

第9章 二元隨機變數及其機率分配 應用統計學 四版

9 二元隨機變數及其機率分配

學習目的

1. 定義或了解二元間斷隨機變數與連續隨機變數的意義及其機率分配。
2. 瞭解邊際機率分配與條件機率分配。
3. 瞭解兩變數間的關係。
4. 瞭解二元隨機變數函數的期望值。
5. 瞭解多元隨機變數。

林嘉坤 陳正倉著 雙葉書局發行 2009

第9章 二元隨機變數及其機率分配 應用統計學 四版

本章結構

二元隨機變數及其機率分配

- 二元間斷隨機變數
 - 聯合機率函數
 - 邊際機率函數
 - 條件機率函數
 - 二元間斷隨機變數的期望值與變異數
- 兩變數間的關係
 - 兩變數獨立
 - 兩變數不獨立—共變與相關
 - 共變數
 - 相關係數
 - 相關係數、共變數與獨立的關係
- Excel 的使用

林嘉坤 陳正倉著 雙葉書局發行 2009

第9章 二元隨機變數及其機率分配 應用統計學 四版

表9.1 初任人員經常性薪資—按教育程度分 2006年半年平均，單位：元

行業別\學歷	國中及以下	高中職	專科	大學	研究所
礦業及土石採取業	20,253	21,580	24,015	28,347	32,316
製造業	19,118	20,729	23,684	26,502	31,188
水電熱氣業	23,212	26,340	31,177	33,506	40,167
營造業	21,815	22,400	24,661	26,800	30,511
批發及零售業	18,467	20,056	22,432	25,026	28,500
住宿及餐飲業	18,621	19,704	21,714	23,194	26,179
運輸、倉儲及通信業	20,779	22,077	24,685	27,669	32,587
金融及保險業	21,896	23,799	27,070	30,569	36,466
不動產及租賃業	19,591	20,440	23,495	26,459	29,489
專業、科學及技術服務業	17,342	19,959	24,086	27,410	33,470
醫療保健服務業	19,992	21,458	25,508	28,876	34,397
文化、運動及休閒服務業	18,387	19,811	22,659	24,871	28,316
其他服務業	18,315	19,656	22,143	24,149	28,002

林嘉坤 陳正倉著 雙葉書局發行 2009

第9章 二元隨機變數及其機率分配 應用統計學 四版

二元隨機變數

○二元隨機變數

二個或二個以上相互有關的事件可定義為二元或多元隨機變數 (bivariate or multivariate random variables)。

林嘉坤 陳正倉著 雙葉書局發行 2009

第9章 二元隨機變數及其機率分配 應用統計學 四版

二元間斷隨機變數

○聯合機率函數

設 X, Y 為二元間斷隨機變數， X 之值為 x_1, x_2, \dots, x_n ， Y 之值為 y_1, y_2, \dots, y_m ，若 $f(x, y)$ 滿足機率的二條件：

- ① $0 \leq f(x, y) \leq 1$
- ② $\sum_x \sum_y f(x, y) = 1$

則 $f(x, y)$ (簡單表示為 $f(x, y)$) 為聯合機率函數。

林嘉坤 陳正倉著 雙葉書局發行 2009

第9章 二元隨機變數及其機率分配 應用統計學 四版

二元間斷隨機變數

○ X 的邊際機率函數

$$f(x) = \sum_y f(x, y) = f(x, y_1) + f(x, y_2) + \dots + f(x, y_m)$$

上式必須滿足下列兩條件：

- ① $0 \leq f(x) \leq 1$
- ② $\sum_x f(x) = 1$

林嘉坤 陳正倉著 雙葉書局發行 2009

第9章 二元隨機變數及其機率分配 應用統計學 四版

二元間斷隨機變數

○ Y 的邊際機率函數

$$f_y(y) = \sum_x f(x, y) = f(x_1, y) + f(x_2, y) + \dots + f(x_n, y)$$

上式必須滿足下列兩條件：

① $0 \leq f_y(y_j) \leq 1$

② $\sum_j f_y(y_j) = 1$

林嘉坤 陳正倉著 雙葉書局發行 2009

第9章 二元隨機變數及其機率分配 應用統計學 四版

二元間斷隨機變數

表9.2 X 與 Y 的聯合機率分配與邊際機率分配表

$X \setminus Y$	y_1	y_2	...	y_j	...	y_m	$f_x(x_i)$
x_1	$f(x_1, y_1)$	$f(x_1, y_2)$...	$f(x_1, y_j)$...	$f(x_1, y_m)$	$f_x(x_1)$
x_2	$f(x_2, y_1)$	$f(x_2, y_2)$...	$f(x_2, y_j)$...	$f(x_2, y_m)$	$f_x(x_2)$
...
x_i	$f(x_i, y_1)$	$f(x_i, y_2)$...	$f(x_i, y_j)$...	$f(x_i, y_m)$	$f_x(x_i)$
...
x_n	$f(x_n, y_1)$	$f(x_n, y_2)$...	$f(x_n, y_j)$...	$f(x_n, y_m)$	$f_x(x_n)$
$f_y(y_j)$	$f_y(y_1)$	$f_y(y_2)$...	$f_y(y_j)$...	$f_y(y_m)$	1

林嘉坤 陳正倉著 雙葉書局發行 2009

第9章 二元隨機變數及其機率分配 應用統計學 四版

二元間斷隨機變數

○ 條件機率函數

設 $f(x, y)$ 為二元機率函數，則在 $Y = y$ 條件下，發生 x 的條件機率表為：

$$f(x_i | Y = y_j) = \frac{f(x_i, y_j)}{f_y(y_j)}$$

在 $X = x_i$ 條件下，發生 y_j 的條件機率表為：

$$f(y_j | X = x_i) = \frac{f(x_i, y_j)}{f_x(x_i)}$$

林嘉坤 陳正倉著 雙葉書局發行 2009

第9章 二元隨機變數及其機率分配 應用統計學 四版

二元間斷隨機變數

表9.3 和碩電訊最新型3G手機的銷售資料

		Y (是否看過簡訊)		合計
		0 (未看過)	1 (看過)	
X (有否購買3G手機)	0 (不買)	198	132	330
	1 (購買)	108	162	270
合計		306	294	600

林嘉坤 陳正倉著 雙葉書局發行 2009

第9章 二元隨機變數及其機率分配 應用統計學 四版

二元間斷隨機變數

表9.4 X 與 Y 的聯合機率分配表

		Y (是否看過簡訊)		合計 $f_x(x)$
		0 (未看過)	1 (看過)	
X (是否購買手機)	0 (不買)	0.33	0.22	0.55
	1 (購買)	0.18	0.27	0.45
合計 $f_y(y)$		0.51	0.49	1.00

林嘉坤 陳正倉著 雙葉書局發行 2009

第9章 二元隨機變數及其機率分配 應用統計學 四版

二元間斷隨機變數

表9.5 $f(x|y)$ 的條件機率

$f(x y)$	$f(x Y=0)$ (未看過)	$f(x Y=1)$ (看過)
$X=0$ (不買)	0.65	0.45
$X=1$ (購買)	0.35	0.55

林嘉坤 陳正倉著 雙葉書局發行 2009

第9章 二元隨機變數及其機率分配 應用統計學 四版

二元間斷隨機變數

表9.6 $f(y|x)$ 的條件機率

$f(y x)$	$f(y X=0)$ (不買)	$f(y X=1)$ (購買)
$Y=0$ (未看過)	0.60	0.40
$Y=1$ (看過)	0.40	0.60

林嘉坤 陳正倉著 雙葉書局發行 2009

第9章 二元隨機變數及其機率分配 應用統計學 四版

二元間斷隨機變數

二元間斷隨機變數的期望值

○ X 的期望值

$$E(X) = \sum_x \sum_y xf(x, y)$$

$$= \sum_x x \sum_y f(x, y) = \sum_x xf_1(x)$$

○ Y 的期望值

$$E(Y) = \sum_x \sum_y yf(x, y)$$

$$= \sum_y y \sum_x f(x, y) = \sum_y yf_2(y)$$

林嘉坤 陳正倉著 雙葉書局發行 2009

第9章 二元隨機變數及其機率分配 應用統計學 四版

二元間斷隨機變數

二元間斷隨機變數的變異數

○ X 的變異數

$$V(X) = \sum_x \sum_y (x - \mu_x)^2 f(x, y)$$

$$= \sum_x (x - \mu_x)^2 f_1(x)$$

$$= \sum_x x^2 f_1(x) - \mu_x^2$$

○ Y 的變異數

$$V(Y) = \sum_x \sum_y (y - \mu_y)^2 f(x, y)$$

$$= \sum_y (y - \mu_y)^2 f_2(y)$$

$$= \sum_y y^2 f_2(y) - \mu_y^2$$

林嘉坤 陳正倉著 雙葉書局發行 2009

第9章 二元隨機變數及其機率分配 應用統計學 四版

兩變數間的關係

表9.7 廣告秒數與銷售收入

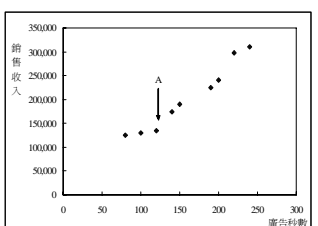
	A	B	C
1	星期	廣告秒數	銷售收入
2	1	120	135,000
3	2	200	240,000
4	3	100	130,000
5	4	150	190,000
6	5	240	310,000
7	6	80	125,000
8	7	220	298,000
9	8	140	175,000
10	9	190	225,000

林嘉坤 陳正倉著 雙葉書局發行 2009

第9章 二元隨機變數及其機率分配 應用統計學 四版

兩變數間的關係

圖9.1 簡訊發放量與銷售收入的散佈圖



林嘉坤 陳正倉著 雙葉書局發行 2009

第9章 二元隨機變數及其機率分配 應用統計學 四版

兩變數間的關係

○ 兩變數獨立的條件

設 X, Y 為二元隨機變數，若 X 與 Y 之值均滿足下列任一條件，則 X, Y 獨立。

- ① $f(x, y) = f_1(x) \cdot f_2(y)$
- ② $f(x|y) = f_1(x)$
- ③ $f(y|x) = f_2(y)$

林嘉坤 陳正倉著 雙葉書局發行 2009

兩變數間的關係

表9.8 繫安全帶與車禍受傷程度統計表

X (受傷程度)	Y (繫安全帶)		合 計
	0 (無)	1 (有)	
1 (輕傷)	380	170	550
2 (重傷)	240	50	290
3 (死亡)	140	20	160
合 計	760	240	1,000

資料來源：參考保險相關報告，數字為虛擬。

兩變數間的關係

表9.9 X與Y的聯合機率分配表

X (受傷程度)	Y (繫安全帶)		合 計 $f_{1j}(x)$
	0 (無)	1 (有)	
1 (輕傷)	0.38	0.17	0.55
2 (重傷)	0.24	0.05	0.29
3 (死亡)	0.14	0.02	0.16
合 計 $f_{2j}(y)$	0.76	0.24	1.00

兩變數間的關係

○ 共變數

共變數是用來衡量兩變數變動的方向及其程度。

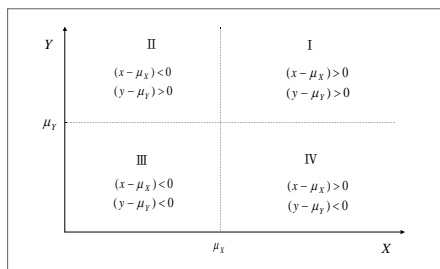
兩變數間的關係

○ 共變數

$$Cov(X, Y) = E[(X - \mu_x)(Y - \mu_y)]$$

兩變數間的關係

圖9.2 $Cov(X, Y)$ 的符號



兩變數間的關係

圖9.3 正向共變

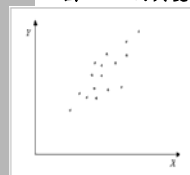


圖9.4 反向共變

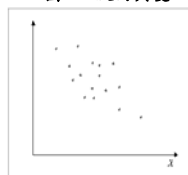


圖9.5 無關係



兩變數間的關係

○ 共變數的優缺點

共變數在衡量共變關係時，有下列幾個優缺點：

- ① 它可看出共變的方向與程度。
- ② $-\infty < Cov(X, Y) < \infty$ ，其值域無限，不易根據 $Cov(X, Y)$ 的大小來判斷其相關程度。
- ③ 它易受衡量單位的影響，例如當以公斤改為以磅衡量體重時，所得的共變數會改變。
- ④ 具有雙重的衡量單位（同時有 X 與 Y 的衡量單位不易解釋）。

兩變數間的關係

○ 相關係數

$$\rho_{XY} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = E\left(\frac{X - \mu_X}{\sigma_X}\right)\left(\frac{Y - \mu_Y}{\sigma_Y}\right)$$

式中： $Cov(X, Y)$ 為共變數， σ_X ， σ_Y 為標準差。

兩變數間的關係

○ 相關係數的數值

① 相關係數的數值可證明是介於 -1 與 $+1$ 之間。

② 當 $\rho_{XY} = +1$ 時，我們稱 X 、 Y 具正的完全直線關係，此時 (X, Y) 的點會落在一條斜率為正的直線上，如圖 8.6。當 $\rho_{XY} = -1$ 時，我們稱 X 、 Y 具負的完全直線關係，此時 (X, Y) 的點會落在一條斜率為負的直線上，如圖 8.7。

③ 當 $\rho_{XY} = 0$ 時， X 、 Y 無線性相關。

圖 8.8 表 X 、 Y 無關。圖 8.9 表示 X 、 Y 無線性相關，但為拋物線關係。圖 8.10 表示 X 、 Y 無線性相關，為圓形關係。圖 8.11 表 X 、 Y 為非線性關係。

④ $|\rho_{XY}|$ 愈大表示 X 、 Y 的線性相關程度愈大。一般而言，若相關係數的數值在 0.5 左右稱為中度相關；若相關係數大於 0.7 ，則認為高度相關；若小於 0.3 則為低度相關。

兩變數間的關係

圖9.6 相關係數 $\rho_{XY} = 1$

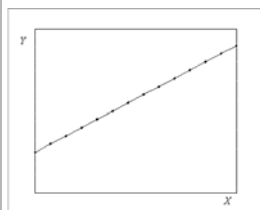
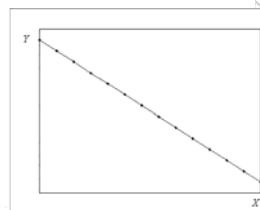


圖9.7 相關係數 $\rho_{XY} = -1$



兩變數間的關係

圖9.8 X、Y無關

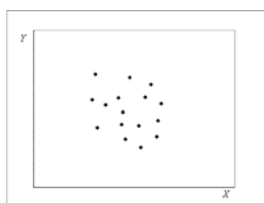


圖9.9 X、Y、為拋物線關係



兩變數間的關係

圖9.10 X、Y、為圓形關係

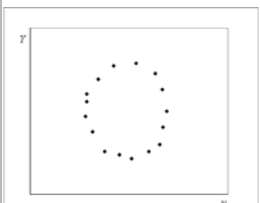


圖9.11 X、Y、為非線性關係



兩變數間的關係

初入職場者教育程度與薪資的關係

	A	B	C	D	E	F
1	X (教育程度)	國中以下	高中職	專科	大學	研究所
2	Y (薪資)	21,815	22,400	24,661	26,800	30,511

兩變數間的關係

圖9.12 教育程度與薪資的關係圖

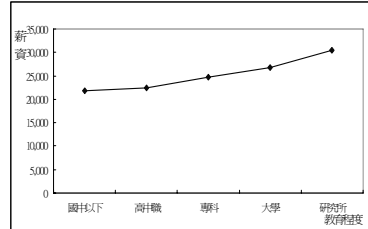


表9.10 廣告秒數與銷售收入之相關係數

	A	B	C
1		廣告秒數	銷售收入
2	廣告秒數	1	
3	銷售收入	0.978463943	1

表9.11 廣告秒數與銷售收入之共變數

	A	B	C
1		廣告秒數	銷售收入
2	廣告秒數	3075	
3	銷售收入	3803750	4914611111