التوزيعات الاحتمالية المشتركة
Joint distributions

مقدمــة

فيما سبق تم عرض التوزيعات الاحتمالية لمتغير عشوائي واحد ، إلا أنه في كثير من النواحي التطبيقية يهتم الباحث بدراسة شكل التوزيع الاحتمالي بين متغيرين ، لمعرفة العلاقة بينهما، أو التنبؤ بأحدها من الآخر.

التوزيع الاحتمالي المشترك لمتغيرين متقطعين Joint distribution for two discrete Variables

بفرض أن (X, Y) متغيرين عشوائيين متقطعين ، حيث أن المتغير (X) يأخذ القيم المنفصلة $(X : x = x_1, x_2, \ldots, x_r)$، والمتغير (Y) يأخذ القيم المنفصلة $(Y : y = y_1, y_2, \ldots, y_c)$.

فإن دالة الاحتمال المشترك والتي تبين احتمال حدوث زوج القيم (x, y) معا تكتب على الصورة التالية:

$$Pr\,(x = x_i,\ y = y_j) = f\,(x_i, y_j) \qquad (1)$$

وتتصف هذه الدالة بالخاصيتين التاليتين:

$$0 < f\,(x_i, y_j) < 1$$

$$\sum_{j=1}^{c} \sum_{i=1}^{r} f\,(x_i, y_j) = 1$$

ويمكن من خلال دالة الاحتمال المشترك في معادلة (1) تكوين جدول التوزيع الاحتمالي المشترك للمتغيرين (X, Y)، وحساب احتمالات حدوث قيم المتغيرين معا ، والتغاير بين المتغيرين ، كما يمكن استنتاج الخصائص الإحصائية التالية:

- دوال الاحتمال الهامشية واستخدامها في حساب متوسط وتباين التوزيع الهامشي.
- التغاير بين المتغيرين.
- حساب معامل الارتباط البسيط بين المتغيرين.
- دالة الاحتمال الشرطية، واستخدامها في حساب متوسط وتباين التوزيع الشرطي.
- استنتاج دالة الانحدار من متوسط التوزيع الشرطي.

وسوف يتم عرض هذه النقاط من خلال التطبيق التالي.

تطبيق (1):

إذا كانت دالة الاحتمال المشتركة بين عدد الأطفال في الأسرة التي بها ما بين طفل إلى ثلاث

أطفال $(X : x = 1,2,3)$، وعدد الوحدات المستهلكة من حليب الأطفال الجاف من نوع معين كل أسبوعين $(Y : y = 0,1,2)$ تأخذ الصورة التالية.

$$f(x,y) = \frac{y - 0.5x + 2}{18}, \quad x = 1,2,3, \quad y = 0,1,2$$

كون جدول التوزيع الاحتمالي المشترك بين المتغيرين سالف الذكر ، ثم أوجد الآتي:

1- احتمال أن أسرة عدد أطفالها 2 ، واستهلاكها من الحليب وحدة واحدة أسبوعيا.

2- كون دوال التوزيع الهامشية لكلا المتغيرين، ثم احسب المتوسط والتباين لكل منها.

3- احسب معامل الارتباط بين عدد الأطفال وعدد الوحدات المستهلكة، وما هو مدلوله.

4- كون التوزيع الاحتمالي الشرطي لعدد الوحدات المستهلكة عند معلومية عدد الأطفال. وما هو متوسط عدد الوحدات المستهلكة إذا علم أن عدد الأطفال في الأسرة طفلان.

5- كون مصفوفة التباين والتغاير للمتغيرين.

الحل(1):

تكوين جدول التوزيع الاحتمالي:

بالتعويض في معادلة الاحتمال المشترك : $f(x,y) = \frac{y - 0.5x + 2}{18}$

عن عدد الأطفال x=1,2,3، عدد الوحدات المستهلكة y=0,1,2 نحصل على الاحتمالات التالية:

		عدد الوحدات المستهلكة y		
		0	1	2
عدد الأطفال x	1	$f(1,0) = \frac{0-0.5(1)+2}{18} = \frac{1.5}{18}$	$f(1,1) = \frac{1-0.5(1)+2}{18} = \frac{2.5}{18}$	$f(1,2) = \frac{2-0.5(1)+2}{18} = \frac{3.5}{18}$
	2	$f(2,0) = \frac{0-0.5(2)+2}{18} = \frac{1}{18}$	$f(2,1) = \frac{1-0.5(2)+2}{18} = \frac{2}{18}$	$f(2,2) = \frac{2-0.5(2)+2}{18} = \frac{3}{18}$
	3	$f(3,0) = \frac{0-0.5(3)+2}{18} = \frac{0.5}{18}$	$f(3,1) = \frac{1-0.5(3)+2}{18} = \frac{1.5}{18}$	$f(2,2) = \frac{2-0.5(3)+2}{18} = \frac{2.5}{18}$

ومن ثم يكتب جدول التوزيع الاحتمالي المشترك على الصورة التالية:

		عدد الوحدات المستهلكة y			$f(x)$
		0	**1**	**2**	
عدد الأطفال x	**1**	(1.5/18)	(2.5/18)	(3.5/18)	**(7.5/18)**
	2	(1/18)	(2/18)	(3/18)	**(6/18)**
	3	(0.5/18)	(1.5/18)	(2.5/18)	**(4.5/18)**
$f(y)$		**(3/18)**	**(6/18)**	**(9/18)**	**1**

1- احتمال أن أسرة عدد أطفالها 2 ، واستهلاكها من الحليب وحدة واحدة أسبوعيا هو:

$$Pr(x=2, y=1) = f(2,1) = \frac{1-0.5(2)+2}{18} = \frac{2}{18} = 0.11$$

أي أن حوالي 11% من الأسر عدد أطفالها 2 وعدد وحدات استهلاكها وحدة واحدة.

2- تكوين دوال التوزيع الهامشية لكلا المتغيرين، وحساب المتوسط والتباين لكل منها.

- دالة الاحتمال الهامشية لعدد الأطفال، أي للمتغير x تحسب بتطبيق المعادلة التالية:

$$Pr(X=x) = f(x) = \sum_{y}^{c} f(x,y) \qquad (4)$$

وبالتطبيق

$$f(x) = \sum_{y}^{c} f(x,y)$$

$$= \sum_{y=0}^{2} \frac{y-0.5x+2}{18} = \frac{0-0.5x+2}{18} + \frac{1-0.5x+2}{18} + \frac{2-0.5x+2}{18} = \frac{9-0.5x}{18}$$

$$Then\ : \quad f(x) = (9-1.5x)/18\ , \quad x = 1,2,3$$

- دالة الاحتمال الهامشية لعدد الوحدات المستهلكة، أي للمتغير الثاني y تحسب بتطبيق المعادلة التالية:

$$Pr(Y=y) = f(y) = \sum_{x}^{r} f(x,y) \qquad (5)$$

وبالتطبيق

$$f(y) = \sum_{x}^{r} f(x,y)$$

$$= \sum_{x=1}^{3} \frac{y-0.5x+2}{18} = \frac{y-0.5(1)+2}{18} + \frac{y-0.5(2)+2}{18} + \frac{y-0.5(3)+2}{18} = \frac{3y+3}{18}$$ ،

$$Then\ : \quad f(y) = (3y+3)/18\ , \quad y = 0,1,2$$

ولحساب المتوسط والتباين للمتغيرين تستخدم دوال التوزيع الهامشية الموضحة بالمعادلتين (4)، (5) ، ويتم تطبيق المعادلات التالية:

	عدد الأطفال في الأسرة x	عدد الوحدات المستهلكة y
المتوسط μ	$\mu_x = \sum_{x}^{r} x\, f(x)$	$\mu_y = \sum_{y}^{c} y\, f(y)$
التباين σ^2	$\sigma_x^2 = \sum_{x}^{r} x^2\, f(x) - \mu_x^2$	$\sigma_y^2 = \sum_{y}^{c} y^2\, f(y) - \mu_y^2$

كما يمكن أيضا اشتقاق التوزيعين الهامشين من جدول التوزيع الاحتمالي المشترك ، وتطبيق معادلتي المتوسط والتباين كما يلي:

$f(x)=(9-1.5x)/18$					$f(y)=(3y+3)/18$			
x	$f(x)$	$x\ f(x)$	$x^2\ f(x)$		y	$f(y)$	$y\ f(y)$	$y^2\ f(y)$
1	(7.5/18)	(7.5/18)	(7.5/18)		0	(3/18)	0	0
2	(6/18)	(12/18)	(24/18)		1	(6/18)	(6/18)	(6/18)
3	(4.5/18)	(13.5/18)	(40.5/18)		2	(9/18)	(18/18)	(36/18)
Total	1	**(33/18)**	**(72/18)**		Total	1	**(24/18)**	**(42/18)**

وبتطبيق معادلتي المتوسط والتباين:

	عدد الأطفال في الأسرة x	عدد الوحدات المستهلكة y
μ المتوسط	$\mu_x=(33/18)=1.83$	$\mu_y=(24/18)=1.33$
σ^2 التباين	$\sigma_x^2=\sum_x^r x^2\ f(x)-\mu_x^2$ $=(72/18)-(1.83)^2=0.6389$	$\sigma_y^2=\sum_y^c y^2\ f(y)-\mu_y^2$ $=(42/18)-(1.33)^2=0.5556$

أي أن متوسط عدد الأطفال في الأسرة التي بها ثلاث أطفال تقريبا 1.83 طفلا، بتباين 0.6389. بينما وجد أن متوسط عدد الوحدات المستهلكة للأسرة 1.33 علبة، بتباين 0.5556.

3- حساب معامل الارتباط بين عدد الأطفال وعدد الوحدات المستهلكة:

يبين معامل الارتباط نوع وقوة العلاقة بين متغيرين ، ولحساب معامل الارتباط بين المتغيرين (X,Y) ، ويرمز له بالرمز ρ_{xy}، تطبق المعادلة التالية:

$$\rho_{xy}=\frac{\sigma_{xy}}{\sigma_x\sigma_y} \tag{8}$$

حيث أن:

σ_{xy} هو التغاير بين المتغيرين (X,Y) ويحسب بالمعادلة التالية

$$\sigma_{xy}=\sum_y^c\sum_x^r xy\ \ f(x,y)-\mu_x\mu_y \tag{9}$$

ويمكن حسابه من جدول التوزيع الاحتمالي المشترك كالتالي:

حساب المجموع $\sum_{y}^{c}\sum_{x}^{r} xy \ f(x,y)$

		عدد الوحدات المستهلكة y		
		0	**1**	**2**
عدد الأطفال x	**1**	$1\times0\times(1.5/18) = 0$	$1\times1\times(2.5/18)=(2.5/18)$	$1\times2\times(3.5/18)=(7/18)$
	2	$2\times0\times(1/18)=0$	$2\times1\times(2/18)=(4/18)$	$2\times2\times(3/18)=(12/18)$
	3	$3\times0\times(0.5/18)=0$	$3\times1\times(1.5/18)=(4.5/18)$	$3\times2\times(2.5/18)=(15/18)$

$$\sum_{y}^{c}\sum_{x}^{r} xy \ f(x,y) = \left\{\left(0+\frac{2.5}{18}+\frac{7}{18}\right)+\left(0+\frac{4}{18}+\frac{12}{18}\right)+\left(0+\frac{4.5}{18}+\frac{15}{18}\right)\right\} = \frac{45}{18} = 2.5$$

إذا التغاير قيمته هي:

$$\sigma_{xy} = \sum_{y}^{c}\sum_{x}^{r} xy \ f(x,y) - \mu_x \mu_y$$

$$= 2.5 - (1.83)(1.33) = 0.0556$$

σ_x: هو الانحراف المعياري لعدد الأطفال x وهو الجذر التربيعي للتباين، أي أن:

$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{0.6389} = 0.7993$$

σ_y: هو الانحراف المعياري لعدد الوحدات المستهلكة y ويحسب أيضا كالتالي:

$$\sigma_y = \sqrt{\sigma_y^2} = \sqrt{0.5556} = 0.7454$$

ويكون معامل الارتباط هو:

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{0.0556}{(0.7993)(0.7454)} = 0.0933$$

يوجد ارتباط طردي ضعيف بين عدد الأطفال في الأسرة وعدد الوحدات المستهلكة من الحليب الجاف.

4- تكوين التوزيع الاحتمالي الشرطي لعدد الوحدات المستهلكة عند معلومية عدد الأطفال.

يعرف التوزيع الشرطي للمتغير، بأنه التوزيع الاحتمالي لأحد المتغيرين عند معلومية المتغير الآخر، ولحساب التوزيع الشرطي لعدد الوحدات المستهلكة y إذا علم عدد الأطفال في الأسرة x تطبق المعادلة التالية:

$$f(y \mid x) = \frac{f(x,y)}{f(x)} = \frac{((y-0.5x+2)/18)}{((9-1.5x)/18)} = \frac{(y-0.5x+2)}{(9-1.5x)}, \quad y = 0,1,2 \tag{10}$$

5- ولحساب متوسط عدد الوحدات المستهلكة y عند معلومية عدد الأطفال x ويرمز له بالرمز $\mu_{y|x}$

تطبق معادلة المتوسط الشرطي باتباع نفس الطريقة المتبعة في حالة الدوال الهامشية ولكن مع احلال دالة الاحتمال الشرطي $f(y \mid x)$ مكان الدالة الهامشية $f(y)$ كما هو مبين بالمعادلة التالية:

$$\mu_{y|x} = \sum_{y}^{c} y \, f(y \mid x) \tag{11}$$

وبالتطبيق نجد أن:

$$\begin{aligned}\mu_{y|x} &= \sum_{y}^{c} y \, f(y \mid x) \\ &= 0 \times f(0 \mid x) + 1 \times f(1 \mid x) + 2 \times f(2 \mid x) \\ &= 0 + \frac{(1-0.5x+2)}{(9-1.5x)} + 2\frac{(2-0.5x+2)}{(9-1.5x)} = \frac{(11-1.5x)}{(9-1.5x)}\end{aligned}$$

أي أن متوسط عدد الوحدات المستهلكة $\left[\mu_{y|x} = (11-1.5x)/(9-1.5x)\right]$ دالة في عدد الأطفال x، ويمكن معرفة معدل التغير في متوسط عدد الوحدات عند تغير قيم x .

6- كون مصفوفة التباين والتغاير للمتغيرين.

تعرف مصفوفة التباين والتغاير بأنها مصفوفة مربعة عناصر قطرها الرئيسي تمثل التباينات وعناصرها الأخري تمثل التغايرات، وحيث أن لدينا متغيرين فقط تكون هذه من الدرجة (2×2) ويعبر عنها كالتالي:

$$\begin{array}{ccc} & x & y \\ x & \sigma_x^2 & \sigma_{xy} \\ y & \sigma_{xy} & \sigma_y^2 \end{array} \quad \text{i.e.} \quad \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix} = \begin{pmatrix} 0.6389 & 0.0556 \\ 0.0556 & 0.5556 \end{pmatrix}$$

تمرين: 1

الجدول التالي يعطي الدالة الاحتمالية المشتركة للمتغيرين Y , X

Y / X	2	3	4

1	**0.06**	**0.15**	**0.09**
2	**0.14**	**0.35**	**0.21**

هل المتغيران مستقلان؟

تمرين 2: اذا كانت دالة الكثافة المشتركة للمتغيرين X , Y هي

$$f(X,Y) = 8\,XY \quad ; \quad 0 \le X \le Y \le 1$$

أوجد

(أ) دوال الكثافة الهامشية ووضح إذا كان المتغيرين مستقلان أو لا.

(ب) $f(X \mid Y = {}^{1}/_{2})$

(ج) $f(Y \mid X = {}^{1}/_{3})$

التوقع الرياضي وعزوم التوزيعات الثنائية

إذا كانت (X,Y) متغيرا عشوائيا ثنائيا مستمرا وكانت دالة الكثافة المشتركة هي $f(X,Y)$

وكانت $g(X,Y)$ دالة في X,Y فإن القيمة المتوقعة لهذه الدالة تعرف كالآتي:

في حالة البيانات المتصلة $$E\{g(X,Y)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(X,Y)\; f(X,Y)\,dX\,dY$$

في حالة البيانات المنفصلة $$\sum\sum g(X,Y)\; f(X,Y)$$

إذا كانت $g\,(X,Y) = X$ فإن

$$E(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} X\, f(X,Y)\,dX\,dY$$

وبالمثل إذا كانت $g\,(X,Y) = Y$ فإن

$$E(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} Y\, f(X,Y)dXdY = \int Yf(Y)dy = \mu_Y$$

تباين المتغير X

إذا كانت $g(X,Y) = (X-\mu_x)^2$ فإن

$$E\left[(X-\mu_x)^2\right] = \iint (X-\mu_x)^2 f(X,Y)dxdy$$

$$= \int (X-\mu_x)^2 f(X)dX = \sigma^2 x = V(X)$$

تباين المتغير Y

$$E\left[(Y-\mu_y)^2\right] = \iint (Y-\mu_y)^2 f(X,Y)dxdy$$

$$= \int (Y-\mu_y)^2 f(Y)dY = \sigma^2 y = V(Y)$$

التغاير:

إن القيمة المتوقعة للدالة $(Y-\mu_y)$ $(X-\mu_x)$ تعرف بالتغاير بين المتغيرين X,Y

ويرمز له بالرمز Cov(X,Y) أو الرمز C(X,Y) أي أن

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

$$= E(XY) - \mu_x \mu_y$$

و هو يقيس درجة الترافق بين المتغيرين

ملاحظة:

إذا كانت X,Y متغيران مستقلان فإن $Cov(X,Y)=0$ (1)

(2) $Cov\,(X,X) = V\,(X) = \sigma^2_x$

(3) $Cov\,(X,Y) = Cov\,(Y,X)$

إذا كانت α,β عددين حقيقيين فإن

$$Cov(\alpha X,\beta Y) = \alpha\beta\, Cov\,(X,Y)$$

معامل الارتباط:

معامل الارتباط بين المتغيرين X,Y يرمز له بالرمز ρ (X,Y) ويعرف

$$\rho = \rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

ان معامل الارتباط يقيس درجة العلاقة بين المتغيرين ومدى ارتباطهما ببعضهما وعلى ذلك فهو يؤدي نفس دور التغاير إلا أن له ميزة مهمة وهو أن وحداته مطلقة.

تمرين3:

إذا كانت الدالة الكثافة المشتركة للمتغيرين X,Y هي

$$f(x,y) = 8\,x\,y \qquad ; \quad 0 \le x \le y \le 1$$

أحسب

(أ) التغاير

(ب) معامل الارتباط

المتوسط المشروط أو التوقع المشروط

نعلم أن دالة الكثافة المشروطة للمتغير Y بشرط أن $X=x$ هي

$$f(\,Y\mid x\,) = \frac{f(x,Y)}{f(x)}$$

ويعرف التوقع المشروط أو المتوسط للمتغير Y (بشرط أن $X=x$) كالاتي

$$E(Y \mid x) = \int_{-\infty}^{\infty} Yf(Y \mid x)\, dY = \int_{-\infty}^{\infty} \frac{Yf(x,Y)dY}{f(x)}$$

التوزيع الطبيعي الثنائي

يقال أن المتغير الثنائي المستمر (X,Y) يتبع توزيعا طبيعيا ثنائيا بمعالم $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$ إذا كانت دالة كثافته المشتركة هي:

$$f(X,Y) = \frac{e^{-q/2}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}; -\infty \leq X, Y \leq \infty$$

حيث أن:

(أ) $$q = \frac{1}{1-\rho^2}\left[\frac{(X-\mu_1)^2}{\sigma_1} - 2\rho\left(\frac{X-\mu_1}{\sigma_1}\right)\left(\frac{Y-\mu_2}{\sigma_2}\right) + \left(\frac{Y-\mu_2}{\sigma_2}\right)^2\right]$$

(ب) $\sigma_1 > 0$, $\sigma_2 > 0$, $-1 \leq \rho \leq 1$

تمرين4:

إذا كان أطوال الأزواج X وأطوال الزوجات Y في مجتمع ما يتبع توزيعا طبيعيا ثنائيا بمعالم

$$\mu_x = 170, \quad \mu_Y = 168\,, \; \sigma_X = \sigma_Y = 8, \quad \rho = 0.6$$

.

اختير زوجا عشوائيا. فإذا علم أن طول الزوج 174 cm فما احتمال أن يكون طول زوجته ما بين (166,172) .