

十四、相關分析.....	2
14.1 相關分析問題型態.....	2
14.2 共變異數(covariance).....	3
14.2.1 母體/族群共變異數(population covariance) :	3
14.2.2 樣本共變異數(sample covariance).....	4
14.2.3 共變異數的意義.....	4
14.3 皮爾遜積差相關分析.....	5
14.3.1 皮爾遜積差相關係數(Pearson product-moment correlation coefficient)/積差相關係數.....	5
14.3.2 點二系列相關係數(Point-biserial correlation coefficient).....	6
14.3.3 史比爾曼等級相關係數(Spearman rank order correlation coefficient).....	6
14.4 相關分析 SPSS 操作方法.....	7
14.5 相關分析研讀報告.....	7

十四、相關分析

Correlation Analysis

14.1 相關分析問題型態

遊客的遊憩體驗、旅遊滿意度、遊憩動機、重遊意願是否有顯著的相關存在？

分析方法

X 變項 (自變數)	Y 變項 (依變數)	適用的相關方法
連續變項 Interval scale Ratio scale	連續變項 Interval scale Ratio scale	皮爾遜積差相關法(Pearson product-moment correlation)
Ordinal scale	Ordinal scale	史比爾曼等級相關(Spearman rank order correlation)
人爲二分變項	人爲二分變項	四分相關(tetrachoric correlation)
真正二分變項	真正二分變項	Φ (phi)相關
人爲二分變項	連續變項	二系列相關(biserial correlation)
真正二分變項 Bi-nominal scale	連續變項 Interval scale Ratio scale	點二系列相關(point-biserial correlation)

連續變項：變項屬性原來是 interval scale 或 ratio scale 的變數，如 Likert scale。若是認知或知識判斷上正確與否，原屬於 nominal scale 的變數，若其相關的認知或知識判斷的項目(問題)數兩個以上時，可將判斷正確給予 3(1)分，不正確給予 1(-1)分，未回答(不知道者)給予 2(0)分，總和計算每位受訪者在此部分的總分，此認知或知識判斷的「總分」可視為 ratio scale 的變數，繼續進行推論性統計分析。

分立變項：變項屬性原來是 nominal scale 或 ordinal scale 的變數。

真正二分變項：變項屬性原來就是二分類別變項或二分次序變項，如性別

人爲二分變項：變項屬性原來是 interval scale 或 ratio scale 的變數，經由人爲操控轉換為二分類別變項或二分次序變項。如學生成就測驗得分原來為連續變項，因研究需要，將成績分為『及格』與『不及格』二類。

Quantitative variables

Qualitative variables

Dichotomous dependent(so-called dummy variables)

Polytomous dependent

Table Correlation coefficients for two nominal variables

No. of categories		X Nominal	
		2	2+
Y Nominal	2	<ul style="list-style-type: none"> ■ Phi^a ■ Yule's Q ■ Lambda^b ■ Goodman and Kruskal's tau^c 	<ul style="list-style-type: none"> ■ Cramer's V ■ Lambda^b ■ Goodman and Kruskal's tau^c
	2+	<ul style="list-style-type: none"> ■ Cramer's V ■ Lambda^b ■ Goodman and Kruskal's tau^c 	<ul style="list-style-type: none"> ■ Cramer's V ■ Lambda^b ■ Goodman and Kruskal's tau^c

^aPhi is the 2x2 version of Cramer's V.

^bComes in both a symmetric and asymmetric version.

^cAsymmetric coefficient will have different values depending on which coefficient is treated as X.

Table Correlation coefficients for two ordinal variables

No. of categories		X Ordinal		
		2	2+	Many
Y Ordinal	2	<ul style="list-style-type: none"> ■ Gamma^a ■ Somers' d^c ■ Kendall's tau-b^d 	<ul style="list-style-type: none"> ■ Gamma^b ■ Somers' d^c ■ Kendall's tau-c^e 	<ul style="list-style-type: none"> ■ No obvious choice^a
	2+	<ul style="list-style-type: none"> ■ Gamma^a ■ Somers' d^c ■ Kendall's tau-c^e 	<ul style="list-style-type: none"> ■ Gamma^a ■ Kendall's tau-b^d 	<ul style="list-style-type: none"> ■ Kendall's rank-order tau^f
	Many	<ul style="list-style-type: none"> ■ No obvious choice^a 	<ul style="list-style-type: none"> ■ Kendall's rank-order tau^f 	<ul style="list-style-type: none"> ■ Spearman's rho ■ Kendall's rank-order tau^f

^aIf the variable with many values can be treated as interval, Pearson's r could be used.

^bGives higher correlations than other coefficients.

^cComes in a symmetric and asymmetric version.

^dUsed when the X and Y variables have the same number of categories.

^eUsed when the X and Y variables have a different number of categories.

^fMore appropriate than Spearman's rho where there are many tied ranks – a situation that occurs when one variable has relatively few categories.

Table Correlation coefficients for two interval variables

No. of categories		X Interval	
		2	2+
Y Interval	2	<ul style="list-style-type: none"> ■ Pearson's r ■ Phi^a ■ Biserial r^a 	<ul style="list-style-type: none"> ■ Pearson's r
	2+	<ul style="list-style-type: none"> ■ Pearson's r ■ Eta 	<ul style="list-style-type: none"> ■ Pearson's r ■ Eta^b

^aReduces to Pearson's r in the 2x2 case.

^bAsymmetric and sensitive to nonlinear relationships.

自變數與依變數(因變數)的選定，應依據研究架構和研究假設的檢定假設為基礎，進行自變數與依變數的檢定配對之排定。

積差相關法的基本假設

1. 受測樣本數最好在 25 人以上
2. 變項間均為連續變項(等距/比率變項)
3. 變項母群體均呈常態分佈
4. 二者相關型態為直線相關，而非曲線相關

14.2 共變異數(covariance)

14.2.1 母體/族群共變異數(population covariance) :

假設母體/族群數值為 $(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), \dots (x_N, y_N)$ ，則母體/族群共變異數 σ_{XY} 或 $\text{Cov}(X, Y)$

$$\sigma_{XY} = \text{Cov}(X, Y) = \frac{\sum_{i=1}^N (x_i - \mu_X)(y_i - \mu_Y)}{N}$$

μ_X : X 變數的算術平均值

μ_Y : Y 變數的算術平均值

14.2.2 樣本共變異數(sample covariance)

假設樣本數值為 $(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), \dots (x_n, y_n)$ ，則樣本共變異數 S_{XY}

$$S_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

\bar{x} : X 變數的算術平均值

\bar{y} : Y 變數的算術平均值

Sample: 假設某旅行社 A、B 兩種套裝行程，在過去半年的銷售數量

月份	1	2	3	4	5	6	平均數
A	5	6	8	4	5	7	5.83
B	8	12	15	13	11	11	11.67

求此兩種套裝行程的樣本共變異數

$$\begin{aligned} S_{AB} &= \frac{\sum_{i=1}^n (A_i - \bar{A})(B_i - \bar{B})}{n-1} = \frac{\sum_{i=1}^6 (A_i - 5.83)(B_i - 11.67)}{6-1} \\ &= \frac{(5-5.83)(8-11.67) + \dots + (7-5.83)(11-11.67)}{5} \\ &= \frac{3.06 + 0.06 + 7.22 - 2.44 + 0.55 - 0.78}{5} = \frac{7.67}{5} = 1.53 \end{aligned}$$

14.2.3 共變異數的意義

I $X - \mu_X < 0$ $Y - \mu_Y > 0$	II $X - \mu_X > 0$ $Y - \mu_Y > 0$
III $X - \mu_X < 0$ $Y - \mu_Y < 0$	IV $X - \mu_X > 0$ $Y - \mu_Y < 0$

$S_{XY} > 0$: 代表 X, Y 同時比平均值大，或 X, Y 同時比平均值小的出現頻率較多，即大部分數值均落於 II、III 兩區域(向限)，故 X, Y 數值之間具有正向線性關係。

$S_{XY} < 0$: 代表 X 大於平均值且 Y 小於平均值，或 X 小於平均值且 Y 大於平均值的出現頻率較多，即大部分數值均落於 I、IV 兩區域(向限)，故 X, Y 數值之間具有負向線性關係。

$S_{XY} = 0$: 代表 X, Y 數值平均分散在 I、II、III、IV 四個區域，故 X, Y 數值之間沒有線性關係。

Reference:

Kim, S. S. and Crompton, J. L. 2001. The effects of different types of information messages on perceptions of price and stated willingness-to-pay. *Journal of Leisure Research*, 33(3):299-318.

14.3 皮爾遜積差相關分析

以皮爾遜積差相關方法分析兩者的相關程度，積差相關係數可作為兩個連續變數間線性相關的指標。

1. **相關係數**介於-1 與+1 之間，正負符號表示相關的方向，負相關表示線性相關的斜率為負，正相關表示線性相關的斜率為正。
2. 相關係數(r)的平方(r²)成爲**決定係數**或**解釋變異量的比例**。
3. 在統計分析中，相關係數的意義與樣本人數大小有關，在推論統計中，若受測的樣本很多，即使相關係數的值很小，也很容易達到顯著。因而在相關分析的解釋過程，除說明兩個變項是否達顯著相關外，也應呈現決定係數的大小，並加以說明。
4. 不論相關係數或決定係數只能說明兩者關係密切的程度，而不能誤認兩者間有因果關係。
5. 若 X 變項與 Y 變項的相關爲 0.50 (p<0.001)，決定係數爲 0.25，意謂著「Y 變項的變異量中，可被 X 變項解釋的變異量百分比爲 25 %」；相對的，也意謂著「X 變項的變異量中，可被 Y 變項解釋變異量百分比也爲 25 %」；而相關係數等於 0.50，則表示 X 變項與 Y 變項間有顯著的正相關。

相關係數(r)	相關程度
0.8 以上	極高
0.6-0.8	高
0.4-0.6	普通
0.2-0.4	低
0.2 以下	極低

14.3.1 皮爾遜積差相關係數(Pearson product-moment correlation coefficient)/積差相關係數

$$\begin{aligned} r_{XY} &= \frac{COV_{XY}}{S_X S_Y} = \frac{XY\text{的共變異數}}{(X\text{的標準差}) \times (Y\text{的標準差})} \\ &= \frac{\frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{n-1}}{\sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n-1}}} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)S_X S_Y} \\ &= \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}} \end{aligned}$$

Question 1: 母體/族群 vs. 樣本數值的 Pearson product-moment correlation coefficient 計算有何異同？

Question 2: 樣本數值加倍(重複)前後的 Pearson product-moment correlation coefficient 計算有何異同？

Sample: 假設某旅行社 A、B 兩種套裝行程，在過去半年的銷售數量

月份	1	2	3	4	5	6	平均數
A	5	6	8	4	5	7	5.83
B	8	12	15	13	11	11	11.67

求此兩種套裝行程的相關係數

$$\begin{aligned}
 r_{AB} &= \frac{\sum_{i=1}^6 (A_i - \bar{A})(B_i - \bar{B})}{\sqrt{\sum_{i=1}^6 (A_i - \bar{A})^2 \sum_{i=1}^6 (B_i - \bar{B})^2}} \\
 &= \frac{(5 - 5.83)(8 - 11.67) + \dots + (7 - 5.83)(11 - 11.67)}{\sqrt{[(5 - 5.83)^2 + \dots + (7 - 5.83)^2][(8 - 11.67)^2 + \dots + (11 - 11.67)^2]}} \\
 &= \frac{7.67}{\sqrt{10.83 \times 27.33}} = \frac{7.67}{17.21} = 0.4455
 \end{aligned}$$

14.3.2 點二系列相關係數(Point-biserial correlation coefficient)

$$r_{AB} = \frac{\bar{X}_A - \bar{X}_B}{S_t} \sqrt{AB}$$

\bar{X}_A : A 族群(男生)某特徵數值(身高、成績、interval scale、ratio scale)的平均值

\bar{X}_B : B 族群(女生)某特徵數值(身高、成績、interval scale、ratio scale)的平均值

S_t : A 和 B 族群(全部)某特徵數值(身高、成績、interval scale、ratio scale)的標準差

A : A 族群(男生)所佔受訪者的比率(0~1)

B : B 族群(男生)所佔受訪者的比率(0~1)

14.3.3 史比爾曼等級相關係數(Spearman rank order correlation coefficient)

$$r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

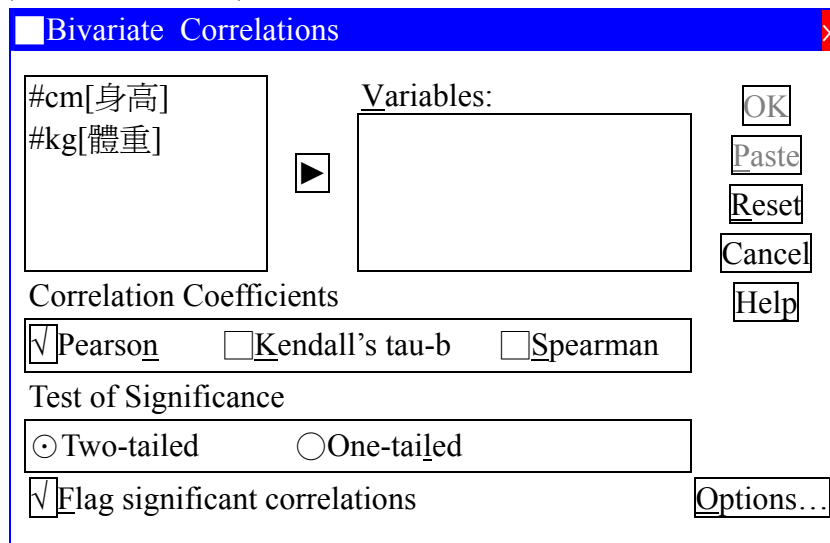
d : 等第差距值

顯著性檢定公式

$$t = \frac{r_s}{\sqrt{\frac{1 - r_s^2}{nN - 2}}}$$

14.4 相關分析 SPSS 操作方法

1. Analyze/Statistics(統計分析) → Correlate(相關) → Bivariate... (雙變數)，打開 Bivariate correlations (雙變數相關分析)對話視窗



2. 出現 Bivariate Correlations (雙變數相關分析)對話視窗
3. 將左邊方塊中欲進行相關分析的變數名稱，勾選進入右邊 Variables:方塊中
4. 在 Correlation Coefficients(相關係數)方塊中勾選「Pearson」(積差相關係數)
5. 在 Test of Significance(顯著性檢定)方塊中勾選「Two-tailed」(雙尾檢定，用於探索性的資料分析)
6. 勾選最下面的「Flag significant correlations」(相關顯著性訊號)，積差相關檢定結果達 0.05 顯著水準，在相關係數旁會以一個「*」符號表示，達 0.01 顯著水準時會以兩個「*」符號表示
7. 按 按鈕，以執行相關分析

Correlations

	cm	Kg
cm Pearson Correlation	1.000	.996**
Sig. (2-tailed)	.	.000
N	7	7
kg Pearson Correlation	.996**	1.000
Sig. (2-tailed)	.000	.
N	7	7

** Correlation is significant at the 0.01 level (2-tailed).

14.5 相關分析研讀報告

- Kim, S. S., Lee, C. K. & Klenosky, D. B. 2003. The influence of push and pull factors at Korean national parks. *Tourism Management*, 24:169-180.
- Redmond, E. C. & Griffith, C. J. 2005. Consumer perceptions of food safety education sources: implications for effective strategy development. *British Food Journal*, 107(7):467-483.