https://cran.r-project.org/web/packages/corrplot/vignettes/corrplot-intro.html

```
library(corrgram)
corrgram(cor(sp_data_used),
         upper.panel = panel.cor,
         main = "corrgram")
```

```
library(corrplot)
corrplot(cor(sp_data_used))
# first principal component order
corrplot.mixed(cor(sp_data_used), order = 'FPC')
```

Pearson Correlation Test

Pearson correlation test, **check the test assumptions**

- Is the covariation linear? In the situation where the scatter plots show curved patterns, we are dealing with nonlinear association between the two variables.
- Are the data from each of the 2 variables (x, y) follow a normal distribution?
  - Use Shapiro-Wilk normality test: `shapiro.test()`
  - and look at the normality plot: `ggpubr::ggqqplot()`

```
> my_cor <- function(x){
+   r <- cor(x, sp_total_score)
+   cor_p <- cor.test(x, sp_total_score)$p.value
+   c(corr = round(r, 5), p.value = round(cor_p, 5))
+ }
> sapply(sp_data_used, my_cor)
             b1      b2      b3      b4      b5      b6
corr    0.60477 0.54523 0.12183 0.60115 0.71333 0.72479
p.value 0.00000 0.00000 0.08570 0.00000 0.00000 0.00000
             b7     b8      b9     b10     b11     b12
corr    0.77078 0.7568 0.79027 0.18534 0.56142 0.73026
p.value 0.00000 0.0000 0.00000 0.00860 0.00000 0.00000
            b13     b14     b15     b16     b17     b18
corr    0.74487 0.72024 0.68985 0.51248 0.75846 0.78178
p.value 0.00000 0.00000 0.00000 0.00000 0.00000 0.00000
            b19
corr    0.67924
p.value 0.00000
```

刪除b3, b10
其它相關皆在0.5以上。

Kendall rank correlation test

The Kendall rank correlation coefficient or Kendall's tau statistic is used to estimate a rank-based measure of association. This test may be used if the data do not necessarily come from a bivariate normal distribution.

`cor.test(x, y,  method = "kendall")`

「修正的項目總相關」：`corr(b1, sum(b2 to b19))`

```
> sapply(1:19, function(x) cor(sp_data_used[,x], apply(sp_data_used[, -x], 1, sum)))
 [1] 0.54993747 0.47953148 0.02225971 0.53017808 0.67501815
 [6] 0.67878075 0.73117057 0.71691918 0.75908436 0.11332615
[11] 0.49904119 0.69523287 0.70467760 0.67634517 0.63362354
[16] 0.44368235 0.72194761 0.74780129 0.62959810
```

8. 同質性檢驗: 信度檢核

- 信度(reliability)代表量表的**一致性**或**穩定性**。
- 信度係數在項目分析中，也可作為同質性檢核指標之一。
- **信度定義: 真實分數的變異數占測量總分數變異數的比例。**
- 若一量表在量測相同的特質，則題數愈多，則信度愈高。
- 採用最多: Cronbach's alpha (克隆巴赫 alpha係數)
  - 又稱**內部一致性alpha係數**。
  - 比較題項刪除前後之信度係數。
  - 信度係數最好在0.8以上。
- $Alpha = \frac{K}{K-1}(1 - \frac{\sum_{k=1}^{K} S_k^2}{S^2})$
  - K: 題數
  - S^2: 量表總分變異數
  - $\sum_{k=1}^{K} S_k^2$: 題項變異數總和

Cronbach's Alpha Range	Internal Consistency of data
>=0.9	Excellent
0.8-0.9	Good
0.7-0.8	Acceptable
0.6-0.7	Questionable
0.5-0.6	Poor
<0.5	Unacceptable

- 採用「一致性alpha係數」，僅適合「單一層面」「單一面向」或「單一構念」之(子)量表。

信度分析

```
> K <- ncol(sp_data_used)
> # ltm: Latent Trait Models under IRT
> library(ltm)
> cronbach.alpha(sp_data_used)

Cronbach's alpha for the 'sp_data_used' data-set

Items: 19
Sample units: 200
alpha: 0.912

> sapply(1:K, function(x) cronbach.alpha(sp_data_used[, -x])$alpha)
 [1] 0.9084485 0.9103116 0.9238492 0.9094739 0.9058152
 [6] 0.9050281 0.9036015 0.9041323 0.9036410 0.9178661
[11] 0.9097500 0.9055538 0.9045618 0.9052406 0.9062305
[16] 0.9112485 0.9043378 0.9036411 0.9064082
```

- 刪除某個題項後，alpha變大，表示此題所測量的特質與其它題項所測的特質，並不同質，可考慮刪除。
- B3: 0.912 => 0.924
- B10: 0.912 => 0.918

Cronbach's Alpha Range	Internal Consistency of data
>=0.9	Excellent
0.8-0.9	Good
0.7-0.8	Acceptable
0.6-0.7	Questionable
0.5-0.6	Poor
<0.5	Unacceptable

```
> my_cronbach_alpha <- function(x){
+   K <- ncol(x)
+   Sk2 <- apply(x, 2, var)
+   S2 <- var(apply(x, 1, sum))
+   cronbach_alpha <- (K/(K-1)) * (1 - sum(Sk2)/S2)
+   cronbach_alpha
+ }
> my_cronbach_alpha(sp_data_used)
[1] 0.9124309
> sapply(1:K, function(x) my_cronbach_alpha(sp_data_used[, -x]))
 [1] 0.9084485 0.9103116 0.9238492 0.9094739 0.9058152
 [6] 0.9050281 0.9036015 0.9041323 0.9036410 0.9178661
[11] 0.9097500 0.9055538 0.9045618 0.9052406 0.9062305
[16] 0.9112485 0.9043378 0.9036411 0.9064082
```

9. 同質性檢驗: 共同性

- 共同性 (communalities)表示題項能解釋共同特質或屬性的變異量。
- **Communality** of a variable is the percentage of that variable's variance that is explained by the factors.
- 例如: 社會參與量表限定一個因素時，表示只有一個心理特質，因而共同性數值愈高，能測量到此心理特質的程度愈高。
- 共同性較低的題項，表示與量表的同質性較低，可考慮刪除。
- 共同性萃取值: 題項在共同因素之因素負荷量的平方加總。
  - 即題項與共同因素間多元相關數的平方。
  - 即迴歸分析中的R^2
  - 共同值若小於0.2，表示題項與共同因素的關係不密切，可刪除 。

```
> sp_pca <- princomp(sp_data_used, cor = TRUE)
> round(sapply(sp_data_used, function(x) cor(x, sp_pca$scores[, 1])^2), 4)
    b1     b2     b3     b4     b5     b6     b7     b8     b9    b10
0.3517 0.2931 0.0006 0.3370 0.5174 0.5480 0.6212 0.6011 0.6504 0.0196
   b11    b12    b13    b14    b15    b16    b17    b18    b19
0.3226 0.5620 0.5707 0.5349 0.4730 0.2426 0.5845 0.6234 0.4626
```

9. 同質性檢驗: 因素負荷量

- 「因素負荷量」(factor loading) 表示題項與因素(心理特質)關係的程度。
- 因素負荷量愈低，表示題項與共同因素(總量表)的關係不密切，同質性低。

```
> # factor loading: eigevectors
> round(sp_pca$loadings[,1], 4)
    b1     b2     b3     b4     b5     b6     b7     b8     b9    b10    b11
0.2057 0.1877 0.0082 0.2013 0.2494 0.2567 0.2733 0.2689 0.2797 0.0485 0.1970
...
> # factor loading: corr(x, comp)
> round(sapply(sp_data_used, function(x) cor(x, sp_pca$scores[, 1])), 4)
    b1     b2     b3     b4     b5     b6     b7     b8     b9    b10    b11
0.5931 0.5414 0.0238 0.5805 0.7193 0.7403 0.7881 0.7753 0.8065 0.1399 0.5680
...
> # sdev: the standard deviations of the principal components. (eigenvalues)
> sp_pca$sdev^2
   Comp.1    Comp.2    Comp.3    Comp.4    Comp.5    Comp.6    Comp.7    Comp.8
8.3164685 1.2914863 1.1209754 1.0342238 0.9420091 0.8639734 0.7361886 0.5962169
...
> # the proportion of variance explained by each PC.
> summary(sp_pca)
Importance of components:
                          Comp.1     Comp.2    Comp.3     Comp.4     Comp.5
Standard deviation     2.8838288 1.13643578 1.0587612 1.01696793 0.97057151
Proportion of Variance 0.4377089 0.06797296 0.0589987 0.05443283 0.04957942
Cumulative Proportion  0.4377089 0.50568183 0.5646805 0.61911337 0.66869279
...
                          Comp.16    Comp.17   Comp.18     Comp.19
Standard deviation     0.55492788 0.48030179 0.4718696 0.408929855
Proportion of Variance 0.01620763 0.01214157 0.0117190 0.008801243
Cumulative Proportion  0.96733819 0.97947976 0.9911988 1.000000000
```

大於1的共同因素有4個。

項目分析後之量表

題項	極端組比較	題項與總分相關		同質性檢驗		
	決斷值	題項與總分相關	校正題項與總分相關	題項刪除後的α值	共同性	因素負荷量
判標準則	≥3.000	≥.400	≥.400	≤量表信度值	≥.200	≥.450

【社會參與量表】

	非常同意	大部分同意	一半同意	大部分不同意	非常不同意	
01.我常會喜歡宗教活動中的一些儀式	□	□	□	□	□	✓
02.我不覺得參加社會服務工作很有意義	□	□	□	□	□	✓
03.我常會選擇自己喜歡的社團活動來參與	□	□	□	□	□	✗
04.我常會參加各種進修學習活動	□	□	□	□	□	✓
05.我常會主動地參加一些休閒娛樂活動	□	□	□	□	□	✓
06.我常會參加一些宗教活動	□	□	□	□	□	✓
07.我覺得參加社會服務工作之後使我的生活更加充實	□	□	□	□	□	✓
08.我參與社團活動時我會積極的投入	□	□	□	□	□	✓
09.我參加學習活動時，常會受到老師和同儕的肯定	□	□	□	□	□	✓
10.我現在常會邀請家人或親友一起從事運動休閒	□	□	□	□	□	✗
11.我常會鼓勵親朋好友一起參加宗教活動	□	□	□	□	□	✓
12.我很願意奉獻自己的專長和經驗來服務別人	□	□	□	□	□	✓
13.我常會做些適合我的運動來增進健康	□	□	□	□	□	✓
14.我參加學習活動時心情都很愉快	□	□	□	□	□	✓
15.我不覺得從事適當的休閒運動後能讓生活更充實	□	□	□	□	□	✓
16.參加宗教活動之後讓我的心靈更為充實	□	□	□	□	□	✓
17.我常會利用時間參加各種社會服務工作	□	□	□	□	□	✓
18.我不覺得我能從社團活動的參與過程中獲得滿足感	□	□	□	□	□	✓
19.我會主動和別人分享進修學習的心得	□	□	□	□	□	✓

✓：題項保留　✗：題項刪除

	非常同意	大部分同意	一半同意	大部分不同意	非常不同意	
01.我常會喜歡宗教活動中的一些儀式	□	□	□	□	□	✓
02.我不覺得參加社會服務工作很有意義	□	□	□	□	□	✓
03.我常會參加各種進修學習活動	□	□	□	□	□	✓
04.我常會主動的參加一些休閒娛樂活動	□	□	□	□	□	✓
05.我常會參加一些宗教活動	□	□	□	□	□	✓
06.我覺得參加社會服務工作之後使我的生活更加充實	□	□	□	□	□	✓
07.我參與社團活動時我會積極地投入	□	□	□	□	□	✓
08.我參加學習活動時，常會受到老師和同儕的肯定	□	□	□	□	□	✓
09.我常會鼓勵親朋好友一起參加宗教活動	□	□	□	□	□	✓
10.我很願意奉獻自己的專長和經驗來服務別人	□	□	□	□	□	✓
11.我常會做些適合我的運動來增進健康	□	□	□	□	□	✓
12.我參加學習活動時心情都很愉快	□	□	□	□	□	✓
13.我不覺得從事適當的休閒運動後能讓生活更充實	□	□	□	□	□	✓
14.參加宗教活動之後讓我的心靈更為充實	□	□	□	□	□	✓
15.我常會利用時間參加各種社會服務工作	□	□	□	□	□	✓
16.我不覺得我能從社團活動的參與過程中獲得滿足感	□	□	□	□	□	✓
17.我會主動和別人分享進修學習的心得	□	□	□	□	□	✓

試題分析:
二元/多元計分試題題項分析

- 古典測驗理論 (Classical Test Theory Functions, CTT)及題目反應理論 (Item Response Theory, IRT)。
- 古典測驗理論Classical Test Theory (CTT): 測量學生能力的方式是透過學生的得分或者是在測驗中的答對率。
  - **題目難(易)度 (Item Difficulty)** : 答對某一題的人數百分比。
  - **測驗信度(Test Reliability)** : 測驗信度是指測驗結果的一致性和穩定性，即測驗在不同時間或不同情況下，是否能夠提供一致的測量結果。在CTT中，常用的信度係數包括克隆巴赫α係數（Cronbach's alpha），斯皮爾曼-布朗係數（Spearman-Brown coefficient）等。信度越高，表示測驗結果越穩定、可信。

$$Cronbach's\ alpha\ = \frac{K}{K-1}\left(1-\frac{\sum_{k=1}^{K}\sigma_k^2}{\sigma_T^2}\right),$$

K: 總題數; σ_k^2: 第k題的變異數; σ_T^2: 總分的變異數

  - **題目鑑別度** (Item Discrimination): 該題得分與整個測驗的總得分之間的線性相關係數，用來評估一個題目的品質。如果學生亂猜某題答案，該相關係數會非常接近0。一個題目的鑑別度如果很高，這反映出高能力的學生在該題普遍答對得1分，而低能力的學生在該題普遍答錯得0分。常用的指標有: (1)該題高分組與低分組答對百分比之差及(2) 點二列相關係數 (point-biserial correlation) 。

- CTT架構之下，一個題目或測驗的統計量是建立在**測驗總分**的概念上。若一群學生的測驗題目，與另外一群學生測驗題目不同，他們的總得分是無法比較的，此時，不適合採用 CTT 理論來進行分析。

點二列相關係數 (Point-Biserial Correlation)

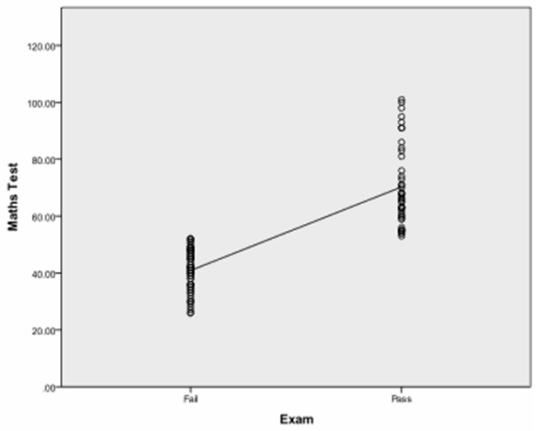
- **The Point-Biserial Correlation** (r-pbis) is a special case of the Pearson Correlation (ranges from −1.0 to 1.0) and is used when you want to measure the relationship between a continuous variable (y) and a dichotomous variable (x), or one that has two values (i.e. male/female, yes/no, true/false).

$$r_{pb} = \frac{M_0 - M_1}{s_y}\sqrt{\frac{n_0}{n}\frac{n_1}{n}}$$

n: total number of observations
n_0 (n_1): number of observations in group 0 (group 1).
M_0 (M_1): mean of the continuous variable for the group 0 (group 1).
S_y: standard deviation of the continuous variable.

  - In a Classical Test Theory, **r-pbis** measures the discrimination or differentiating strength, of the item.
  - It is a correlation of item scores and total raw scores.
  - Consider a scored data matrix (multiple-choice items converted to 0/1 data), r-pbis is the correlation between the item column and a column that is the sum of all item columns for each row (a person's score).

- **Assumptions**
  - No outliers (continuous variable).
  - Approximately normally distributed (continuous variable).
  - Homogeneity of variance of the continuous variable between both groups of the dichotomous variable.



二元計分試題題項分析

```
> exam1  <- read.csv("data/exam1_student_answer.csv")
> head(exam1)
     ID P01 P02 P03 P04 P05 P06 P07 P08 P09 P10 P11 P12 P13 P14 P15 P16
1   ANS   3   2   4   2   2   3   4   4   1   2   1   4   1   2   3   1
2 ST001   3   2   4   3   4   3   4   4   1   3   2   4   3   2   4   1
...
  P35 P36 P37 P38 P39 P40
1   2   4   2   2   2   1
2   2   4   2   2   2   1
...
> tail(exam1)
> dim(exam1)
[1] 45 41
> n <- nrow(exam1) - 1
> p <- ncol(exam1) - 1
> exam1_response <- exam1[-1, -1]
> exam1_solution <- exam1[1, -1]
> exam1_ID <- exam1[-1, 1]
> exam1_01 <- t(apply(exam1_response, 1, function(x) ifelse(x - exam1_solution == 0, 1, 0)))
> colnames(exam1_01) <- colnames(exam1_solution)
> rownames(exam1_01) <- exam1_ID
> head(exam1_01)
      P01 P02 P03 P04 P05 P06 P07 P08 P09 P10 P11 P12 P13 P14 P15 P16 P17
ST001   1   1   1   0   0   1   1   1   1   0   0   1   0   1   0   1   0
...
      P35 P36 P37 P38 P39 P40
ST001   1   1   1   1   1   1
...
> rowSums(exam1_01) # total score
ST001 ST002 ST003 ST004 ST005 ST006 ST007 ST008 ST009 ST010 ST011 ST012
   29    20    31    29    36    20    18    25    31    35    36    31
...
```

- 測驗(考試)最主要的目的: (1) 了解學生在所測驗領域的能力程度。(2)了解測驗與題目的相關特徵(例如: 題目難度(題目鑑別度)、測驗是否能區分學生的能力程度(測驗信度))
- 計算(低分組、高分組、分組)難度、鑑別度、CR值、試題與總分相關係數、刪題後之alpha值。

誘答力分析 (Distractor Analysis)

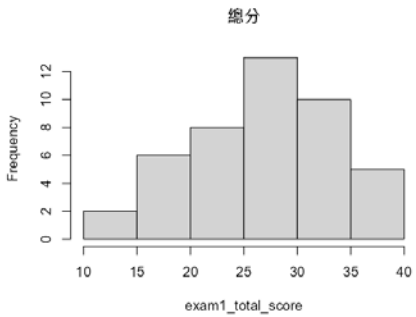
```
> library(CTT) # Classical Test Theory Functions
> distractor.analysis(exam1_response, exam1_solution)
You will find additional features and better formatting by using
distractorAnalysis().
$P01
        score.level
response lower middle upper
       1     2      0     1
       2     3      3     0
      *3     9      9    13
       4     1      2     1
...

> distractorAnalysis(exam1_response, exam1_solution, digits = 2)
$P01
  correct key  n rspP  pbis discrim lower mid50 mid75 upper
1           1  3 0.07 -0.28   -0.07  0.17  0.00  0.00   0.1
2           2  6 0.14 -0.25   -0.25  0.25  0.13  0.14   0.0
3       *   3 31 0.70  0.20    0.32  0.58  0.67  0.71   0.9
4           4  4 0.09 -0.02    0.00  0.00  0.20  0.14   0.0
...
```

- **correct**: An * indicates the correct response
- **key**: The response option being described
- **n**: The number of responsdents choosing that option
- **rspP**: The proportion of respondents with that response
- **pBis**: The point-biserial correlation between that reponse and the total score with that item removed
- **discrim**: The upper proportion minus the lower proportion
- **lower**: The proportion of respondents choosing that response that are from the lowest score group
- **upper**: The proportion of respondents choosing that response that are from the highest score group

計算難易度、鑑別度、CR值 、p值

```
> library(psych)
> describe(exam1_01) # mean is Item Difficulty
    vars  n mean   sd median trimmed mad min max range  skew kurtosis   se
P01    1 44 0.70 0.46      1    0.75   0   0   1     1 -0.87    -1.28 0.07
P02    2 44 0.66 0.48      1    0.69   0   0   1     1 -0.65    -1.61 0.07
...
P40   40 44 0.95 0.21      1    1.00   0   0   1     1 -4.22    16.15 0.03
```

```
> # 計算難易度、鑑別度、CR值 、p值
> exam1_total_score <- apply(exam1_01, 1, sum)
> exam1_total_score
ST001 ST002 ST003 ST004 ST005 ST006 ST007 ST008 ST009 ST010
   29    20    31    29    36    20    18    25    31    35
...
ST041 ST042 ST043 ST044
   24    21    29    22
> hist(exam1_total_score, main = "總分")
>
> bks <- quantile(exam1_total_score, probs = c(0, 0.27, 1 - 0.27, 1))
> bks
   0%   27%   73%  100%
14.00 21.61 31.00 40.00
> exam1_LMH <- cut(exam1_total_score, breaks = bks,
+                  labels = c("低分組(<= 21.61)", "不分組(21.61 - 31.00)", "高分組(> 31)"),
+                  include.lowest = T, right = F)
> table(exam1_LMH)
exam1_LMH
     低分組(<= 21.61) 不分組(21.61 - 31.00)          高分組(> 31)
                   12                    17                    15
```



計算難易度、鑑別度、CR值 、p值

```
> low_id <- exam1_LMH == "低分組(<= 21.61)"
> high_id <- exam1_LMH == "高分組(> 31)"
> LH_group <- factor(c(rep("L", length(which(low_id))),
+                      rep("H", length(which(high_id)))))
> exam1_01_LH <- rbind(exam1_01[low_id, ], exam1_01[high_id, ])
>
> exam1_mean_LH <- aggregate(exam1_01_LH, by = list(LH_group), mean)
> exam1_mean_LH <- t(exam1_mean_LH[, -1])
> head(exam1_mean_LH)
         [,1]      [,2]
P01 0.8666667 0.5833333
...
> my_t_test2 <- function(x){
+   is.var.equal <- var.test(x ~ LH_group)$p.value > 0.05
+   output <- t.test(x ~ LH_group, var.equal = is.var.equal)
+   round(c(output$statistic, output$p.value), 3)
+ }
> CR_p <- t(apply(exam1_01_LH, 2, my_t_test2))
>
> report <- data.frame(低分組難易度 = exam1_mean_LH[, 2],
+                      高分組難易度 = exam1_mean_LH[, 1],
+                      分組難度 = (exam1_mean_LH[, 2] + exam1_mean_LH[, 1]) / 2,
+                      鑑別度 = exam1_mean_LH[, 1] - exam1_mean_LH[, 2],
+                      CR值 = CR_p[, 1],
+                      p值 = CR_p[, 2])
> round(report, 3)
     低分組難易度  高分組難易度  分組難度  鑑別度  CR值   p值
P01        0.583        0.867   0.725  0.283 1.696 0.102
P02        0.583        0.733   0.658  0.150 0.801 0.431
...
P39        0.250        1.000   0.625  0.750 5.745 0.000
P40        0.833        1.000   0.917  0.167 1.483 0.166
```

修正的項目總相關
刪題後信度

```
> library(psychometric) # Applied Psychometric Theory
> exam1_item.exam <- item.exam(exam1_01, discrim = T)
> round(exam1_item.exam$Item.Tot.woi, 4) #Correlation of item with total test score (scored without item)
 [1] 0.1972 0.1112 0.5495 0.3466 0.2140 0.0907 0.1704 0.4394 0.2635
...
[37] 0.4355 0.2476 0.6588 0.3490
> item_corr <- sapply(1:40, function(x) cor(exam1_01[, x], rowSums(exam1_01[, -x])))
> names(item_corr) <- colnames(exam1_01)
> round(item_corr, 4)
   P01    P02    P03    P04    P05    P06    P07    P08    P09    P10
0.1972 0.1112 0.5495 0.3466 0.2140 0.0907 0.1704 0.4394 0.2635 0.1103
...
   P31    P32    P33    P34    P35    P36    P37    P38    P39    P40
0.2794 0.3389 0.3644 0.2988 0.2973 0.4917 0.4355 0.2476 0.6588 0.3490
```

```
> library(psych)
> reliability_output <- psych::alpha(exam1_01, check.keys = T)
> data.frame(總量表信度_刪題 = round(reliability_output$alpha.drop$raw_alpha, 3),
+            題目信度 = round(item.exam(exam1_01, discrim = T)$Item.Rel.woi, 3))
   總量表信度_刪題 題目信度
1            0.854    0.090
2            0.856    0.053
3            0.846    0.201
4            0.850    0.173
5            0.854    0.107
...
```

check.keys

if TRUE, then find the first principal component and reverse key items with negative loadings. Give a warning if this happens.

item.exam {psychometric}

item.exam {psychometric}

Item Analysis: Conducts an item level analysis. Provides item-total correlations, Standard deviation in items, difficulty, discrimination, and reliability and validity indices.

- **Sample.SD**: Standard deviation of the item (standard deviation (SD) of the observed score (0/1) on an item
- **Item.total**: Correlation of the item with the total test score
- **Item.Tot.woi**: Correlation of item with total test score (scored without item)
- **Difficulty**: Mean of the item (p) (proportion of examinees who responded correctly on the item)
- **Discrimination**: Discrimination of the item $(u - l)/n$. (the point-biserial correlation between item score (0/1) and total score)
- **Item.Criterion**: Correlation of the item with the Criterion (y)
- **Item.Reliab**: Item reliability index (題項一致性信度) (the product of item-score SD and discrimination)
- **Item.Rel.woi**: Item reliability index (scored without item) (the product of item-score SD and the correlation between the item score (0/1) and the criterion score (Y)

```
> head(exam1_item.exam)
    Sample.SD Item.total Item.Tot.woi Difficulty Discrimination Item.Criterion
P01 0.4615215  0.2645232   0.19723309  0.7045455      0.2857143             NA
P02 0.4794950  0.1832099   0.11115963  0.6590909      0.2142857             NA
P03 0.3699894  0.5883484   0.54950698  0.8409091      0.4285714             NA
...
    Item.Reliab Item.Rel.woi Item.Validity
P01  0.12068788   0.08998697            NA
P02  0.08684419   0.05269132            NA
P03  0.21519480   0.20098813
...
```

```
apply(exam1_01, 2, sd)  # item_sd

colMeans(exam1_01)  # difficulty

total_score <- rowSums(exam1_01)
apply(exam1_01, 2, function(x) cor(x, total_score))  # discrimination
```

itemAnalysis {CTT}

```
> exam1_iA <- itemAnalysis(exam1_01, hardFlag = 0.25, pBisFlag = 0.15)
> exam1_iA
 Number of Items
 40
 Number of Examinees
 44
 Coefficient Alpha
 0.854
> exam1_iA$itemReport
   itemName  itemMean       pBis       bis alphaIfDeleted hard lowPBis
1       P01 0.7045455 0.19723309 0.2606209      0.8537216
2       P02 0.6590909 0.11115963 0.1436576      0.8560502            X
3       P03 0.8409091 0.54950698 0.8291385      0.8460662
...
40      P40 0.9545455 0.34895802 0.7606776      0.8511936
> str(exam1_iA)
List of 6
 $ nItem     : int 40
 $ nPerson   : int 44
 $ alpha     : num 0.854
 $ scaleMean : num 27.2
 $ scaleSD   : num 6.55
 $ itemReport:'data.frame':      40 obs. of  7 variables:
  ..$ itemName      : chr [1:40] "P01" "P02" "P03" "P04" ...
  ..$ itemMean      : num [1:40] 0.705 0.659 0.841 0.523 0.523 ...
  ..$ pBis          : num [1:40] 0.197 0.111 0.55 0.347 0.214 ...
  ..$ bis           : num [1:40] 0.261 0.144 0.829 0.435 0.268 ...
  ..$ alphaIfDeleted: num [1:40] 0.854 0.856 0.846 0.85 0.854 ...
  ..$ hard          : chr [1:40] "" "" "" "" ...
  ..$ lowPBis       : chr [1:40] "" "X" "" "" ...
 - attr(*, "class")= chr "itemAnalysis"
```

hardFlag (easyFlag)
If a numeric value is provided, a flag is added to itemReport for each item with a mean less (greater) than the value. `itemReport = TRUE` must also be set.

itemReport:
Returned if `itemReport = TRUE`. Returns a data frame with key item analysis results: item mean (itemMean), point-biserial (pBis), biserial (bis), Cronbach's alpha if item removed, and any item flags indicated in the function call.