١٣

التوزيعات الاحتمالية المشتركة

Joint distributions

مقدمـــة

في كثير من النواحي التطبيقية، يكون اهتمام الباحث هو دراسة شكل التوزيع الاحتمالي بين متغيرين ، لمعرفة العلاقة بينهما، أو التنبؤ بأحدهما من الآخر.

أولا: التوزيع الاحتمالي المشترك لمتغيرين متقطعين

Joint distribution for two discrete Variables

بفرض أن (x_2, x_1) متغيرين عشوائيين متقطعين، حيث أن:

$$(x_1, x_2 : (x_{11}, x_{21}), \ldots, (x_{1n}, x_{2n}))$$

فإن دالة الاحتمال المشترك تكتب على الصورة $f(x_1, x_2) = P(x_1 = x_1, x_2 = x_2)$

ومن خصائص هذه الدالة ما يلي:

١- $0 < f(x_1, x_2) < 1$

٢- $\sum_{x_1} \sum_{x_2} f(x_1, x_2) = 1$

ويمكن من خلال دالة التوزيع الاحتمالي تكوين جدول التوزيع الاحتمالي المشترك للمتغيرين (x_2, x_1)

تطبيق

إذا كانـت دالـة الاحتمـال المـشتركة بـين عـدد أفـراد الأسـرة (x_1)، وعـدد الوحـدات المستهلكة من سلعة معينة خلال الأسبوع (x_2) تأخذ الصورة التالية.

$$f(x_1, x_2) = \frac{x_1 + x_2}{27} \quad , \quad x_1 = 1,2,3 \quad , \quad x_2 = 0,1,2$$

كون جدول التوزيع الاحتمالي، ثم أوجد الآتي:

١- احتمال أن أسرة ما يكون عدد أفرادها 2، وعدد وحدات استهلاكها الشهري وحدة واحدة أسبوعيا.

٢- إذا كان لدينا 500 أسرة، ما هو عدد الأسر المتوقع أن يكون عدد أفرادها واحد، واستهلاكها 2 وحدة أسبوعيا؟

الحــل

تكوين جدول التوزيع الاحتمالي المشترك.

يتم التعويض بجميع قيم (x_2, x_1) في الدالة أعلاه كما يلي:

$$f(1,0) = (1+0)/27 = 1/27 \quad , \quad f(1,1) = (1+1)/27 = 2/27 \quad , \quad f(1,2) = (1+2)/27 = 3/27$$

$$f(2,0) = (2+0)/27 = 2/27 \quad , \quad f(2,1) = (2+1)/27 = 3/27 \quad , \quad f(2,2) = (2+2)/27 = 4/27$$

$$f(3,0) = (3+0)/27 = 3/27 \quad , \quad f(3,1) = (3+1)/27 = 4/27 \quad , \quad f(3,2) = (3+2)/27 = 5/27$$

جدول التوزيع الاحتمالي المشترك

		عدد الوحدات المستهلكة (x_2)			$f(x_1)$
		0	**1**	**2**	
عدد أفراد الأسرة (x_1)	**1**	1/27	2/27	3/27	6/27
	2	2/27	3/27	4/27	9/27
	3	3/27	4/27	5/27	12/27
	$f(x_2)$	6/27	9/27	12/27	**1**

١- احتمال أن أسرة عدد أفرادها 2، واستهلاكها الشهري وحدة واحدة أسبوعيا هو.

$$P(x_1 = 2, x_2 = 1) = f(2,1) = 3/27$$

٢- عدد الأسر المتوقع أن يكون عدد أفرادها واحد واستهلاكها 2 هو:

$$= 500 \times f(1,2)$$
$$= 500 \times (3/27) = 55.6 = 56$$

دوال كثافة الاحتمال الهامشية

هي الدالة الخاصة بمتغير عشوائي واحد، ويمكن اشتقاقها من التوزيع المشترك كما يلي:

إذا كان $f(x_1, x_2)$ هي دالة كثافة الاحتمال المشترك بين المتغيرين (x_1, x_2) فإن دالة كثافة الاحتمال الخاصة بكل متغير على حدة تشتق كما يلي:

$$f(x_2) = \sum_{x_1} f(x_1, x_2) \quad ، \quad f(x_1) = \sum_{x_2} f(x_1, x_2)$$

وتسمى الدالة $f(x_1)$ ، الدالة $f(x_2)$ بالدوال الهامشية.

تطبيق

في التطبيق السابق أوجد الآتي:

١- دالة كثافة الاحتمال الهامشية لعدد أفراد الأسرة في المجتمع.

٢- احسب القيمة المتوقعة، والتباين لعدد أفراد الأسرة.

دالة كثافة الاحتمال الشرطية.

في بعض الحالات قد يكون لدينا معلومات عن أحد المتغيرين، ويراد معرفة احتمالات حدوث قيم المتغير الآخر في ظل هذه المعرفة، في هذه الحالة يتم اشتقاق دالة تسمى بدالة الاحتمال الشرطي، كما يلي:

إذا كان $f(x_1, x_2)$ هي دالة كثافة الاحتمال المشترك بين المتغيرين (x_1, x_2) فإن دالة كثافة الاحتمال الشرطي للمتغير (x_1) إذا توافرت لدينا معلومات عن المتغير (x_2) هي:

$$f(x_1|x_2) = \frac{f(x_1, x_2)}{f(x_2)}$$

تطبيق

في التطبيق السابق أوجد الآتي:

١- دالة كثافة الاحتمال الشرطي لعدد الوحدات المستهلكة، إذا كان عدد أفراد الأسرة في المجتمع فردان.

٢- احسب التوقع الشرطي لعدد الوحدات المستهلكة

معامل الارتباط الخطي البسيط

يمكن حساب معامل الارتباط بين المتغيرين (x_1, x_2) ، حيث يحسب هذا المعامل بتطبيق المعادلة التالية:

$$\rho = \frac{\sigma_{12}}{\sigma_1 \sigma_2}، \qquad -1 < \rho < 1$$

حيث أن:

σ_{12} : هو التغاير بين المتغيرين (x_1, x_2) ، ويحسب من التوزيع الاحتمالي المشترك $f(x_1, x_2)$ كما يلي:

$$\begin{aligned} \sigma_{12} &= E[(x_1 - \mu_1)(x_2 - \mu_2)] \\ &= E(x_1 x_2) - \mu_1 \mu_2 \end{aligned}$$

حيث أن $E(x_1 x_2)$ هو التوقع المشترك بين المتغيرين، ويحسب من التوزيع المشترك $f(x_1, x_2)$ كالتالي:

$$E(x_1 x_2) = \sum_{x_1} \sum_{x_2} x_1 x_2 f(x_1, x_2)$$

أما μ_1, μ_2 فهما متوسطي المتغيرين العشوائيين، ويحسبا من التوزيع الاحتمالي الهامشي:

$$\mu_2 = \sum_{x_2} x_2 f(x_2) \quad ، \mu_1 = \sum_{x_1} x_1 f(x_1)$$

σ_1 : هو الانحراف المعياري للمتغير العشوائي x_1 ، σ_2 : هو الانحراف المعياري للمتغير العشوائي x_2

تطبيق

في التطبيق السابق احسب معامل الارتباط بين المتغيرين.

جدول توزيع احتمالي مشترك بين المتغيرين x_2, x_1

		x_1			$f(x_2)$	$x_2\ f(x_2)$	$x_2^2\ f(x_2)$
		1	**2**	**3**			
x_2	**1**	**0.1**	**0.05**	**0.05**	**0.20**	**0.20**	**0.20**
	2	**0.05**	**0.1**	**0.15**	**0.30**	**0.60**	**1.20**
	3	**0.15**	**0.15**	**0.2**	**0.50**	**1.50**	**4.50**
$f(x_1)$		**0.30**	**0.30**	**0.40**	**1.0**	**2.3**	**5.90**
$x_1\ f(x_1)$		**0.30**	**0.60**	**1.20**	**2.1**		
$x_1^2\ f(x_1)$		**0.30**	**1.20**	**3.60**	**5.1**		

$$
\begin{aligned}
E(x_1x_2) &= 1\times 1\times 0.10 + 1\times 2\times 0.05 + 1\times 3\times 0.05 \\
&\quad + 2\times 1\times 0.05 + 2\times 2\times 0.10 + 2\times 3\times 0.15 \\
&\quad + 3\times 1\times 0.15 + 3\times 2\times 0.15 + 3\times 3\times 0.20 \\
&= 0.1 + 0.10 + 0.15 + 0.10 + 0.40 + 0.90 + 0.45 + 0.90 + 1.80 \\
&= 4.9
\end{aligned}
$$

$$\sigma_{x_1}^2 = E(x_1^2) - \mu_{x_1}^2 = 5.1 - (2.1)^2 = 0.69$$

$$\sigma_{x_2}^2 = E(x_2^2) - \mu_{x_2}^2 = 5.9 - (2.3)^2 = 0.61$$

$$\sigma_{x_1 x_2} = E(x_1 x_2) - \mu_{x_1}\mu_{x_2} = 4.9 - (2.1 \times 2.3) = 0.07$$

$$\rho_{x_1 x_2} = \frac{\sigma_{x_1 x_2}}{\sqrt{\sigma_{x_1}^2}\sqrt{\sigma_{x_2}^2}} = \frac{0.07}{\sqrt{0.69}\sqrt{0.61}} = 0.11$$

يوجد ارتباط طردي بين المتغيرين

ثانيا: التوزيع الطبيعي الثنائي:

إذا كان المتغيرين x ، y لهما توزيع طبيعي مشترك، وكان تباين المتغير x هو σ_x^2 ، وتباين والمتغير y هو σ_y^2 ، والتغاير بينهما هو σ_{xy} ، فإن معامل الارتباط في المجتمع ρ_{xy} هو:

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} ، \qquad -1 < \rho_{xy} < 1$$

الاستدلال الإحصائي حول الارتباط الخطي البسيط

في المجتمع (ρ_{xy})

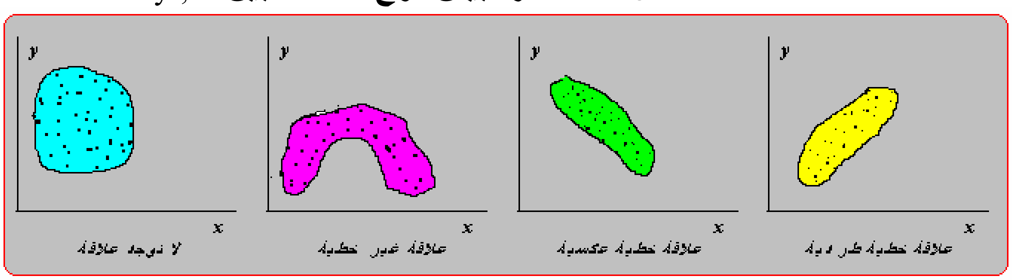
تهتم هذه المحاضرة بالاستدلال الإحصائي لمعامل الارتباط في المجتمع فإذا كان اهتمام الباحث هو دراسة العلاقة بين متغيرين استخدم لذلك أسلوب تحليل الارتباط، ومن الأمثلة على ذلك:

١- الإنفاق، والدخل العائلي.
٢- سعر السلعة، والكمية المطلوبة منها.
٣- الفترة الزمنية لتخزين الخبز، وعمق طراوة الخبز.
٤- تقديرات الطلاب في مقرر الإحصاء، وتقديراتهم في مقرر الرياضيات.
٥- كميات السماد المستخدمة، وكمية الإنتاج من محصول معين تم تسميده بهذا النوع من السماد.
٦- عدد مرات ممارسة نوع معين من الرياضة البدنية، ومستوى الكلسترول في الدم.
٧- وزن الجسم، وضغط الدم.

والأمثلة على ذلك في المجال التطبيقي كثيرة، فإذا كان لدينا المتغيرين $(x \; , \; y)$ ، **وتم جمع بيانات عن أزواج قيم هذين المتغيرين، وتم تمثيلها بيانيا فيما يسمى بشكل الانتشار، فإن العلاقة بينها تأخذ أشكالا مختلفة على النحو التالي :**

شكل(٦-١)

شكل الانتشار لبيان نوع العلاقة بين x , y



التقدير بنقطة لمعامل الارتباط الخطى البسيط Simple Correlation

إذا كان الغرض من التحليل هو تحديد نوع وقوة العلاقة بين متغيرين ، يستخدم تحليل الارتباط ، وأما إذا كان الغرض هو دراسة وتحليل أثر أحد المتغيرين على الآخر ، يستخدم تحليل الانحدار، وفي هذا الفصل يتم عرض أسلوب تحليل الارتباط الخطي البسيط، أي في حالة افتراض أن العلاقة بين المتغيرين تأخذ الشكل الخطي ، وسوف يجرى حسابه في حالة البيانات الكمية.

٦/٢/١ الغرض من تحليل الارتباط الخطى البسيط

الغرض من تحليل الارتباط الخطي البسيط هو تحديد نوع وقوة العلاقة بين متغيرين، ويرمز له في حالة المجتمع بالرمز ρ (رو)، وفي حالة العينة بالرمز r، وحيث أننا في كثير من النواحي التطبيقية نتعامل مع بيانات عينة مسحوبة من المجتمع، سوف نهتم بحساب معامل الارتباط في العينة r كتقدير لمعامل الارتباط في المجتمع، ومن التحديد السابق للغرض من معامل الارتباط، نجد أنه يركز على نقطتين هما:

- **نوع العلاقة:-** وتأخذ ثلاث أنواع حسب إشارة معامل الارتباط كما يلي:

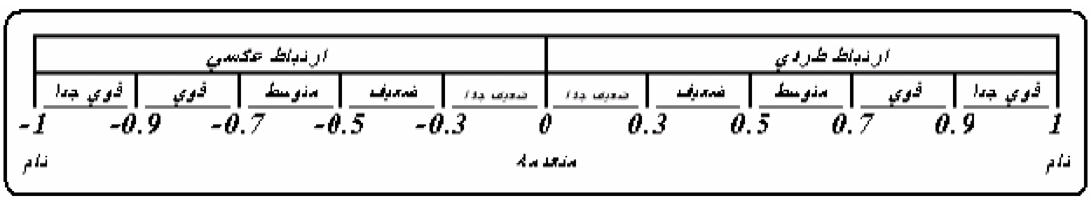
١- إذا كانت إشارة معامل الارتباط سالبة $(r < 0)$ توجد علاقة عكسية بين المتغيرين، بمعنى أن زيادة أحد المتغيرين يصاحبه انخفاض في المتغير الثاني، والعكس.

٢- إذا كانت إشارة معامل الارتباط موجبة $(r > 0)$توجد علاقة طردية بين المتغيرين، بمعنى أن زيادة أحد المتغيرين يصاحبه زيادة في المتغير الثاني، والعكس .

٣- إذا كان معامل الارتباط قيمته صفرا $(r = 0)$ دل ذلك على انعدام العلاقة بين المتغيرين.

- **قوة العلاقة:-** ويمكن الحكم على قوة العلاقة من حيث درجة قربها أو بعدها عن (± 1)، حيث أن قيمة معامل الارتباط تقع في المدى$(-1 < r < 1)$، وقد صنف بعض الإحصائيين درجات لقوة العلاقة يمكن تمثيلها على الشكل التالي:

شكل (٦-٢)
درجات قوة معامل الارتباط



في حالة جمع بيانات عن متغيرين كميين $(y \ , \ x)$، يمكن قياس الارتباط بينهما، باستخدام طريقة "بيرسون" Pearson، ومن الأمثلة على ذلك: قياس العلاقة بين الوزن والطول، والعلاقة بين الإنتاج والتكلفة، والعلاقة بين الإنفاق الاستهلاكي والدخل، والعلاقة

بن الدرجة التي حصل عليها الطالب وعدد ساعات الاستذكار، وهكذا الأمثلة على ذلك كثيرة.

ولحساب معامل الارتباط في العينة، كتقدير لمعامل الارتباط في المجتمع تستخدم صيغة " بيرسون" التالية :

$$r = \frac{S_{xy}}{S_x\, S_y} = \frac{\dfrac{\sum (x-\bar{x})(y-\bar{y})}{(n-1)}}{\sqrt{\dfrac{\sum (x-\bar{x})^2}{(n-1)}}\sqrt{\dfrac{\sum (y-\bar{y})^2}{(n-1)}}} \qquad (٦-١)$$

حيث أن :

$S_{xy} = \sum (x-\bar{x})(y-\bar{y}) / (n-1)$ **: هو التغاير بين** $(x \,,\, y)$ ،

$S_x = \sqrt{\sum (x-\bar{x})^2 / (n-1)}$ **: هو الانحراف المعياري لقيم** (x) ،

$S_y = \sqrt{\sum (y-\bar{y})^2 / (n-1)}$ **: هو الانحراف المعياري لقيم** (y).

ويمكن اختصار الصيغة السابقة على النحو التالي:

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2}\sqrt{\sum (y-\bar{y})^2}} \qquad (٦-٢)$$

مثـــال (٦-١)

فيما يلي المساحة المنزرعة بالأعلاف الخضراء بالألف هكتار، وإجمالي إنتاج اللحوم بالألف طن، خلال الفترة من 1995**حتى عام** 2002.

السنة	1995	1996	1997	1998	1999	2000	2001	2002
المساحة	305	313	297	289	233	214	240	217

الكمية	592	603	662	607	635	699	719	747

والمطلوب: حساب معامل الارتباط بين المساحة والكمية، وما هو مدلوله ؟

الحــل

بفرض أن (x) هي المساحة المنزرعة، (y) هي الكمية، ولحساب معامل الارتباط بين (x , y) يتم تطبيق المعادلة (٦-٢)، وذلك على النحو التالي:

- **حساب الوسط الحسابي لكل من المساحة، والكمية $(\bar{x} , \bar{y})$.**

$$\bar{x} = \frac{\sum x}{n} = \frac{2108}{8} = 263.5 \ , \ \bar{y} = \frac{\sum y}{n} = \frac{5264}{8} = 658$$

- **حساب المجاميع**

x	y	$x-\bar{x}$	$(x-\bar{x})^2$	$y-\bar{y}$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
305	592	41.5	1722.25	-66	4356	-2739
313	603	49.5	2450.25	-55	3025	-2722.5
297	662	33.5	1122.25	4	16	134
289	607	25.5	650.25	-51	2601	-1300.5
233	635	-30.5	930.25	-23	529	701.5
214	699	-49.5	2450.25	41	1681	-2029.5
240	719	-23.5	552.25	61	3721	-1433.5
217	747	-46.5	2162.25	89	7921	-4138.5
2108	5264	0	12040	0	23850	-13528

$$\sum(x-\bar{x})^2 = 12040 \ , \ \sum(y-\bar{y})^2 = 23850 \ ,$$
$$\sum(x-\bar{x})(y-\bar{y}) = -13528$$

إذا معامل الارتباط قيمته هي:

$$r=\frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2}\sqrt{\sum(y-\bar{y})^2}}=\frac{-13528}{\sqrt{12040}\sqrt{23850}}$$

$$=\frac{-13528}{(109.727)(154.434)}=\frac{-13528}{16945.619}=-0.798$$

- **يوجد ارتباط عكسي قوي بين المساحة المنزرعة، وكمية إنتاج اللحوم.**

تبسيط العمليات الحسابية:

في بعض الأحيان، يكون استخدام صيغة المعادلة (٦-٢) في غاية الصعوبة، خاصة إذا لازم العمليات الحسابية قيما كسرية، من أجل ذلك يمكن تبسيط الصيغة (٦-٢) إلى صيغة أسهل تعتمد على مجموع القيم وليس على انحرافات القيم عن وسطها الحسابي، وهذه الصيغة هي:

$$\boxed{r=\frac{\sum xy-\frac{\sum x\sum y}{n}}{\sqrt{\left(\sum x^2-\frac{(\sum x)^2}{n}\right)\left(\sum y^2-\frac{(\sum y)^2}{n}\right)}}} \qquad (٦-٣)$$

وبالتطبيق على بيانات المثال السابق ، يتبع الآتي :

- **حساب المجاميع:**

x	y	xy	x^2	y^2
305	592	180560	93025	350464
313	603	188739	97969	363609
297	662	196614	88209	438244
289	607	175423	83521	368449
233	635	147955	54289	403225
214	699	149586	45796	488601
240	719	172560	57600	516961
217	747	162099	47089	558009

المجاميع المطلوبة
$\sum x=2108\ ,\ \sum y=5264$
$\sum xy=1373536$
$\sum x^2=567498$
$\sum y^2=3487562$

	3487562	567498	1373536	5264	2108

- **حساب معامل الارتباط:**

باستخدام المجاميع السابقة، وبالتطبيق على المعادلة (٦-٣) أعلاه، نجد أن معامل الارتباط قيمته هي:

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)\left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}}$$

$$= \frac{1373536 - \frac{(2108)(5264)}{8}}{\sqrt{\left(567498 - \frac{(2108)^2}{8}\right)\left(3487562 - \frac{(5264)^2}{8}\right)}}$$

$$= \frac{-13528}{\sqrt{(12040)(23850)}} = \frac{-13528}{16945.619} = -0.798$$

وهي نفس النتيجة السابقة: