الجزء الخامس

الإحصاءات الوصفية المتعددة

- الفصل الخامس عشر: التوسع في جداول التقاطع: إضافة متغيرات التحكم
- الفصل السادس عشر: الارتباط الجزئي والانحدار المتعدد

الفصل الخامس عشر

التوسع في جداول التقاطع: إضافة متغيرات التحكم

مقدمة:

تناولنا في الفصل السادس تحليل العلاقة بين متغيرين، وقد تم افتراض أن أي علاقة تمتْ مشاهدتها في البيانات بين متغيرين هي ناشئة عن علاقة بسيطة ومباشرة. إن التطابق القوي في الجداول الثنائية، على أية حال، لا يعني بالضرورة أن هذه العلاقة البسيطة المباشرة موجودة واقعياً فالأمر يتعلق فقط، بالكيفية التي قمنا بها في تفسير هذه البيانات. ويمكننا القول، إن هناك علاقات أكثر تعقيداً تخفيها البيانات، لكننا لم نبذل جهداً كافياً لاكتشاف هذه العلاقات.

إن أسهل طريقة للتوسع في العلاقة، هي اكتشافها في جدول التوافق وذلك من خلال النظر إلى احتمالية تأثير متغير ثالث على العلاقة الثنائية الأصلية. واعتماداً على مخرجات هذا التوسع، فإنه يمكننا أن نعدل من نموذج العلاقة بين المتغيرين الأصليين آخذين في الاعتبار تأثير المتغير الثالث.

هناك ثلاث نتائج يمكن الوصول إليها عند إدخال المتغير الثالث في عملية التحليل:

1- النتيجة الأولى أنه ليس للمتغير الثالث أي تأثير على العلاقة الأصلية المباشرة.

2- قد يكون لإدخال المتغير الثالث تأثير في العلاقة الموجودة.

3- أو أن هناك علاقة مشروطة[1].

في هذا الفصل سوف نقوم باستقصاء هذه النتائج المحتملة، وذلك من خلال إعطاء أمثلة لكل نمط من أنماط هذه العلاقة.

1- العلاقة المباشرة:

دعنا ننظر إلى النقطة الأولى التي مفادها أنه لا يوجد أي تأثير لمتغير التحكم في العلاقة الثنائية الأصلية. ويبرهن لنا، هذا، على أن النموذج المباشر البسيط هو النموذج الملائم لوصف العلاقة.

على سبيل المثال، يمكننا إقامة علاقة بين الدخل ومشاهدة الإذاعة المرئية كما هي مرتبة في جدول التقاطع.. (انظر جدول 15 – 1).

جدول (15 ـ 1): مشاهدة الإذاعة المرئية حسب مستوى الدخل

مشاهدة الإذاعة المرئية	الدخل		المجموع
	منخفض	عال	
منخفض	115 % 57	95 % 32	210
عال	88 % 43	204 % 68	292
المجموع	203	299	502
جاما = 0.47 = Gamma			

إن العلاقة الأصلية في هذا النموذج بين هذين المتغيرين هي علاقة مباشرة كما في الشكل التالي:

مستوى الدخل ← مشاهدة الإذاعة المرئية

شكل (15 ـ 1) علاقة مباشرة

إن النموذج النظري الذي اعتمدنا عليه يوضح أن الدخل يؤثر مباشرة في كمية مشاهدة الإذاعة المرئية للشخص، وأن الإحصاءات تؤكد لنا أن هذه العلاقة الموجبة متوسطة إلى قوية، وعندما نجادل أن هناك علاقة مباشرة بين هذين المتغيرين فإننا بشكل فعّال نجادل في أن العلاقة سوف تكون نفسها بغض النظر عن أي متغير آخر يمكن أن يسبب تباين الحالات فيما بينها، في هذا المثال نعتقد أن الدخل يؤثر في مشاهدة الإذاعة المرئية بنفس الطريقة وبنفس الدرجة بغض النظر عن أي متغير آخر يمكنه أن يسبب التباين في الحالات مثل النوع، العمر... الخ.

إن هذه العلاقة الثنائية المباشرة على - أي حال - يمكن أن تظهر لنا أنها علاقة بسيطة. بالتأكيد توجد متغيرات أخرى قد تؤثر على كمية مشاهدة الإذاعة المرئية. فقد يشعر باحث آخر على سبيل المثال، بأن (المستوى التعليمي) يؤثر أيضاً في مشاهدة الأفراد للإذاعة المرئية. ولتقييم احتمالية تأثير هذا المتغير الجديد (المستوى التعليمي) أن له تأثير على العلاقة المشاهدة بين الدخل وكمية مشاهدة الإذاعة المرئية، يمكننا تقسيم العينة إلى مجموعتين فرعيتين:

المجموعة الأولى: تمثل أولئك الذين لم يكملوا التعليم الثانوي، والمجموعة الثانية: تمثل أولئك الذين أكملوا بعضاً من التعليم العالي. ومن الناحية الفنية فإننا قد أدخلنا متغير المستوى التعليمي كمتغير للتحكم. إن متغير التحكم هو ذلك المتغير الذي يحلل البيانات إلى مجموعات فرعية تستند على فئات متغير التحكم، إن تأثير متغير التحكم يؤدي بنا إلى توليد جداول متقاطعة منفصلة لكل المجموعات الفرعية والتي تم تحديدها من خلال متغير التحكم، في هذا المثال: دعنا نأخذ أولاً تلك (الحالات التي ليس لها تعليم عالٍ) ونصمم جدول تقاطع بين دخولهم ومشاهدتهم للإذاعة المرئية متغاضين عن تلك الحالات التي أكملت بعضاً من التعليم العالي، وبعد ذلك نأخذ فقط الحالات التي لديها بعضاً من

التعليم العالي ونصمم جدول تقاطع بين دخولهم ومشاهدة الإذاعة المرئية متغاضين عن أولئك الذين ليس لديهم تعليم عال.

إن ما قمنا به من توليد لهذه الجداول يطلق عليه الجداول الجزئية[2]. ويمكننا أن نولد من الجداول الجزئية العدد الذي نريده كفئات لمتغير التحكم (انظر جدول 15-2، 15-3) فإننا في هذين الجدولين قد أدخلنا المستوى التعليمي كمتغير للتحكم وتم قياس التعليم بفئتين فقط. وعليه يمكننا أن نولد جدولين جزئيين. (إذا كان لدينا ثلاث فئات لمتغير التحكم مثل: لا تعليم جامعي، بعض التعليم الجامعي، كثير من التعليم الجامعي في هذه الحالة يمكن توليد ثلاثة جداول جزئية).

جدول (15 ـ 2)

مشاهدة الإذاعة المرئية حسب مستوى الدخل

التحكم في مستوى التعليم (الذين ليس لديهم تعليم عالٍ)

مشاهدة الإذاعة المرئية	الدخل		المجموع
	منخفض	عال	
منخفض	78 A % 57	22 B % 31	100
عال	58 C % 43	48 D % 69	106
المجموع	136	70	206
جاما = 0.49 = Gamma			

جدول (15 ـ 3)

مشاهدة الإذاعة المرئية

التحكم في مستوى التعليم (الذين أكملوا بعضا من التعليم)

مشاهدة الإذاعة المرئية	الدخل		المجموع
	منخفض	عال	
منخفض	37 % 55	73 % 32	110
عـــال	30 % 45	156 % 68	186
المجموع	67	229	296
جاما = 0.45 = Gamma			

وبهذه النتيجة يمكننا القول إن العلاقة الأصلية قد ولَّدت لنا تقريباً علاقة متساوية في كل جدول جزئي، وإن قيمة جاما لكل واحد من هذين الجدولين الجزئيين تقرب كثيراً من قيمة جاما في الجدول الأصلي قبل التحكم في مستوى التعليم. بمعنى آخر، بغض النظر عن المستوى التعليمي فإن العلاقة بين الدخل ومشاهدة الإذاعة المرئية ما زالت قائمة، إن العلاقة المباشرة التي شاهدناها في البداية لازالت باقية على ما هي عليه حتى بعد التحكم في المتغير الثالث. وبغض النظر عن كيف تتباين الحالات طبقاً لمستوى الدخل، فالعلاقة الثنائية المباشرة تبقى الشيء نفسه، وعليه فلن نغير من نموذجنا الأصلي الذي يوضح أن هناك علاقة مباشرة بين الدخل ومشاهدة الإذاعة المرئية[3].

إجراءات توليد جداول التقاطع مع التحكم في المتغيرات باستخدام SPSS:

1- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Analyze/ descriptive Statistics / Crosstabs (يعطيك مربع جداول التقاطع).

2- انقر على TV Watching (مشاهدة الإذاعة المرئية).

3- انقر على ◄ التي تشير إلى القائمة المحددة المعنونة: Row (s) (الصف أو الصفوف).

نقوم بلصق TV Watching (Pasts) في القائمة المحددة للصف أو الصفوف Row (s)

4- انقر على Income.

5- انقر على ◄ التي تشير إلى القائمة المحددة المعنونة Column(s)

نقوم بلصق Income (Pasts) في القائمة المحددة للأعمدة Column(s).

6- انقر على مستوى التعليم Education Level.

7- انقر على ◄ التي تشير إلى القائمة المحددة أسفل Layer 1 of 1، نقوم بلصق (Pasts) Education Level في القائمة المحددة التي تحتوي المتغير المتحكم The Control Variable جدول التقاطع سوف يولد كل قيمة للمتغير في القائمة.

8- انقر على زر Statistics وقم باختيار Gamma هذه العملية ستولد قيمة جاما Gamma لكل جدول جزئي.

9- انقر على الخلايا Cells واختر نسب العمود Column Percentages.

10- انقر على Ok.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

Cross tabs TV Watching "Income" Education Level Crosstabulation

Education Level		Income		ToTAL
		Low	High	
No Post. Secondary	TV Watching Low Count	78	22	100
	% Within Income	57.4 %	34.4 %	48.5 %
	High Count	58	48	106
	% Within Income	42.6 %	88.6 %	51.5 %
	TOTAL Count	136	70	206
	% Within Income	100 %	100 %	100 %
Post. Secondary	TV Watching Low Count	37	73	110
	% Within Income	55.2 %	31.9 %	37.2 %
	High Count	30	156	186
	% Within Income	44.8 %	68.1 %	62.8 %
	TOTAL Count	67	229	296
	% Within Income	100 %	100 %	100 %

Symmetric Measures

Education Level		Value	Asymp. std. Error[a]	Approx.T[b]	Approx. Sig.
No Post. Secondary	ordinal by Ordinal GAMMA	.492	.118	3.657	.000
	N of Valid cases	206			
Post. Secondary	ordinal by Ordinal GAMMA	.450	.113	3.317	.001
	N of Valid cases	296			

a. Not assuming The null hypothesis.

b. Using The asymptotic standard Error assuming the null hypothesis.

المصدر: George Argyrous, Statistics for Social & Health Research, op.cit. , p. 465.

شكل (2-15) مخرجات SPSS لجداول التقاطع بإدخال متغير التحكم

تفسير مخرجات SPSS لجداول التقاطع مع التحكم في المتغيرات.

إذا نظرنا إلى هذه المخرجات نجد في الحقيقة جدولين متقاطعين جمعا في جدول واحد فالنصف الأول من الجدول، هو جدول التقاطع للدخل مع مشاهدة الإذاعة المرئية للحالات التي لم تتحصل على التعليم العالي. وتحت هذا الجدول مباشرة جدول التوافق لتلك الحالات التي أكملت بعضاً من التعليم العالي. إن نسبة الحالات التي تشاهد مستوىً معيناً من الإذاعة المرئية هي الشيء نفسه لكل فئات الدخل بغض النظر عن المستوى التعليمي. وقد عززت هذه النتيجة بقيم جاما Gamma كما ظهرت في الجدول المُعَنْوَن بـِ Symmetric Measures Table. وهذه القيم هي الشيء نفسه كالتي تم حسابها سابقاً. إن العلاقة بين الدخل ومشاهدة الإذاعة المرئية احتفظت بنفس القوة والاتجاه لكل جدول من الجداول الجزئية[4].

2- العلاقة الكاذبة أو الدخيلة:

نفترض أنه عندما أدخلنا المستوى التعليمي في التحليل تحصلنا بدلاً من الجدول (15-2 و15 -3) على جدول (15-4 و 15-5) من خلال هذين الجدولين (4 و 5) نلاحظ أن العلاقة بين الدخل ومشاهدة الإذاعة المرئية والتي شاهدناها في الجدول الأصلي قد اختفت فجأة. إنه بوضوح يمكننا أن نرى أنه لا يوجد ارتباط يمكن الحديث عنه بين الدخل ومشاهدة الإذاعة المرئية، وما إن أدخلنا مستوى التعليم كمتغير للتحكم حتى إن الارتباط الأصلي بين المتغيرين الدخل ومشاهدة الإذاعة المرئية قد اختفى بإدخال هذا المتغير وأن الطريقة الدقيقة للوصول لهذه النتيجة هي حساب جاما الجزئية من قيم جاما لكل جدول جزئي.

جدول (15 ـ 4) مشاهدة الإذاعة المرئية وفقا لمستوى الدخل: التحكم في مستوى التعليم (لأولئك الذين لم يتحصلوا على التعليم العالي)

مشاهدة الإذاعة المرئية	الدخل		المجموع
	منخفض	عال	
منخفض	102 % 75	50 % 71	152
عال	34 % 25	20 % 29	54
المجموع	136	70	206
جاما = 0.09 = Gamma			

جدول (15 ـ 5) مشاهدة الإذاعة المرئية وفقا لمستوى الدخل: التحكم في مستوى التعليم (الذين أكملوا بعضا من التعليم العالي)

مشاهدة الإذاعة المرئية	الدخل		المجموع
	منخفض	عال	
منخفض	13 % 19	45 % 20	58
عال	54 % 81	184 % 80	238
المجموع	67	229	296
جاما = 0.007- = Gamma			

جاما الجزئية Partial Gamma:

إن قيم جاما التي تم حسابها لكل جدول من الجدولين الجزئيين كانت ذات فائدة لكشف النقاب عن العلاقة الكاذبة أو الدخيلة. ويمكننا أن نرى أن هذه القيم المتعلقة

بجاما ضعيفة من حيث القوى باختلاف قيم جاما المشتركة بالجدول الثنائي الأصلي. في الجدول الأصلي، حيث إن الحالات لم تكن منفصلة وفقاً لمستوى التعليم، فقد وصلت قيمة جاما إلى 0.47. إلا أن قيم جاما لكل جدول من الجداول الجزئية تقرب من الصفر.

إن المنطق وراء حساب جاما الذي تَمَّتْ مناقشته في متن هذا الكتاب المتعلق بالأزواج المتوافقة والأعداد المتعلقة بالأزواج غير المتوافقة. فالأزواج المنسجمة أو المتوافقة كما بَيَّنَّا سابقاً هي مجموعة الحالات الزوجية التي تم ترتيبها بشكل متساوٍ مع كل متغير من هذين المتغيرين، وبالتالي شملت العلاقة الإيجابية بين هذين المتغيرين، بينما على الجانب الآخر، فإن الأزواج المتنافرة من الحالات والتي رتبت بشكل مختلف عن المتغيرين قد اشتملت على علاقة سالبة بين المتغيرين.

إذا ما أضفنا الأزواج المتقاربة أو المنسجمة عبر الجدولين الجزئيين وأضفنا الأزواج المتنافرة عبر نفس الجدولين فإنه باستطاعتنا حساب جاما الجزئية التي تقيس العلاقة المباشرة بين المتغيرين اللذين بدأنا بهما مع إدخال متغير ثالث كمتحكم، قد تم حساب جاما لجمع الأزواج المنسجمة والأزواج غير المنسجمة عبر الجداول الجزئية.

تجدر الإشارة إلى أننا لازلنا نستخدم كل الحالات لتحديد جاما الجزئية، ولكننا الآن نقوم بهذا العمل بعد أن تم فصل الحالات إلى جدولين جزئيين منفصلين. إن عملية حساب جاما الجزئية لهذه البيانات يمكن توضيحها في الجدول (15 – 6).

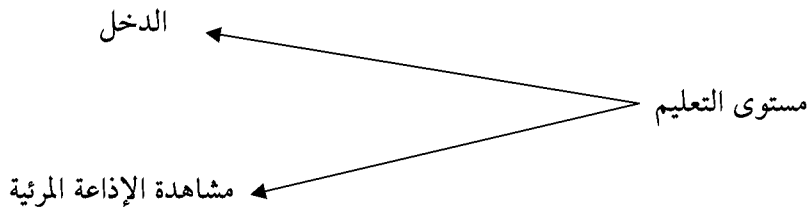
جدول (15 ـ 6) حساب جاما الجزئية

	الأزواج المنسجمة	الأزواج غير المنسجمة	جاما
الجدول الثنائي الأصلي	23.460 = 115 × 204	8360 = 95 × 88	0.47
الجدول الجزئي (1)	2040 = 102 × 20	1700 = 50 × 34	0.09
الجدول الجزئي (2)	2392 = 184 × 13	2430 = 45 × 54	-0.007
المجموع عبر الجداول الجزئية	4432 = 2392 + 2040	4130 = 2430+1700	0.04

إن قيمة جاما الجزئية لهذه البيانات هي فقط 0.04 مشيرة إلى وجود علاقة مباشرة ضئيلة بين الدخل ومشاهدة الإذاعة المرئية عندما تم إضافة متغير مستوى التعليم كمتغير متحكم[5].

العلاقة الكاذبة أو الدخيلة:

عندما كانت جاما الجزئية على درجة منخفضة إذا ما قورنت بقيمة جاما الأصلية التي تم حسابها في جدول التوافق المشترك يتوجب علينا أن نستنتج وجود إما علاقة كاذبة أو دخيلة بين المتغيرين الأصليين. وقبل تفسير كل نمط من أنماط هذه العلاقة، يحتم علينا الإشارة لأن نقرر أياً من النماذج يفسر النتائج المتعلقة بتوسيع الجداول هل هو تفسير نظري وليست مسألة إحصائية وعندما وجدنا أن العلاقة الأصلية قد اختفت بعد توسيع جداول التوافق، فالأمر يبقى لنا لنقرر كيف تتوافق هذه المتغيرات الثلاثة معاً استناداً إلى فهمنا إلى الواقع، فعلى سبيل المثال، يمكننا الاعتقاد أن النموذج الموضح أدناه (شكل 15 -3) هو أفضل تفسير للنتائج التي تم تحليلها الآن.



شكل رقم (15 ـ 3) علاقة كاذبة

إن هذا النموذج يوضح لنا العلاقة الكاذبة بين الدخل ومشاهدة الإذاعة المرئية بإدراك أن هذه العلاقة غير موجودة بين المتغيرين الأصليين وأن أن هذه العلاقة مجرد نتيجة إحصائية فقط استناداً على خصوصية علاقتهما بالمتغير المتحكم. فالمستوى التعليمي يؤثر بشكل منفصل على الدخل ومشاهدة الإذاعة المرئية، إلا أن المتغيرين الأخيرين ليسا مرتبطين ببعضهما البعض.

وقد نجد باحثاً آخر ينظر إلى هذه النتائج التي تحصلنا عليها من خلال توسيع جداول التقاطع بين الدخل ومشاهدة الإذاعة المرئية وبدلاً من ذلك يمكنه وصف العلاقة كما هي موضحة قي الشكل رقم (4).

الدخل ← المستوى التعليمي ← مشاهدة الإذاعة المرئية

شكل رقم (15 ـ 4) علاقة دخيلة

إن هذا الباحث يمكنه أن يجادل في أن أولئك الذين يكسبون دخلاً عالياً بإمكانهم الحصول على التعليم الجامعي، وهذا التعليم الجامعي يؤثر على كمية مشاهدة الإذاعة المرئية. ومهما كان اعتقادنا بأن هذه المناقشة جيدة أو غير ذلك، فهي مسألة ترجع في الأساس إلى كونها مسألة نظرية. أما إذا كان هذا التفسير أكثر ملاءَمةً لنتائج توسيع الجداول أكثر من نموذج العلاقة الكاذبة فهو أمر متروك لنقاش. لكن التحليل الإحصائي في حد ذاته لا يمكنه أن يقرر بين المسألة. فنتائج التحليل الإحصائي هي مجرد إشارة إلى أن واحداً من هذه النماذج يمكنه أن يفسر النتائج بشكل جيد.

العلاقة المشروطة:

نفترض أن باحثاً ما يرغب في معرفة إلى أي مدى يستجيب المرضى لبرنامج رياضي يهدف إلى تحسين نظام الأوعية الدموية (القلب) وقد قسم الباحث المرضى إلى مجموعتين: مجموعة ذات مستوى منخفض من الممارسة الرياضية والأخرى ذات مستوى عالٍ من الممارسة الرياضية ولاحظ ما إذا كان هناك أي تحسن في أنظمة الأوعية الدموية (انظر جدول رقم 7).

جدول (15 ـ 7) تحسن أنظمة الأوعية الدموية وفقا لمستوى الممارسة الرياضية

التحسن	المستوى الرياضي		المجموع
	منخفض	عال	
لا	38 % 73	11 % 34	49
نعم	14 % 27	21 % 66	35
المجموع	52	32	84

ومن خلال جدول (15 – 7) يمكننا أن نلاحظ أن خلايا النموذج لكل عمود يقترح أن هناك علاقة قوية وموجبة بين المتغيرين. إن البرنامج الرياضي يبدو أنه ذو فعالية. ولتعزيز هذه الفكرة فقد تم حساب جاما التي أنتجت قيمة تساوي 0.68.

إن الباحث يمكنه أن يتوقف عند هذه النتيجة، ويصل إلى قرار أن هناك علاقة مباشرة بين المتغير المستقل (مستوى التمارين الرياضية) والمتغير التابع (مستوى التحسن). وعلى أية حال، فإن الباحث يعتقد أن العلاقة الفعلية أكثر تعقيداً من هذا، إنه ربما توجد عوامل أخرى لم يضعها الباحث في هذا التحليل والتي يمكنها أن تقود إلى ما إذا كان قد تحسن نظام الأوعية الدموية (القلب) لدى المريض.

وبشكل خاص، يمكن للباحث أن يعتقد فيما إذا كان الشخص من المنتظمين في التدخين سوف يؤثر على فرصتهم للاستجابة لبرنامج الممارسة الرياضية. ومن هنا يتوجب على الباحث أن يعيد حساب جاما في جدول التوافق، وإدخاله متغير مستوى التدخين هذه المرة. والنتائج لهذا التحليل تظهر في الجدولين 8 و 9.

جدول (15 ـ 8): التحسن في الأوعية الدموية وفقا لمستوى الممارسة الرياضية (المدخنون فقط)

التحسن	مستوى الرياضة		المجموع
	منخفض	عال	
لا	28 % 74	7 % 70	35
نعم	10 % 26	3 % 30	13
المجموع	38	10	48
جاما = 0.09			

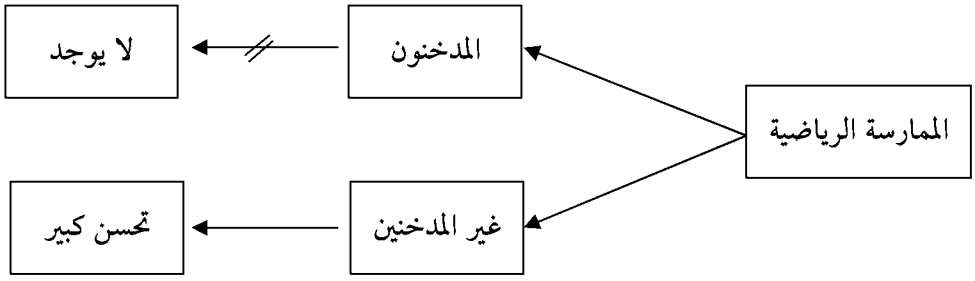
جدول (15 ـ 9) التحسن في نظام الأوعية الدموية وفقا لمستوى الممارسة الرياضية (غير المدخنين)

التحسن	مستوى الرياضة		المجموع
	منخفض	عال	
لا	10 % 71	4 % 18	14
نعم	4 % 29	18 % 82	22
المجموع	14	22	36
جاما = 0.84			

وعند مقارنة هذين الجدولين الجزئيين بالجدول الأصلي فإنه من الواضح أن العلاقة بين هذه المتغيرات تتباين اعتماداً على تاريخ الشخص المدخن. فالأفراد المنتظمين في التدخين لم يسجلوا أي تحسن في مستويات صحتهم كنتيجة لبرنامج الممارسة الرياضية. بينما سجل الأفراد غير المدخنين علاقة أقوى مما كانت عليه عند التعامل مع الجدول الأصلي. فالنتيجة قد ضعفت نتيجة لتضمين أولئك المدخنين مما أدى إلى عدم ثبات العلاقة وهذا يتأكد لنا

من خلال قيم جاما لكل جدول من هذه الجداول. فالجدول المتعلق بغير المدخنين فقط وصلت قيمة جاما إلى 0.84 مقارنة بـِ 0.68 للجدول الأصلي ككل. أما فيما يتعلق بالمدخنين، فإنه لا يوجد أي تحسن من برنامج الممارسة الرياضية، وبالتالي نجد أن النتيجة قد اختفت نتيجة لتأثير المتغير المتحكم. إن مقياس التطابق يعتبر مقيداً للغاية نظراً لأن مقياس التطابق يقيس التغيرات التي حدثت عندما أضيف متغير التحكم.

وكنتيجة لهذه الملاحظة فقد عمد الباحث إلى تغيير النموذج الذي يمكن أن يربط المتغيرات بعضها ببعض. وبدلاً من العلاقة البسيطة المباشرة ذات الاتجاه الواحد، فالباحث يمكنه أن يصف الارتباط في إطار العلاقة المشروطة كما هي مبينة في الشكل التالي:



شكل رقم (15 ـ 5) العلاقة المشروطة

إن العلاقة المشروطة يطلق عليها في بعض الأحيان (التفاعل)، فالتفاعل يكون موجوداً عندما تكون العلاقة بين متغيرين معتمدة على قيم خاصة بمتغير ثالث. وفي بعض الأحيان يمكننا أن نجد العلاقة عكسية اعتماداً على قيمة المتغير المتحكم؛ فقد تكون العلاقة لمجموعة فرعية علاقة موجبة، في حين يمكن أن تكون العلاقة سلبية في مجموعة أخرى فرعية.

مثال:

الآن يمكننا أن نستقصي العلاقة بين مقياس الذكاء (IQ) والدخل. فالذكاء مقياس وفقاً لمعيار اختبار IQ ويمكننا تقسيم المبحوثين إلى مجموعتين: مجموعة ذات مستوى ذكاء

منخفض والأخرى ذات مستوى ذكاء عالٍ. كذلك يمكننا تقسيم هؤلاء المبحوثين أيضاً إلى مجموعتين: مجموعة ذات دخل منخفض ومجموعة ذات دخل عالٍ وذلك طبقاً لمتوسط الدخل القومي لهذا المجتمع أو ذاك؛ إن النتائج المجمعة لـِ: 1000 شخص تم مسحهم يوضحها الجدول رقم (15 - 10)، ويوضح هذا الجدول أن هناك ارتباطاً متوسطاً بين الذكاء - كما تم قياسه باختبار IQ - والدخل، وقد تقود هذه النتيجة إلى تفسير إلى أن التباين في الذكاء يسبب التباين في مستوى الدخل. فقدرة الناس على الكسب تحدد من خلال ذكائهم الخاص. فهم إلى حد ما ومن أجل تحاشي مثل هذه النتيجة يمكننا أن نجادل بأن اختبار (IQ) كقياس للذكاء يكون اختباراً متحيزاً، إننا قد نشعر بشكل خاص أن درجات IQ هي انعكاس لخلفية الوضع الاجتماعي، وأن هذا المتغير هو الأساس في تحديد الدخل ولتقييم هذا الأمر يمكننا بناء جدولين جزئيين من خلال تقسيم الـ 1000 مبحوث إلى فئة اجتماعية عالية وفئة اجتماعية منخفضة، وإن هذه المجموعات الفرعية ولدت لنا النتائج التالية كما توضحها الجداول 15 - 10، 15 - 11، 15 - 12.

جدول (15 ـ 10) الدخل وعلاقته بالذكاء

الذكاء IQ	الدخل		المجموع
	منخفض	عال	
منخفض	165 %36	95 %18	260
عال	295 %64	445 %82	740
المجموع	460	540	1000
جاما = 0.45			

جدول رقم (15 ـ 11) الدخل وعلاقته بالذكاء
(الفئة الاجتماعية العليا فقط)

الذكاء IQ	الدخل		المجموع
	منخفض	عال	
منخفض	20 %18	60 %14	80
عال	90 %82	380 %86	470
المجموع	110	440	550
جاما = 0.17			

جدول رقم (15 ـ 12) الدخل وعلاقته بالذكاء
(الفئة الاجتماعية الدنيا فقط)

الذكاء IQ	الدخل		المجموع
	منخفض	عال	
منخفض	145 %41	35 %35	180
عال	205 %59	65 %65	270
المجموع	350	100	450
جاما = 0.13			

ومن خلال هذه الجداول يتضح أن هناك علاقةً ثنائيةً قويةً قد اختفت بشكل كبير عندما تم إدخال الخلفية الاجتماعية كمتغير للتحكم. وكما أنه يمكن ملاحظة أنَّ هناك فرقاً بسيطاً في نمط التوزيعات النسبية عبر الجدولين الجزئيين. في حقيقة الأمر أنه عندما تم حساب جاما الجزئية على أساس الجداول الجزئية وجدنا أن قيمة جاما فقط 0.15 في هذه الحالة يمكننا القول بأنّ لدينا علاقة كاذبة أو علاقة دخيلة [6].

الخلاصة:

في هذا الفصل قد تمت مناقشة الطريقة التي من خلالها تم إدخال متغير ثالث يمكنه إحداث تغيير في العلاقة التي تم ملاحظتها سابقاً بين متغيرين. حقاً، قد يكون الأمر أكثر تعقيداً عندما يسمح بوجود تأثير متغيرات أكثر على العلاقة الثنائية الأصلية. آخذين في الاعتبار احتمالية تأثير متغيرات أخرى متعلقة بالتحليل المتعدد، والذي سوف نتناوله في الفصل اللاحق. ولمساعدة القارئ في فهم هذا الفصل يمكننا أن نرسم بعض النتائج من خلال توسعنا في جداول التوافق (انظر جدول 15-13) الذي يقدم لنا دليلاً لاتخاذ القرار.

جدول (15 ـ 13) النتائج المحتملة عند إدخال متغير التحكم

الجداول الجزئية عند مقارنتها بجداول التقاطع يتبيّن:	النموذج	التطبيقات لتحليل أبعد	احتمالية الخطوة التالية في التحليل الإحصائي	التطبيقات النظرية
العلاقات نفسها بين X و y	علاقة مباشرة	غُض النظر عن متغير التحكم	اختيار متغير تحكم آخر لزيادة اختبار العلاقة المباشرة	نموذج يشير إلى أن متغير X يسبب y في علاقة مباشرة تم دعمها
علاقة ضعيفة أو لا توجد علاقة بين X و y	علاقة كاذبة أو دخيلة	إدخال متغير التحكم أو إدخال متغير متحكم	• التركيز على العلاقة بين هذه المتغيرات الثلاثة • التركيز على العلاقة بين هذه المتغيرات الثلاثة	• النموذج يشير إلى أن X تسبب y لمتغير لم يتم دعمه • النموذج الذي يشير إلى أن X تسبب y جزئياً ولكن يجب تعديل النموذج لأخذ متغير التحكم في الاعتبار
علاقات مختلطة	التفاعل / علاقة مشروطة	إدخال متغير التحكم	تحليل المجموعات الفرعية استناداً على متغير التحكم بشكل منفصل	النموذج الذي يشير إلى أن X تسبب y ثم دعمه جزئياً، ولكن يتوجب علينا تعديل النموذج من أجل أخذ متغير التحكم في الحسبان

المصدر: Joseph F. Healey , Statistics: A Tool for Social Research , Wadsworth Cengage Learning , USA , 1993 , P. 428.

أسئلة للمراجعة:

1- ما هي النتيجة التي يمكنك استنتاجها حول العلاقة بين X و y استناداً على البيانات التالية:

كل الحالات:

Y	X		مجـ
	1	2	
1	177	146	323
2	51	346	397
مجـ	228	492	720

التحكم في C1:

Y	X		مجـ
	1	2	
1	153	52	205
2	44	123	167
مجـ	197	175	372

التحكم في C2:

Y	X		مجـ
	1	2	
1	24	94	118
2	7	223	230
مجـ	31	317	348

2- دراسة أوضحت العلاقة بين العمر، الاهتمام بالبيئة، والانتماء السياسي.. أنتجت لنا قيم جاما التالية:

- جاما (العمر مع الاهتمام بالبيئة): 0.57-
- جاما (العمر مع الاهتمام بالبيئة) (الليبراليين فقط): 0.22-
- جاما (العمر مع الاهتمام بالبيئة) (المحافظين فقط) 0.67-

ما هي النتيجة المستنتجة حول هذه العلاقة؟ إذا كانت هناك علاقة بين هذه المتغيرات.

الهوامش والمصادر:

أولا: الهوامش:

1- George Argyrous , Statistics for Social and Health Research With a Guide to SPSS , Sage Publications , London ,2001 , P.461

2- انظر: عبد الله عامر الهمالي، التقنيات الإحصائية ومناهج البحث، ط 1، منشورات جامعة قاريونس، بنغازي، 2008، ص ص 88-94.

3- George Argyrous , op.cit, PP. 462 - 463.

4- Ibid , PP. 464 - 465.

5- Ibid , PP. 466 - 467.

6- Ibid , PP. 470 - 471.

ثانيا: المصادر:

1- George Argyrous, Statistics for Social and Health Research With Guide to SPSS, Sage Publications, London, 2001.

2- Joseph F. Healey, Statistics for Social Research, Wadsworth Cengage Learning, USA, 1993.

3- _____ , The Essentials of Statistics: A Tool for Social Research, Wadsworth Cengage Learning, USA, 2010.

الفصل السادس عشر

الارتباط الجزئي والانحدار المتعدد

مقدمة:

نحاول في هذا الفصل أن نتناول بعض التقنيات المستخدمة في تحليل العلاقات السببية التي تساعد الباحث في الوصول إلى تنبُّؤَات حول الظاهرة التي يسعى لدراستها. وتستند هذه التقنيات على ارتباط بيرسون (r) الذي تمتْ مناقشته في الفصل التاسع وتستخدم هذه التقنيات بشكل ملائم وعلى درجة عالية من الدقة، وتقاس متغيرات هذه التقنيات على مستوى المقياس ذي المسافات والنسبي.

إن أول هذه التقنيات التي سوف نناقشها في مستهل هذا الفصل هي تحليل الارتباط الجزئي.

أولاً: الارتباط الجزئي:

تعريفه: الارتباط الجزئي هو عبارة عن تقنيات إحصائية تسمح للباحث أن يفحص العلاقة الثنائية بين متغيرين، بينما يتحكم في المتغير الثالث، والذي يمكن الإشارة إليه في هذا الشأن بـِ (A) أو متغير التحكم.

حساب الارتباط الجزئي:

لحساب الارتباط الجزئي - بادئ ذي بدء - ينبغي على الباحث إجراء ارتباط بيرسون (r)، وبعد ذلك يمكنه الدخول في إجراءات حساب الارتباط الجزئي. وتجدر الإشارة هنا إلى الاختلاف بين الارتباط الثنائي (r) والارتباط الجزئي، حيث يفترض الباحث أن هناك تأثيراً للمتغير الثالث على العلاقة الثنائية بين المتغيرين تحت الاستقصاء. فعلى سبيل المثال، إذا كان الزوج المتعلم، والزوج غير المتعلم لديهما استجابات مختلفة فيما يتعلق بعدد الأطفال في الأسرة. فإن الارتباط الجزئي سيختلف في القوة، وربما في الاتجاه إذا قورنا بمعامل الارتباط الثنائي.

وقبل الدخول مباشرة في حساب معامل الارتباط الجزئي، دعنا نأخذ بعين الاعتبار العلاقات بين معامل الارتباط الجزئي ومعامل الارتباط الثنائي، وماذا تعني هذه الارتباطات حيث توجد ثلاثة أنماط محتملة من هذه العلاقات.

أنماط العلاقات[1]:

1- العلاقات المباشرة:

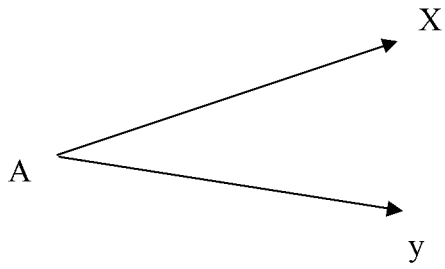
في الحقيقة إنه عندما يُدخل الباحث متغير التحكم في تحليل بياناته، فإنه قد يتحصل على أحد الاحتمالات، وهي أن معامل الارتباط الجزئي في الأساس قيمته مساوية للمعاملات الثنائية. فعلى سبيل المثال، إذا تم التحكم في تعليم الزوج، فإننا نجد أن قيمة معامل الارتباط الجزئي تصل إلى 0.49+ إذا قورنت بارتباط بيرسون (r) 0.50+. وهذا يعني إن إدخال المتغير الثالث لا تأثير له على العلاقة الأصلية (عدد الأطفال وعدد الساعات المتعلقة برعاية هؤلاء الأطفال). بمعنى آخر، أنه بغض النظر عن مستوى التعليم، فإن الأزواج يستجيبون بطريقة واحدة لمسألة رعاية الأطفال. إن مثل هذه النتيجة تكون متوافقة مع النتيجة التي تشير إلى العلاقة المباشرة أو السببية بين X و y.

y ⟶ X

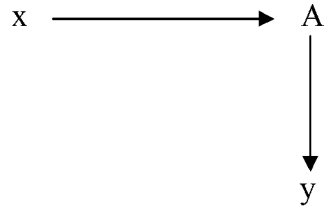
وأن المتغير الثالث لا علاقة له بالاستقصاء. وفي هذه الحالة ينبغي على الباحث أن يغض النظر عن المتغير الثالث (A) في التحليلات الإحصائية اللاحقة.

2- العلاقات الكاذبة والدخيلة:

إن الاحتمال الآخر الذي يمكن أن يحدث هو عندما تكون قيمة معامل الارتباط الجزئي ضعيفة إذا ما قورنت بمعامل الارتباط الثنائي، وربما يصل هذا الارتباط إلى صفر. وقد تكون النتيجة متساوية مع علاقات مختلفة بين المتغيرين. أولهما العلاقة الكاذبة Spurious Relationship. فالتحكم في متغير (A) هو سبب في وجود كلا المتغيرين، المتغير المستقل (X) والمتغير التابع (y)، كما هو موضح في الشكل التالي:



وتعني هذه النتيجة أنه لا توجد علاقة حقيقية بين X و y، ويظهر هذان المتغيران أنهما مرتبطان بعضهما ببعض فقط بوجود متغير التحكم A. وعندما يضع الباحث في اعتباره متغير التحكم (A)، فإن العلاقة المرئية بين X و y ستختفي. وثانيهما العلاقة الدخيلة بين X و y هذا النمط (يكون الارتباط الجزئي أكثر ضعفاً من معامل الارتباط الثنائي) متسقاً Consistent مع العلاقة الدخيلة بين المتغيرين كما هو موضح في الشكل التالي:



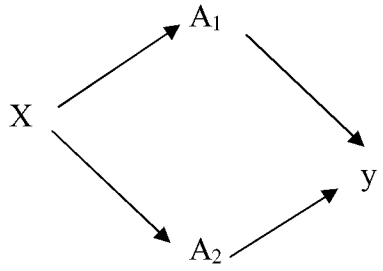
في مثل هذا الموقف، نجد أن X و y متغيران لا يرتبطان بعضهما ببعض مباشرة

ولكنهما - سبباً - مرتبطان بوجود التحكم (A). مرة أخرى، إنه عند التحكم في متغير (A) فإن العلاقة المرئية بين X و yستختفي.

إن السؤال الذي يمكن طرحه في هذا الخصوص هو: كيف يمكن للباحث أن يميز بين العلاقة الكاذبة والعلاقة الدخيلة؟ للإجابة عن هذا السؤال، يمكننا القول، إن التمييز بينهما لا يمكن القيام به من خلال الأسس الإحصائية: فالعلاقة الكاذبة والعلاقة الدخيلة هما شيء واحد من الناحية الإحصائية: فالباحث يمكنه أن يميز بين هاتين العلاقتين فقط من حيث ترتيب المتغيرات، من حيث الوقت (بمعنى أي من المتغيرين قد جاء أولاً) أو يمكنه التمييز بين هاتين العلاقتين على أسس نظرية *Theoretical grounds* وليس على أسس إحصائية.

التفاعل Interaction:

وأخيراً، فإن الاحتمال الثالث للعلاقة بين المتغيرين يمكن الإشارة إليه هنا، بالرغم من صعوبة اكتشافه من خلال تحليل الارتباط الجزئي، ويطلق على هذه العلاقة: التفاعل. وتحدث علاقة التفاعل بين X و y وتتغير بشكل ملحوظ تحت قيم متنوعة لـِ A. فعلى سبيل المثال، قد تكون هناك علاقة موجبة بين X و y لأحد فئات A، وعلاقة سالبة للفئة الأخرى كما هو مبين في الشكل التالي:



حساب وتفسير معامل الارتباط الجزئي:

المصطلحات والمعادلات: تتطلب معادلة الارتباط الجزئي بعض المصطلحات الجديدة. وسنتعامل هنا مع أكثر من واحد من العلاقات الثنائية، وبالتالي، نحتاج أن نميز بين هذه

العلاقات من خلال رموز معينة. وعليه، فإن الرمز ryx سوف يشير لمعامل الارتباط بين المتغير y والمتغير x، و ryA يشير إلى معامل الارتباط بين متغير y ومتغير A، و rxA يشير إلى معامل الارتباط بين X و A.

معامل الارتباط الجزئي:

إن أول علاقة جزئية يمكن الإشارة إليها بالرموز الجبرية التالية:

ryx.A

يمثل المتغير A إلى اليمين من النقطة، المتغير المتحكم وعليه، ryx.A تشير إلى معامل الارتباط الجزئي التي تقيس العلاقة بين المتغير x والمتغير y مع التحكم في متغير A. والمعادلة المتعلقة بالارتباط الجزئي:

$$ryx.A = \frac{ryx - (ryA)(rxA)}{\sqrt{1 - r^2\, yA}\sqrt{1 - xA}}$$

ملاحظة: ينبغي على الباحث أن يجري أولاً معامل الارتباط بين كل زوجين من المتغيرات (المتغيرات X و y و X و A، و y و X) قبل استخدام هذه المعادلة.

حساب معامل الارتباط الجزئي:

لتوضيح حساب معامل الارتباط الجزئي يمكننا استخدام المثال التالي:

مثال: البيانات التالية تمثل التدخين، الإصابة بالزكام، والإجهاد:

X التدخين	Y الإصابة بالزكام	A الإجهاد
70	2	4
105	6	3
35	1	2
105	4	3
0	2	2
70	7	3
35	3	1
140	6	5
0	4	5
140	4	3
$\overline{X}=70$	$\overline{y}=3.9$	$\overline{A}=3.10$
$S^2X=2450.0$	$S^2y=3.49$	$S^2A=1.49$
$SX=49.50$	$Sy=1.87$	$SA=1.22$

المصدر: George Diekhoff , Statistics for Social and Behavioral Sciences , Univariate , Bivariate , Multivariate , wm. c. Brown Publishers ,USA , 1992 , PP. 257 - 258

rxy =.53

rAy =.40

rAx =.29

$$ryx.A=\frac{ryx-(ryA)(rxA)}{\sqrt{1-r^2}\,yA\sqrt{1-r^2}\,xA}$$

$$=\frac{(.53)-(.40)(.29)}{\sqrt{1-(.40)^2}\sqrt{1-(.29)^2}}$$

$$= \frac{0.414}{\sqrt{0.84}\sqrt{0.92}}$$

$$= \frac{0.414}{(0.92)(0.96)}$$

$$= \frac{0.414}{0.883}$$

$$ryx.A = 0.47$$

التفسير: إن هذه القيمة التي تقيس العلاقة بين التدخين والإصابة بالزكام مع التحكم في متغير الإجهاد هي أقل من معامل الارتباط (r) ryx.53.

إن مربع الارتباط الجزئي هو $.47^2 = .22$ rxy.A =. وتشير هذه الدرجة إلى أن 22% من التباين للتعرض للإصابة بالزكام يمكن التنبؤ به من خلال عدد مرات التدخين عندما يتم التحكم في متغير الإجهاد المهني[2].

اختبار الدلالة للارتباط الجزئي:

إن الارتباط الجزئي بين التدخين (X) والإصابة بالزكام (y) والتحكم في الإجهاد المهني (A) تم حسابها .47= rxy.A.

إن السؤال الذي يمكن طرحه هنا هو: ما إذا كانت هذه القيمة للارتباط الجزئي كافية. بمعنى هل هذه القيمة من الارتباط الجزئي دالة إحصائياً؟

دعنا نرى ذلك من خلال اتّباع الخطوات التالية:

الخطوة الأولى: صياغة الفرض الصفري والفرض البديل:

Ho: إن القيمة المشاهدة لِـ rxy.A مرجعها إلى خطأ المعاينة. إن الارتباط الجزئي بهذا الحجم نسبياً يكون مرجحاً أنه يوجد في عينة سحبت من مجتمع يكون فيه قيمة rxy.A تساوي صفراً. rxy.A = 0

Hi: إن القيمة المشاهدة لـ rxy.A كبيرة جداً لأن تكون مرجحة لخطأ المعاينة. إن هذه القيمة من غير المحتمل، أن العينة المسحوبة من مجتمع تكون فيه rxy.A تساوي صفراً

وتعطي درجة عالية للارتباط الجزئي. وعليه، فإن احتمال العينة قد جاء من المجتمع الذي تكون فيه 0 < rxy.A، وأن هناك احتماليةً كبيرةً بأنه في حالة إعادة الدراسة فإن درجة الارتباط الجزئي تكون أكبر من صفر (0).

الخطوة الثانية: إحصاء الاختبار The Test Statistic:

إن الارتباط الجزئي rxy.A يساوي 47. يمثل إحصاء الاختبار للعلاقة بين X و y بعد التحكم في متغير A.

الخطوة الثالثة: تحديد الاحتمالية المرتبطة بإحصاء الاختبار.

بمقارنة القيمة المحسوبة لـِ rxy.A المساوية لـِ 47. لتوزيع المعاينة rxy.A نستطيع عندئذ rxy.A تحديد الاحتمالية القريبة بأن العينة التي تعطي هذه القيمة من الارتباط الجزئي قد جاءت من مجتمع فيه تكون قيمة rxy.A تساوي صفراً (0). ولدرجة حرية df=N-J=10-3=7، حيث (N) تشير إلى عدد الحالات و(J) تشير إلى عدد المتغيرات الخاضعة للتحليل، وباعتماد اختبار أحادي الجانب، فإن القيمة الحرجة لـِ (r) كما هو موضح في جدول توزيع قيمة (r) تحت مستوى دلالة 05. تكون القيمة مساوية لـِ rxy.A تساوي 47. تقع تحت القيمة الحرجة، فإننا بالتالي نصل إلى نتيجة مفادها أن الاحتمالية أكبر من 0.05. إن العينة المسحوبة من مجتمع تكون فيه rxy.A تساوي صفراً (0) ستعطينا ارتباطاً جزئياً يصل إلى 47. = rxy.A، وعليه فإن العينة المدروسة إلى حد كبير يرجح أنها تمثل هذا المجتمع. بمعنى آخر، إن العلاقة بين التدخين والإصابة بالزكام ليست دالة إحصائياً عند التحكم في الإجهاد المهني إحصائياً[3].

الارتباط الجزئي باستخدام SPSS:

1- إدخال البيانات Data Entry:

إدخال قيمة X في العمود الأول، وقيم y في العمود الثاني و A في العمود الثالث في محرر بيانات Spss.

2- تحليل البيانات Data Analysis:

1- انقر على Anayze ⟸ واختر Correlate وانقر على Partial.

2- في الوقت نفسه يلقى الضوء على قائمة العمود لـِ X و y على الجانب الأيسر من العمود ويتم تحريكها إلى مربع المتغيرات.

3- أُلْقِ الضوء على قائمة العمود لـِ A ويتم التحرك إلى مربع التحكم.

4- انقر على OK.

3- مخرجات Spss:

تمدنا هذه المخرجات بمصفوفة الارتباط Correlation Matrix، الارتباطات المحتملة بما فيها الارتباط بين X و X، والارتباط بين y و y. نحن نرغب في الارتباط الجزئي بين X و y التي يكون في أعلى زاوية من يمين المصفوفة. وتحت الارتباط يظهر مستوى الدلالة Significance Level. وعندما يكون مستوى الدلالة أقل من 05. فإن ذلك يشير إلى أن معامل الارتباط معامل دالة.

ثانياً: تحليل الانحدار المتعدد:

مراجعة الانحدار الثنائي Bivariate Regression:

في المثال التالي يمكننا أن نستقصي العوامل التي تؤثر في قيمة مبيعات ما. لقد قامت إحدى الوكالات للخدمات العقارية بجمع بيانات حول متغيرين لعدد 12 منزلاً. وقد جاءت النتائج وفقاً للجدول التالي:

جدول (16 ـ 1) حجم المنزل وأسعار البيع

سعر البيع ($،000)	حجم المنزل	سعر البيع ($،000)	حجم المنزل
260	20	287	24
240	18	252	20
245	20	270	23
210	13	275	25
230	18		
242	14		
295	28		
235	16		

المصدر: George Argyrous, Statistics for Social & Health Research, with a Guide to SPSS , SAGE Publications, London ,2001 , p. 475

إن الغرض من هذا التحليل هو تفسير التباين في سعر بيع هذه المنازل التي تم التعامل معها كمتغير تابع. وتعتقد وكالة الخدمات العقارية أن العامل الأساسي المفسر للتباين في أسعار البيع يرتبط بالتباين في مساحات هذه المنازل. وبالتالي يمكن أن نطلق عليه نموذج العوامل المحددة لسعر المنزل. وبما أن هذا النموذج هو نموذجٌ نظريٌّ يصف العلاقة التي يمكن أو لا يمكن أن تصمد أمام التدقيق الأمبيريقي.

دعنا نقارن، على سبيل المثال، منزلين من عينة هذه المنازل، كالمنزل الذي تم بيعه بقيمة 252,000 ألف دولار، بالمنزل الآخر الذي تم بيعه بقيمة 230,000 ألف دولار. من خلال هذه المقارنة، يمكننا القول، بأن قيمة البيع مرتبطة بحجم المنزل. ومن هنا نجد أن هذين المنزلين يبدو أنهما متناغمان مع نموذج هذه الوكالة للخدمات العقارية.

إن السؤال الذي يمكن طرحه هنا هو: هل هذه العلاقة تبقى صحيحة في كل المنازل الأثنى عشر؟

وبتطبيق تحليل الانحدار البسيط لهذه البيانات الواردة في الجدول أعلاه مستخدمين طريقة أقل المربعات المألوفة Method of Ordinary Least Square، تتولد لدينا النتائج التالية:

$$y = 157 + 4.88\ X$$

$$r = 0.92$$

$$r^2 = 0.85$$

وعلى ضوء هذه النتائج يمكننا التوصل إلى:

- هناك علاقة موجبة بين حجم المنزل وسعر البيع.
- أن أي زيادة في حجم المنزل تقود إلى الزيادة في سعره بحوالي 4880 دولاراً.
- إن هذه العلاقة، علاقة قوية وبدرجة عالية من الثقة لإجراء عملية التنبؤ.
- إن التباين في حجم المنزل ليس بالضرورة أن يكون على نحو كامل للتنبؤ بسعر المنزل، فمعامل التحديد، معامل قيمته عالية (0.85)، ولكنها ليست مساوية لواحد صحيح. ومع ذلك يمكن القول، بأن هناك عوامل أخرى تؤثر في سعر عينة هذه المنازل.

إن ثمن البيع الفعلي لأي منزل من هذه المنازل يمكننا التعبير عنه بالمعادلة التالية:

$$\text{ثمن البيع} = a + b\ (\text{حجم المنزل}) + e$$

وتشير هذه المعادلة إلى أن ثمن مبيعات هذه المنازل تتباين بشكل أساسي بسبب الفروق في أحجامها، وكذلك بسبب عوامل عشوائية تمثلت من خلال مصطلح الخطأ (e). ويشير مصطلح الخطأ (e) إلى الفرق بين ما تم التنبؤ به فيما يتعلق بثمن المنزل من خلال حجمه المحدد، والثمن الذي بَيْع به فعلاً.

وتجدر الإشارة إلى أن هناك عدة عوامل تؤثر في مبيعات هذه المنازل، إلا أن النموذج المعتمد لدى وكالة الخدمات العقارية ترى أنه بالرغم من العوامل الأخرى التي تؤثر في مبيعات هذه المنازل، إلا أنه من بين هذه العوامل لا يوجد متغير واحد (حجم المنزل) الذي يلعب دوراً أساسياً في تحديد ثمن البيع بطريقة نظامية وثابتة. وهذا ما حدا بنا إلى التركيز على متغير حجم المنزل وإعطائه وضعاً بيّناً في المعادلة. إلا أنه مع ذلك يمكننا القول، أيضاً إننا لا نريد أن نتجاهل العوامل الأخرى المؤثرة في ثمن بيع هذه المنازل.

إن مصطلح الخطأ (e) يخدم كل هذه العوامل الأخرى، فالعوامل التي تؤثر في ثمن بيع المساكن في كيفما اتفق وبطريقة غير نظامية. فقد يباع أحد المنازل بقيمة عالية، لأن وكالة الخدمات العقارية كانت، بشكل خاص، تطرح هذه المنازل للبيع بطريقة مغامرة ووضع ثمن معين لهذا المنزل، في حين يباع منزل آخر، لأن المشتري قد جذبه بشكل خاص نظام الألوان؛ في حين، يمكن أن يباع منزل آخر بثمن بخس لأن البائع Vendor يريد بيع المنزل بشكل سريع نظراً للدين محدد السداد، ونتيجة لهذه العوامل وعوامل أخرى، تنبثق عشوائياً من بيع إلى بيع آخر الذي لا ينبغي علينا أن نتعامل معها كمتغيرات مستقلة ومنفصلة. ولكن بالسماح لمصطلح الخطأ أن ينتزع تأثير هذه العوامل مجتمعة. ففي بعض الأحيان، هذه العوامل العشوائية تُسَبِّبُ زيادةً عاليةً في عملية البيع أكثر مما نتنبأ به استناداً على المعرفة المتعلقة بحجم المنزل، وفي أحيان أخرى تتسببُّ هذه العوامل في انخفاض ثمن البيع بحيث يكون أقل من القيمة المتوقعة. وبمعرفة حجم المنزل سوف يُسْمَح لنا بالتنبؤ بقيمة البيع التي سوف تكون قريبة من الثمن المستهدف. إلا أننا نسلّم بأن منزلاً معيناً فإن تأثير هذه العوامل العشوائية سوف تعني أن ثمن البيع الحقيقي ليس بالضرورة أن يكون مساوياً للثمن الذي تمَّ التّنُّبؤ به[4].

مقدمة للانحدار المتعدد:

تجدر الإشارة إلى أنه يمكن أن ننظر إلى النموذج الثنائي Bivariate Model بشكل مبسط أكثر مما ينبغي. فقد يتولد لدينا شعورُ بأن هناك عواملَ أخرى غيرَ عامل (حجم المنزل) قد لا تكون عواملَ عشوائيةً، ولكنها عواملُ تعملُ بطريقةٍ نظاميةٍ مُسَبِّبَةٍ التباينَ في قيمة مبيعات المنازل وتتباين هذه العوامل بشكل مستقل عن حجم هذه المنازل. بمعنى آخر، إذا قارنا منزلين لديهما نفس الحجم، فالفرق في ثمن البيع لهذين المنزلين قد لا يكون راجعاً فقط للعوامل العشوائية التي ناقشناها للتو. فإذا نظرنا إلى الجدول رقم (1). فإننا نجد أن هناك ثلاثةَ منازلَ في عينة الدراسة - على سبيل المثال - لديها نفس الحجم. فقد تم بيع أحد هذه المنازل بـِ 260,000 ألف دولاراً، والآخر بـِ 245,000 ألف دولاراً، والثالث بـِ 252,000 ألف دولاراً. فالسؤال الذي يمكن طرحه هنا هو لماذا هذه الفروق في أسعار البيع؟ بالرغم من أن هذه المنازل متساوية من حيث الحجم؟ إذا وضعنا ثقتنا

التامة في النموذج الثنائي والقول بأنَّ العوامل العشوائية مسئولة عن تفسير هذه الفروق. أو يمكننا القول، بأنَّ نموذجاً آخَرَ يمكن أن يسمح لفعالية متغيرات أخرى بشكل نظامي للتأثير على سعر البيع خَرَجْنَا بتفسيرات أفضل.

ويمكننا، على سبيل المثال، الاعتقاد بأن عمر المنزل أيضاً (وبشكل جزئي) يفسّر هذا التباين في أسعار هذه المنازل. بمعنى، أن عمر المنزل ليس متغيراً أحياناً يمكن أن يؤثر على ثمن بيع المنزل، ولكن بدلاً من ذلك، يعتبر عامل شائع وأنه على نحو منتظم يؤثر على ثمن بيع المنازل.

إن النموذج الجديد يمكن أن يظل قائماً بشكل معقول لتوقع أنه كلما زاد عمر المنزل قل سعره. بمعنى، أننا نتوقع أن هناك علاقة سلبية بين سعر المنزل وعمره. وإذا ما اعترانا الشك فإن ذلك كذلك، فإن الأمر يتطلب منا أن نوسع من تحليل الانحدار ليشمل هذا التحليل العملية المتعلقة بهذا المتغير الآخر، وهذه العملية تشبه إلى حد كبير نفس الطريقة التي اتبعناها في الفصل الخامس عشر المتعلق بالتوسع في التحليل الثنائي لجداول التقاطع والأخذ في الاعتبار التأثير المتوقع للمتغير الثالث. وعند التعامل مع بيانات المقياس ذي المسافات والنسبي كما هو الحال الآن يكون الهدف من وراء تحليل الانحدار المتعدد *Multiple Regression*. فالانحدار المتعدد هو ذلك التحليل الذي يستقصى العلاقة بين مجموعة المتغيرات المستقلة (متغيرين أو أكثر) ومتغير تابع.

واستناداً على هذا النموذج المتعدد، توضح البيانات الواردة في الجدول (16 – 2) عمر الاثنى عشر منزلاً.

جدول (16 ـ 2) ثمن البيع، حجم المنزل، عمر المنزل (12 منزلاً)

ثمن البيع ($0.000)	حجم المنزل	عمر المنزل (بالسنوات)
260	20	5
240	15	12
245	20	9
210	13	15
230	18	9
242	14	7
295	28	1
235	16	12
287	24	2
252	20	5
270	23	5
275	25	5

المصدر: George Argyrous, Statistics for Social & Health Research, op.cit , p. 478

بشكل عام، يمكننا التعبير عن العلاقة بين أي متغير تابع، وأي عدد من المتغيرات المستقلة (K) بالطريقة التالية:

$$y = a + b_1 x_1 + b_2 X_2 + \ldots b_k x_k + e$$

ومن خلال المثال الذي بين أيدينا، فإنه يمكننا أن نوضح نموذج العلاقة بالتعبيرات التالية:

$$a + b_1 \text{ (حجم المنزل)} + b_2 \text{ (العمر)} + e \text{ (ثمن البيع)}$$

بمعنى آخر، إننا نعتقد بأن ثمن بيع المنزل ينطلق في اتجاه واحد أو آخر وفقاً لعمر وحجم المنزل. وإننا بذلك نتوقع أن المنزل القديم ذا الحجم الصغير نسبياً سيئُول سعره نحو الاتجاه الأسفل عبر العملية المستقلة لكلا المتغيرين العمر والحجم. وبشكل عكسي، فإننا نتوقع أن منزلاً حديثاً وكبير الحجم نسبياً سيتجه سعر بيعه نحو الأعلى. وفي شواهد أخرى، فإن حجم وعمر المنزل من الممكن أن يتجها في اتجاهات معاكسة.

وبالرجوع إلى المعادلة السابقة. فإننا لازلنا نسمح للعوامل العشوائية أن يكون لها تأثير. وبالتالي فإن عمر وسعر المنزل لا يحددان تماماً ثمن البيع في كل حالة. ولكن إذا ما كان هذا النموذج المتعدد Multivariate Model هو الأفضل الذي يمكن الركون إليه في تفسير قيمة بيع المنازل، إذا ما قورن بالنموذج الثنائي Bivariate Model، فإن كمية التباين قد تركت لكي تفسر من خلال مصطلح الخطأ (e)، ستكون أصغر في النموذج المتعدد إذا ما قورنت بالنموذج الثنائي الذي بدأنا به في مستهل هذا الفصل. إلا أنه على الجانب الآخر، فإن إدخال العمر في المعادلة لم يكن قادراً على التقليل من نسبة التباين في ثمن البيع الذي يعزى إلى الخطأ (e). إذاً من خلال معرفتنا بعمر المنزل، ليس بالضرورة، أن يُحسِّن من قدرتنا بالتنبؤ بسعر بيعه. ومن هنا نكون قد أضعنا الوقت والجهد في جمع معلومات حول متغير لا فائدة ترجى منه.

إن مهمة الانحدار المتعدد هي أن يوزع التباين في سعر المنازل لكل واحد من هذه العوامل المتنافسة في التأثير على المتغير التابع، هل يهيمن أحد هذه العوامل على تحديد سعر البيع، كقولنا إن عمر أو حجم المنزل يكون بشكل واضح أكثر أهمية. أم أن هذين العاملين يكونا لديهما نفس التأثير؟

والسؤال الآخر الذي يمكن طرحه في هذا السياق، ما الدور الباقي للعوامل العشوائية؟ إن التحليل المتعدد، وفقاً لحساب معامل الانحدار أو الارتباط الجزئي لكل متغير يمدنا بمقاييس دقيقة للتأثير الخاص لهذه المتغيرات المستقلة على المتغير التابع. إنه من الممكن استخدام تقنيات عددية لحساب معامل الانحدار بين كل واحد من هذه المتغيرات المستقلة والمتغير التابع. ومع ذلك، فإن هذه التقنيات ستكون مُرْهِقَةً بشكل كبير ومُسْتَهْلِكَةً للوقت، خاصة إذا كان الباحث يتعامل مع مجموعة كبيرة من البيانات. وتجدر الإشارة هنا إلى أنه، ومع تقدم تقنيات الحاسوب، أصبح من السهولة بمكان على الباحث حساب الانحدار المتعدد من خلال البرامج المعدة لذلك مثل SPSS الذي يقوم بإنهاء العمليات الحسابية بشكل سريع ودقيق. وتبقى مهمة الباحث الأساسية هي تفسير هذه النتائج[5].

إجراءات الانحدار المتعدد باستخدام SPSS:

إن إجراء حساب الانحدار المتعدد يشبه إجراء حساب الانحدار الثنائي الذي تناولناه في الفصل التاسع من هذا الكتاب، ما عدا أننا في تحليل الانحدار المتعدد نقوم بلصق Paste أكثر من متغير إلى المتغيرات المستقلة Independent (s) إلى القائمة المستهدفة Target List.

الإجراء:

1- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Analyze ⟸ Regression ⟸ Linear

(يعطيك صندوق الحوار لـ Linear Regression

Dialog box)

2- انقر على Selling Price (سعر البيع) في قائمة المتغير Variable List.

3- أنقر على ◄ الذي يشير إلى المتغير التابع Dependent: القائمة المحددة للمتغير. هذه العملية ستقود إلى لصق Selling Price (Pastes) كمتغير تابع.

4- انقر على House Size (حجم المنزل) في قائمة المتغير، وبينما تضغط على مفتاح Shift، انقر على العمر بالسنوات Age in years.

5- انقر على ◄ الذي يشير إلى المتغير المستقل (المتغيرات) Independent (s) في القائمة المحددة للمتغيرات. هذه العملية ستقود إلى عملية لصق كلا المتغيرين House Size وعمر المنزل بالسنوات Age in years كمتغيرين مستقلين.

6- انقر على OK.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

Variables Entered / Removed[b]

Model	Variables Entered	Variables Removed	Method
I	Age in years House size (Squares[a])		ENTER

a. All Requested Variables Entered

b. Dependent Variable: Selling Price ($0.000)

Model Summary

Model	R	R Square	Adjusted R Square	Std Error of the Estimate
I	.959[a]	.919	.901	.807

a. Predictor: (Constant) , age in years , House size (Squares)

Anova[b]

Model		Sum of Squares	Df	Mean Square	F	Sig.
I	Regression	.7596248	2	24.37931	.29851	.000
	Residual	.158548	9	.90660		
	Total	.9176796	11			

a. Predictors (Constant) , age in years , House size (in Squares)

b. Dependent Variable: Selling Price ($0.000)

Coeffi cients[a]

Model	Unstandarized coefficients		Standariz ed coeffi cients	T	Sig.
	B	Std Error	Beta		
I (Constant)	.290224	.22226		.5538	.000
House size (Squares)	.5782	.973	.487	.6502	.026
Age in years	-2.974	.0761	-.503	-2.764	.022

a. Dependent Variable: Selling Price ($0.000)

المصدر: George Argyrous, Statistics for Social & Health Research, op.cit, p. 481.

شكل رقم (16 ـ 1) مخرجات SPSS للا نحدار الخطي

إذا نظرنا إلى هذه المخرجات، فإننا نجد كثيراً من المعلومات قد تولدت من خلال برنامج SPSS، وعليه، ينبغي على الباحث أن يركز على أهم الأجزاء في هذه المخرجات. فالجدول المُعَنْوَنُ بـِ Variables Entered / Removed، متغيرات أدخلت / أبعدت، يبين لنا وصفاً لفظياً بسيطاً للنماذج التي تمكننا من التكهن.

إن برنامج SPSS، بإمكانه أن يقوم بحساب عدة انحدارات متعددة بشكل تلقائي، وذلك من خلال استخدام عدة مجموعات من المتغيرات المستقلة ليحدد أيًّا من هذه المجموعات أو التوليفات تعطينا أفضل تفسير للتباين في المتغير التابع. هنا في هذا المثال، تنبأنا بنموذج واحد أطلق عليه نموذج (1)، الذي استخدم فيه متغير العمر بالسنوات وحجم المنزل بالمربعات كمتنبئ للمتغير التابع قيمة البيع (000.$).

أما الجدول الثاني، والذي أطلق عليه خلاصة النموذج Model Summary فإنه يبين لنا قيمة معامل الارتباط التي تشير إلى قوة العلاقة بين مجموعة المتغيرات المتوالفة في النموذج Model، والمتغير التابع. وتشير قيمة R إلى 959.، علاقة قوية. أن قيمة R هي الارتباط المتعدد مرادفاً لمعامل الارتباط الثنائي، (r). وأن R^2 (معامل الانحدار المتعدد للتحديد) تصل قيمتها إلى 919. وعندما يتم استخدام النموذج الثنائي Bivariate Model لتفسير سعر البيع (عندما تجرى عملية الارتباط فقط مع حجم المـنزل House Size) فإن قيمة معامل التحديد قد وصلت إلى 0.85. وعندما نستخدم كلا المتغيرين: حجم المنزل وعمره للتنبؤ بسعر البيع، فإن معامل التحديد ترتفع لتصل إلى 919.. وتشير هذه القيمة إلى قدرتنا على التفسير (أو التنبؤ) بقيمة بيع المنزل التي زادت عندما توفرت لدينا معلومات حول عمر وحجم المنزل. إن جزءاً من التباين في قيمة البيع التي أرجعناها سابقاً إلى عوامل عشوائية، هي في حقيقة الأمر تعزو للتأثير النظامي لعمر المنزل.

أما الجدول المُعَنْوَن بـِ ANOVA أنوفا فهو جدولٌ يحتوي على إحصائيات استدلالية inferential Statistics، تسمح لنا بإجراء استدلالات من خلال العينة على المنازل الأثنى عشرة. إلى هذه النقطة فإن التركيز يدور حول الإحصاءات الوصفية للعينة، وعليه، ينبغي على الباحث أن يتجاوز هذا الجزء من هذه المخرجات. والنظر إلى الجدول المُعَنْوَن بـِ Coefficients الذي يوفر عناصر معادلة الانحدار Elements of the regression equation

التي تم التنبؤ بها والتي يمكن كتابتها بالشكل التالي:

(حجم المنزل بالمربعات) 2.578 + 224.29 = قيمة البيع (\$.000)

(عمر المنزل بالسنوات) 2.974

وعند قراءة معادلة الانحدار، فإنه من الأهمية بمكان، أن نضع في أذهاننا وحدات القياس التي على ضوئها تم قياس المتغيرات. ففي المثال الذي بين أيدينا أن أي زيادة في كل مربع واحد في حجم المنزل تؤدي إلى الزيادة في سعر المنزل بقيمة تصل إلى 2578 دولاراً. وبالتحرر من هذه العلاقة، فقد وجدنا أيضاً أن زيادة كل سنة في عمر المنزل تقلل من قيمة البيع (لاحظ العلاقة السالبة) على نحو 2974 دولاراً.

دعنا الآن نتعامل مع شيء عملي. نفترض أننا نود أن نشتري منزلاً حجمه 15 متراً مربعاً وعمره خمس سنوات. ما هي توقعاتنا بالقيمة المطلوبة لبيعه؟ نضع المعلومات في معادلة الانحدار:

قيمة البيع (\$.000) = 224.29 + 2.578 (15) – 2.974 (5)

= 224.29 + 38.67 – 14.87

= 248.09

بطبيعة الحال، فإننا لا نتوقع أن يكون 248,090دولاراً الثمن الحقيقي الذي يتحقق عندما يتم بيع المنزل فعلياً، حيث إن العوامل العشوائية لازالت تلعب دوراً كبيراً في هذا الشأن. ولكن، وبوجود درجة عالية لمعامل التحديد، فإن هذه العوامل العشوائية ينبغي ألا تكون علة الثمن الفعلي للانحراف كثيراً عن القيمة المتوقعة.

إنه من الصعوبة بمكان استخدام معامل الانحدار لنقيّم الأهمية النسبية لكل متغير مستقل في تحديد قيمة المتغير التابع طالما أن كل متغير مستقل قد تم قياسه بوحدات مختلفة (أحدهما تم قياسه بعدد السنوات، والآخر بالمربعات). وإذا قمنا بقياس حجم المنزل بوحدة أخرى كالقدم المربع، حينئذٍ ستكون قيمة معامل الانحدار لهذا المتغير مختلفة وذلك

لاختلاف وحدة القياس التي تم اعتمادها في قياس هذا المتغير. بمعنى آخر، لا يمكننا القول بأن ذلك بسبب معامل حجم المنزل التي وصلت إلى 2.578. في حين أن قيمة المعامل المتعلقة بالعمر، قد وصلت إلى 2.974-. من هنا نجد أن متغير العمر متغيرُ قوي وله تأثير فاعل على قيمة البيع. وعمود المعامل المعيارية Standardized Coefficients توضح ذلك. وبدون الغوص في تفاصيل كيف تم حساب هذه المعامل المعيارية (يطلق عليها أيضاً أوزان بيتا beta - weights ، فإننا ببساطة نلاحظ أن هذه الأوزان (المعامل) قد أبعدت تأثير وحدات القياس. إنه باستطاعتنا أن نرى أن العمر (508.-) لديه قوة تأثير طفيفة على قيمة البيع، أكثر من التأثير الذي يحدثه حجم المنزل (487.).

الجدول التالي يلخص الدور الذي تلعبه المقاييس المختلفة التي يولدها برنامج SPss:

جدول رقم (16 ـ 3) تفسير مخرجات SPSS

معامل الانحدار (Regression Coefficients) ←	تسمح لنا بإجراء التنبؤ على المتغير المعياري (التابع) استناداً على قيم المتغيرات المستقلة في إطار الوحدات الأصلية للقياس.
المعامل المعيارية (Standardized Coefficients) ←	تسمح لنا أن نميز بين الأهمية النسبية لكل متغير في تحديد قيمة المتغير التابع.
R ←	تشير إلى قوة العلاقة بين مجموعة المتغيرات المستقلة (كتلة واحدة) والمتغير التابع (المعياري).
R^2 ←	تشير إلى كمية التباين في المتغير الذي تم تفسيره خلال مجموعة المتغيرات المستقلة في النموذج (1)، بذلك إشارة إلى ما إذا كان النموذج (1) نموذجاً جيداً للتنبؤ بالمتغير المعياري.

المصدر: George Argyrous, Statistics for Social & Health Research, op.cit, p. 482.

اختبار الدلالة للنموذج المتعدد:

إن المعلومات الحرجة لاختبار الاستدلال يوضحه الجدول المعنون بـ ANOVA. ويقوم برنامج SPSS بإجراء اختبار F على النموذج بكامله لاختبار الفرضية المتعلقة بمعامل الارتباط Correlation Coefficients لكل المتغيرات التي يحويها النموذج Model. وهي صفر (0). في هذا المثال الذي بين أيدينا، فإن إحصاء اختبار F لهذا النموذج يكون مستوى الدلالة فيه 0.000 وهذا بين لنا أنه على الأقل أن أحد الارتباطات بين كل واحد من المتغيرات المستقلة والمتغير المعياري ليست مساوية لصفر في المجتمع. وتؤكد هذه النتيجة ما تم الحصول عليه في جدول معامل (Coefficients). حيث إن قيمة t الإحصائية لكل متغير مستقل أنها ذات دلالة على مستوى 0.05.

وعليه، فإنه باستخدامنا لاختبار F (F-test) نرى ما إذا كان على الأقل بعض المتغيرات المستقلة في النموذج هي دالة، وأن إحصاءات t (t-Stataitics) لكل متغير على حدة يشير إلى أي من هذه المتغيرات دالة إحصائياً[6].

الانحدار التدرجي Stepwise Regression:

قد يلاحظ مكتب الخدمات العقارية أن هناك متغيرات أخرى ذات أهمية كتلك المتغيرات التي تعاملنا معها، تلعب هي الأخرى دوراً أساسياً في تحديد قيمة بيع تلك المنازل في هذه المنطقة. وبالرغم من القوة التفسيرية للنموذج ذي المتغيرين المستقلين، فإن مكتب الخدمات العقارية يمكنه أن يجادل في أن قدرتنا في التنبؤ بأسعار هذه المنازل ستكون أفضل عند إدخال متغير مساحة الأرض المقام عليها المنزل، كمتغير آخر مستقل. وعليه، فإن مكتب الخدمات العقارية سيقوم بجمع معلومات إضافية حول مساحة الأرض المقامة عليها هذه المنازل الأثنا عشر. والجدول التالي يوضح هذه البيانات الإضافية.

جدول (16 ـ 4) سعر البيع، حجم المنزل، عمر المنزل
والمساحة المقام عليها المنزل (12 منزلاً)

سعر البيع ($0.000)	حجم المنزل (مربعات)	عمر المنزل (بالسنوات)	مساحة الأرض المقام عليها المنزل (بالمتر المربع)
260	20	5	420
240	15	12	640
245	20	9	600
210	13	15	590
230	18	9	700
242	14	7	720
295	28	1	624
235	16	12	590
287	24	2	710
252	20	5	630
270	23	5	700
275	25	5	710

المصدر: George Argyrous, op.cit , p. 483.

الآن، لدينا ثلاثة نماذج لتفسير سعر بيع هذه المنازل. إن أول هذه النماذج: النموذج الثنائي الذي أخذ في اعتباره حجم المنزل، وعلاقته بسعر البيع. وثاني هذه النماذج النموذج المتعدد الذي أخذ في اعتباره حجم وعمر المنزل وعلاقتهما بسعر البيع. في حين يركز النموذج المتعدد الثالث على متغيرات حجم المساحة كمتغير تفسيري:

نموذج (1) سعر البيع = $a + b_1$ (حجم المنزل) + e

نموذج (2) سعر البيع = $a + b_1$ (حجم المنزل) + b_2 (عمر المنزل) + e

نموذج (3) سعر البيع = $a + b_1$ (حجم المنزل) + b_2 (عمر المنزل) + b_3 (مساحة الأرض) + e

إننا لِلتّوِّ قَدْ ناقشنا الطريقة التي من خلالها يمكن الحكم ما إذا كان متغير ما يضيف إلى القوة التفسيرية للنموذج من خلال النظر إلى تأثيره الضمني على قيمة مربع R. فإذا زادت قيمة R تربيع بشكل دالٍ عندما أضيف متغير إلى النموذج. حينئذٍ فالمعلومات الإضافية التي وفرها لنا هذا المتغير المضاف زادت من قدرة النموذج في تفسير التباين في سعر البيع.

تجدر الإشارة إلى أن هناك طريقةً واحدةً لنقرر من خلالها بين عدة نماذج، وهي الشروع في إجراء انحدارات خطية منفصلة استناداً على مجموعة مؤتلفة من المتغيرات المستقلة التي نود أن يشملها النموذج. وبعد ذلك، يمكننا أن نقارن قيم R تربيع لنرى إلى أي مدى، تكون أيٌّ من قدرتنا قادرةً على تفسير التباين في سعر البيع الذي يصل إلى مداه الأعلى عن طريق كل واحد من هذه المجموعة المؤتلفة من المتغيرات المستقلة. على سبيل المثال، إذا قمنا بإجراء الانحدار المتعدد مضافاً إلى هذه الأجزاء مساحة الأرض، فإن قيمة R تربيع ستصل إلى 0.922 وهي قيمة مساوية للقيمة التي تحصلنا عليها في النموذج (2). الذي يحتوي على متغيرين فقط هما عمر المنزل وحجمه. بمعنى آخر، إن مساحة الأرض المقام عليها المنزل لا تؤدي إلى زيادة قدرتنا في تفسير سعر البيع، فالوقت والجهد في قياس هذا المتغير، هما في واقع الأمر عمل ضائع[7].

إن المشكلة الكامنة في هذه الطريقة، أنها طريقة مملة في القيام بإجراء انحدارات منفصلة لكل واحد من هذه النماذج الممكن بناؤها. إضافة إلى ذلك، صعوبة الحكم على مقدار الزيادة في R تربيع لتبرير ما يتضمنه أي متغير في النموذج. إلا أنه من حسن الطالع أن برنامج SPSS قادر على أن يمدنا بطريقة يطلق عليها الانحدار التدرجي حيث تسمح لنا هذه الطريقة بتحديد أي من المجموعة المتآلفة من المتغيرات المستقلة تقدم لنا أفضل تفسير للمتغير المعياري. وتتم هذه العملية بإضافة وحذف متغيرات من العملية الحسابية طبقاً إلى ما إذا كان كل واحد من هذه المتغيرات يحدث دلالة إحصائية تكون سبباً في تغير قيمة R تربيع[8].

إجراءات الانحدار المتعدد التدرجي باستخدام SPSS:

1- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Analyze ⟸ Regression ⟸ Linear

(تقود هذه العملية إلى مربع الحوار (Linear Regression

2- انقر على Selling Price (سعر البيع) في قائمة المتغير.

3- انقر على ◄ التي تشير إلى Dependent في القائمة المحددة للمتغير (هذه العملية تقود إلى لصق Selling Price (Pasts) (كمتغير تابع).

4- انقر على House Size (حجم المنزل) في قائمة المتغير، مع الإبقاء ضاغطاً على مفتاح Shift والنقر على Age in years (العمر بالسنوات) وبعدها انقر على Land Size (مساحة الأرض).. (هذه العملية تلقي الضوء على House Size , Age in years and Land Size).

5- انقر على الـ ◄ التي تشير إلى المتغيرات المستقلة Independent (s) في القائمة المحددة للمتغيرات. (هذه العملية تقود إلى لصق Pasts، House Size , Age in years , Land Size كمتغيرات مستقلة).

6- انقر على ◄ القريبة من Enter (تظهر قائمة Adrop - down menu).

7- في القائمة drop-down menu انقر على Stepwise (تقود هذه العملية إلى اختيار الانحدار التدرجي Stepwise regression كطريقة لتضمن وتبعد المتغيرات من الانحدار.

8- انقر على OK.

فيما يلي المخرجات الناتجة عن هذا الإجراء :

Regression

Variables Entered / Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Age in years		Stepwise (criteria) Probability-of-f-to Enter=.050 Probability-of-f-to remove=1.00
2	House size (Squares)		Stepwise (criteria) Probability-of-f-to Enter=.050 Probability-of-f-to remove=.100

a. Dependent Variable: Selling Price ($0.000)

M0del Summary

Model	R	R Square	Adjusted R Square	Std Error of the Estimate
1	.925a	.856	.842	.999
2	.959b	.919	.901	.807

a. Predictors: (Constant) , age in years.

b. Predictors: (Constant) , age in years , House size (Squares).

ANOVA[b]

Model	Sum of Square	Df	Mean Square	F	Sig
1 Regression	.9995820	1	.9995820	.84859	.000 a
Residua	5.91897	10	.59297		
Total	6.917879	11			
2 Regression	.7596248	2	24.37831	.29851	.0006 b
Residua	.158548	9	.90880		
Total	.9176796	11			

a. Predictors (Constant) , age in years.

b. Predictors (Constant) , age in years , House size (Squares)

c. Dependent Variable: Selling Price ($0.000)

Coeffi cients[a]

Model	Unstandarized coefficients		Standarized coefficients	T	Sig.
	B	Std Error	Beta		
1 (Constant)	.702292	.8325		.19350	.000
Age in years	-5.419	702	-.925	-7.723	.000
2 (Constant)	.290224	.22226		.5638	.000
Age in years	-2.974	.0761	-.509	-2.764	.022
House size (Squares)	.5782	397	.487	.6502	.026

a. Dependent Variable: Selling Price ($0.000)

Excluded Variables[a]

Model	Beta IN	t	Sig.	Partial correlation	Collinearit Statistics
					Tolorance
1 House size (Squares)	.487a	.6502	.026	.662	.265
Land size in meters squares	.015	.115	.911	.038	.973
2 Land size in meters squares	.015b	.147	.886	.052	.973

a. Predictors in the model (Constant) , age in years.

b. Predictors in the model (Constant) , age in years , House size (Squares)

c. Dependent Variable: Selling Price ($0.000)

المصدر: .George Argyrous, Statistics for Social & Health Research, op.cit , p. 486

شكل (16 ـ 2): مخرجات SPSS للا نحدار الخطي

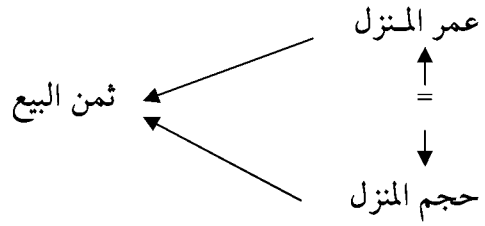
ولتفسير هذه المخرجات:

الجدول الأول المُعَنْوَن بـِ متغيرات / أدخلت وأبعدت Variables Entered / Removed في هذا الجدول تم توليد نموذجين من ثلاثة متغيرات تم افتراضها: الأول متضمناً متغير العمر بالسنوات (النموذج الثنائي الأساسي) والذي أطلق عليه في برنامج SPSS ، نموذج (1) (Model 1)، والنموذج الآخر يشتمل على متغير العمر وحجم المنزل (بالمربعات)، والذي أطلق عليه في برنامج SPSS نموذج (2) (Model 2)، أما باقي المخرجات فهي أساساً نفس المخرجات السابقة التي تم توليدها بشكل منفصل من قِبل كل واحد من هذه النماذج، وهنا لدينا نموذجان تم تقديمهما بنفس التحليل. أما الجزء الجديد من المخرجات فيوضحه الجدول الأخير المُعَنْوَنُ بـِ "Excluded Variables"، المتغيرات المستثناه. ويوضح هذا الجدول، وعلى أساس اختبار F وعلى أساس التغيرات في R^2. إن مساحة الأرض المقام عليها المنزل ليست متغيراً مهماً يمكن أن يضمن في أي من هذه النماذج Models. ويعني هذا كله، أن تعقيد النموذج بإضافة هذا المتغير الجديد لن يُؤَمن لنا أي نوع من الدقة فيما يتعلق بتقدير المتغير التابع، وبالتالي ينبغي الاستغناء عنه[9].

الافتراضات التي ينبغي مراعاتها عند استخدام الانحدار المتعدد:

على الرغم من قوة الانحدار المتعدد كأداة قوية لتقييم تأثير مجموعة من المتغيرات المستقلة على المتغير المعياري، إلا أن هناك مجموعةً من الافتراضات حول هذا التحليل تقلل من تطبيقاته. إن كل الافتراضات التي تمت مناقشتها عند الحديث عن الانحدار الثنائي فهي قائمة في تطبيقها على حالة الانحدار المتعدد. إضافة إلى هذه القائمة فإننا بحاجة إلى إضافة افتراض آخر مهم وهو أن التحليل المتعدد يفترض أن كل متغير من المتغيرات المستقلة مستقل عن الآخر. أي تسامت متعدد Mutticollinearity (أي التعدد الذي يقع على نفس الخط).

في المثال الذي بين أيدينا يمكن توضيحه بالشكل التالي:



إن كلاً من عمر وحجم المنزل يؤثران في سعر بيعه، ولكنهما لا يؤثران في بعضهما البعض. وقد يبدو لنا أن هذا الأمر معقول لهذه المتغيرات تحديداً: فإذا تم بشكل مفاجئ توسيع المنزل، فإن هذا الأمر لا يقود بشكل مفاجئ لأنْ يجعل من هذا المنزل منزلاً عتيقاً أو جديداً! وبشكل مشابه إذا ما زاد عمر المنزل، لا يعني بالضرورة أن يزيد حجمه أو يصغر. إن هذه الفرضية تحتوي ضمناً الانحدار المتعدد، مما يجعله إلى حد ما أكثر تعقيداً من التقنيات المتعددة الأخرى التي تناولناها في الفصل السابق، حيث تم استخدام التحليل المتعدد، لتحديد أي من هذه النماذج يقدم لنا أفضل تفسير للعلاقة بين المتغيرات الثلاثة أو أكثر منها[10]. ومن خلال تحليل الانحدار، فإننا نفترض نموذجاً محدداً كما تم في مخرجات SPSS شكل رقم (1)[11].

حدود استخدام الانحدار المتعدد والارتباط:

يعتبر الارتباط الجزئي، والارتباط المتعدد، تقنيات إحصائية قوية لتحليل العلاقات المتبادلة بين ثلاثة أو أكثر من المتغيرات. وتسمح هذه التقنيات للباحث بالقدرة على التنبؤ بدرجات متغير واحد من اثنين أو أكثر من متغيرات أخرى.

ولما كانت هذه التقنيات على درجة عالية من القوة فهي بالتالي تتطلب جهداً كبيراً، ودرجة عالية من البيانات، وقياساً على المستويين ذي المسافات والنسبي، وهذه متطلبات، عادة يصعب على الباحث إنجازها. إضافة إلى ذلك، فإن هذه التقنيات تستند على افتراضات: أن العلاقة المتبادلة بين المتغيرات تتبع شكلاً محدداً.

أولاً، تفترض هذه التقنيات أن كل متغير مستقل لديه صبغة العلاقة الخطية مع المتغير التابع. ولكي يتعرف الباحث على هذه العلاقة الخطية عليه القيام بمراجعة لشكل الانتشار.

ثانياً: تفترض هذه التقنيات أن تأثير المتغيرات المستقلة هي تأثيرات إضافية additive. وهذا يعني، أنه ينبغي علينا أن نفترض أفضل تنبؤ للمتغير التابع (y) يمكن الحصول عليه، وذلك من خلال إضافة درجات المتغير المستقل بعضِها إلى بعضٍ، على سبيل المثال (الفقر عندما يوجد مع الانعزالية) يولد درجة عالية من معدلات الجريمة.

ثالثاً: أن تقنيات الانحدار المتعدد، ومعامل الارتباط تفترض أن المتغيرات المستقلة متغيرات غير مرتبطة. هذه الحالة تعني أن معامل الارتباط بين كل أزواج المتغيرات المستقلة ينبغي أن تكون صفراً. في الممارسة العملية، ومع ذلك نعمل كما لو أن هذا الافتراض قد تم تحقيقه، إذا كانت الارتباطات المتبادلة بين المتغيرات المستقلة على درجة منخفضة.

وأخيراً، يجب الإشارة إلى أننا في هذا الفصل قد غطينا فقط التطبيقات البسيطة للارتباط الجزئي، والانحدار المتعدد، من حيث المنطق والتفسير، ولزيادة التفصيل لهذه التقنيات عندما يكون لدينا متغيرات مستقلة كثيرة واضحة المعالم نسبياً، ومع ذلك، فإن حسابات مثل هذه المواقف تعتبر معقدة جداً. وعندما يواجه الباحث مثل هذه المواقف لمتغيرات تزيد عن ثلاثة حينئذٍ يتطلب منه الاستعانة ببرنامج SPSS أو SAS للتعامل مع هذا الكم الهائل من المتغيرات[12].

أسئلة للمراجعة:

1- من البيانات التالية:

المدينة	الاحتجاج (y)	معدل البطالة (x)	% الدعاية السلبية
A	55	5	60
B	60	8	63
C	65	9	55
D	68	9	53
E	70	10	44
المتوسط	63.6	8.2	$\bar{x}$ = 55.8
الانحراف	5.5	1.7	S = 5.3
الاحتجاج	0.95 معدل البطالة		% الدعاية السلبية −0.87
معدل البطالة	-		−0.70

المطلوب:

a) حساب الارتباط الجزئي بين y و x، والتحكم في A ما هو التأثير الذي يحدثه متغير التحكم على العلاقة الثنائية. هل العلاقة بين الاحتجاج ومعدلات البطالة علاقة مباشرة.

b) احسب معامل الارتباط الجزئي بين (y) الاحتجاج، و (x) الدعاية السلبية متحكماً في معدل البطالة (A). ما هو التأثير الذي يحدثه متغير التحكم على العلاقة الثنائية. هل العلاقة بين الاحتجاج والدعاية السلبية علاقة مباشرة؟

c) اكتب ملخصاً قصيراً يبين نتائج هذه العلاقات بين المتغيرات الثلاثة.

2- من البيانات التالية: احسب معامل الارتباط الجزئي لعدد 12 حالة، مستخدماً اختبار أحادي الجانب:

rxy =.79

rAx =.81

rAx =.72

3- قام باحث بدراسة العوامل التي من الممكن أن تقود إلى التقليل من عدد أيام العمل والتي فقدت نتيجة للمرض، في مصنع معين. عشرة أشخاص تمت دراستهم. والبيانات التالية حول عدد الساعات التي يمارسون فيها التمارين الرياضية في الأسبوع، وكذلك أيام العمل التي غابوا فيها نتيجة لتعرضهم للمرض، يوضحها الجدول التالي:

عدد ساعات التمرين	عدد الأيام المفقودة	العمر بالسنوات
3	12	36
8	10	35
1	10	54
0	15	42
0	18	41
4	7	25
7	7	32
2	14	39
5	9	43
0	16	29
9	8	32
3	10	50

المطلوب:

- بين المتغير المعياري.
- ماذا تتوقع أن تكون الإشارة أمام المتغيرات المستقلة.
- ما هي معادلة الانحدار؟
- احسب معامل الانحدار المتعدد.
- هل تضمين العمر في النموذج يضيف شيئاً لقدرتنا على التنبؤ بعدد الساعات التي فقدت بسبب المرض.

4- بين كيف تم تطبيق تحليل الانحدار التدرجي في مجال البحث الاجتماعي، وما هي أسس هذا التطبيق؟

الهوامش والمصادر:

أولا: الهوامش:

1- Joseph F. Healey , The Essentials of Statistics: A Tool for Social Research , Wadsworth Cengage Learning , USA , 2010. pp. 362 - 364.

2- George DiEkhoff, Statictics for social and Behavioral Sciences: Univariat, Bivariate, Multivarite, wm. c. Brown Publishers , USA , 1992 , P. 261.

3- Ibid , P. 263.

4- George Argyrous , Statistics for Social and Health Research With Guide to SPSS , Sage Publications ,London , 2001 , P. 477.

5- Ibid , P. 479.

6- Ibid , P. 483.

7- Ibid , P. 484.

8- Ibid , P. 484.

9- Ibid , P. 485.

10- Ibid , P. 487.

11- Ibid , P. 487.

12- Joseph F. Healey , op. Cit. PP. 375 , 377 , 379.

ثانيا: المصادر:

1- George Argyrous , Statistics for Social and Health Research With Guide to SPSS , Sage Publications , London , 2001.

2- George Diekhoff, Statistics for Social and Behavioral Sciences, Univariate, Bivariate, Multivariate, wm. c. Brown Publishers, USA, 1992.

3- Joseph F. Healey, The Essentials of Statistitics: A Tool for Social Research, Wadsworth Cengate Learning, USA. 2010.

4- جولي بالانت (ت) خالد العامري، التحليل الإحصائي باستخدام برامج SPSS، ط 2، دار الفاروق للنشر والتوزيع، القاهرة، 2009.

5- عبد الله عامر الهمالي، التحديث الاجتماعي: معالمه ونماذج من تطبيقاته، منشورات جامعة قاريونس، 2008 م.