الفصل الثامن

تحليل الارتباط

Correlation Analysis

(8-1) مقدمة :

تعرضنا في الفصول السابقة لدراسة ظاهرة واحدة كالاستهلاك، الدخل، المبيعات، العمر، والوزن وبدأنا ببيان كيفية عرض البيانات التي جمعناها في جدول أو أشكال بيانية مختلفة، واستطردنا بعد ذلك إلى كيفية تلخيص تلك البيانات من خلال مقاييس النزعة والتشتت لمجموعة واحدة أو مجموعات مختلفة، وكيفية تقدير تلك المقادير، وكيفية إجراء اختبارات الفروض عليها، والآن نبحث كيفية إيجاد العلاقات الرياضية التي تربط المتغيرات ببعضها، وما مقدار هذا الارتباط من خلال ما يسمى مقاييس الارتباط المختلفة التي سوف نتعرض لها في هذا الفصل، كما نناقش في الفصل الآتي كيف يمكن التنبؤ بأحد المتغيرات لقيمة محددة للمتغير الآخر وهي ما تسمى معادلة الانحدار.

نستطيع القول إن تحليل الارتباط والانحدار هو أداة إحصائية نستفيد منها في تحديد العلاقة بين متغيرين أو أكثر للتنبؤ بأحد المتغيرات استنادا إلى قيم المتغير أو المتغيرات الأخرى. فمثلا إذا علمنا العلاقة بين مصروفات الدعاية والمبيعات، فيمكننا الاستفادة من تحليل الارتباط للتنبؤ بالمبيعات حالما تتوافر لنا قيمة نفقات الدعاية.

وسوف نتناول في هذا الفصل قضية الارتباط بين متغيرين سواء كانا كميين أو غير ذلك، بينما في الفصل القادم نتناول كيفية التنبؤ بأحد المتغيرات في ضوء متغير آخر أو أكثر من خلال دارسة الانحدار.

(8-2) الارتباط الخطي البسيط Simple Correlation

تسمى العلاقة بين ظاهرتين بالارتباط Correlation مثلاً العلاقة بين الدخل والاستهلاك، فمن البدهي أن زيادة دخل الفرد تؤدي إلى زيادة استهلاكه من السلع والخدمات (علاقة طردية)، كما أن ارتفاع سعر سلعة ما يؤدي إلى تدني الطلب عليها (علاقة عكسية) علماً بأن الارتباط قد يكون خطياً Linear أو غير خطي Non Linear. إن المقياس المستخدم الذي يقيس درجة الارتباط يعرف بمعامل الارتباط Correlation Coefficients ويرمز له r وتراوح قيمته بين 1- و 1 $[-1 \le r \le 1]$.

ومن الأمثلة على ذلك:

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

يحسب معامل الارتباط الخطي البسيط بافتراض وجود علاقة خطية بين اثنين من المتغيرات فقط مع العلم أن الحصول على قيمة صغيرة (قريبة من الصفر) لهذا المعامل لا يعني عدم وجود علاقة بين المتغيرين، فقد توجد علاقة من الدرجة الثانية (ارتباط غير خطي). ويختلف نوع المقاييس الذي نستخدمه في حساب معامل الارتباط طبقًا لنوع البيانات، وسوف يجري حسابه في حالة البيانات الكمية، والبيانات الوصفية المقاسة بمعيار ترتيبي.

الغرض من تحليل الارتباط الخطي البسيط

الغرض من تحليل الارتباط الخطي البسيط هو تحديد نوع وقوة العلاقة بين متغيرين، ويرمز له في حالة المجتمع بالرمز ρ (وتقرأ "رُو")، وفي حالة العينة بالرمز r، وحيث إننا 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 ثلاثة أنواع حسب إشارة معامل الارتباط كما يلي:

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

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

✓ إذا كانت قيمة معامل الارتباط صفرًا ($r = 0$) دل ذلك على انعدام العلاقة بين المتغيرين.

قوة العلاقة:.

يمكن الحكم على قوة العلاقة من حيث درجة قربها أو بعدها عن $[\pm1]$، حيث إن قيمة معامل الارتباط تقع في المدى $[\,-1 \le r \le 1\,]$.

معامل الارتباط الخطي البسيط "لبيرسون" Karl Pearson

في حالة جمع بيانات عن متغيرين كميين (x , y)، 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 بين الدرجة التي حصل عليها الطالب وعدد ساعات الاستذكار، وهكذا الأمثلة على ذلك كثيرة.

إذا افترضنا أن لدينا عينة مكونة من n من أزواج المشاهدات $(x_1 , y_1), (x_2 , y_2),\dots (x_n , y_n)$ لمتغيرين X وY، فإن معامل الارتباط لبيرسون يعرف بأنه:

$$r = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}}{\sqrt{\left[\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}\right]\left[\sum_{i=1}^{n} Y_i^2 - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}\right]}}$$

(1-8)

فإذا نظرنا إلى المعادلة السابقة عرفنا مقدار المعاناة عند حساب مقدار معامل الارتباط، فضلاً عن حجم التحليل والحسابات إذا زاد حجم العينة، والآن نتعرف كيف يمكن تطبيق ذلك بمنتهى السهولة باستخدام البرنامج وذلك من خلال المثال القادم:

مثال (8-1):

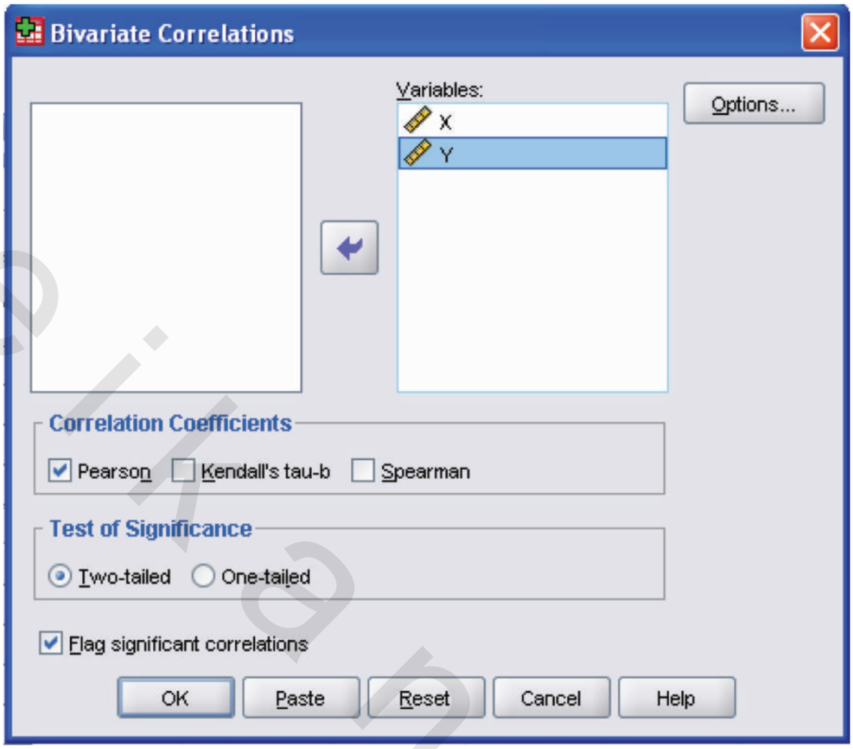
احسب معامل الارتباط المناسب بين درجة الطالب في الرياضيات ودرجته في الإحصاء باستخدام البيانات الآتية لعينة من خمسة طلاب:

جدول (8-1)

درجة الرياضة (x)	15	19	11	3	7
درجة الإحصاء (y)	18	20	16	14	12

الحـل:

من شريط قوائم نختار Analyze ثم نختار Correlate ثم نختار Bivariate فيظهر الشكل الآتي:



شكل (8-1)

ونرى أن الشكل السابق يتضمن مجموعة من الأجزاء:

- عمود المتغيرات المتاحة وهو العمود الأول من جهة اليسار، ومنه يتم نقل المتغيرات إلى العمود الذي بجواره لنبدأ حساب معامل الارتباط وهنا نقوم بنقل X، Y.
- معامـل الارتبـاط (*Correlation Coefficients*): وهو يتيح لنـا تحديد نوع معامل الارتباط الذي نريده ونلاحظ أنه متوافر لدينا 3 اختيارات:

✓ Pearson: وهو معامل ارتباط الظواهر الكمية سواء كانت مبوبة أو غير مبوبة.

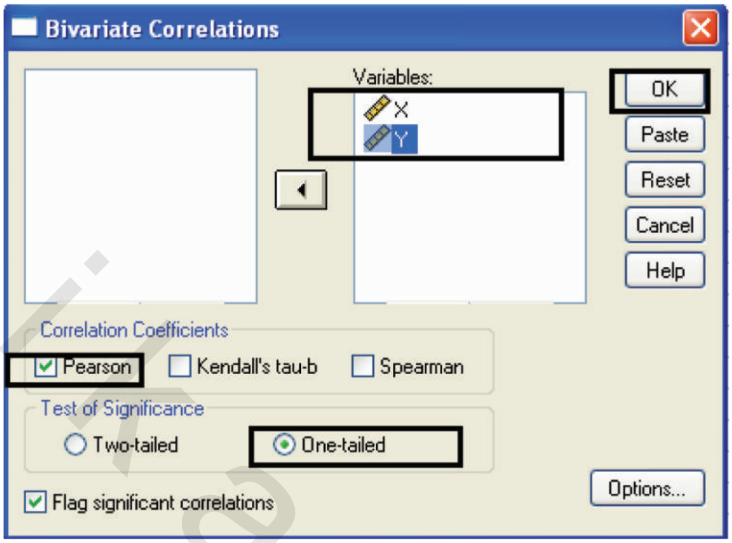
✓ Kendall's tau: وهـو معامـل ارتبـاط الظواهر الترتيبية، ويفضل للعينـات الصغيرة والبيانات غير المبوبة.

✓ Spearman: وهـو أيضًا معامل الارتباط بين الظواهر الكمية والترتيبية، ويفضل للعينات الصغيرة، ويستخدم في حالة البيانات غير المبوبة. وهنا نقوم بتحديد معامل بيرسون.

✓ Test of Siginificance: ولأننا نستخدم بيانات عينة فإننا نحتاج إلى التأكد من قيمة معامل الارتباط فنحتاج إلى اختبار هذه القيمة، وهنا نختار إما الاختبار من جهة أو جهتين.

Flag Significant Correlation: هنـا تستخـدم لتعليم الارتباط بعلامة نجمة (star) في حالة الارتباط المعنوي (الدال) أي الارتباط الحقيقي الذي يؤكده الاختبار.

وبعد الاختيار السابق يظهر الشكل (8-3) في الشكل الآتي:



شكل (8-2)

ثم ننقر على ok فيظهر الجدول الآتي:

جدول (8-2)

Correlations

		X	Y
X	Pearson Correlation	1	.900*
	Sig. (2-tailed)		.037
	N	5	5
Y	Pearson Correlation	.900*	1
	Sig. (2-tailed)	.037	
	N	5	5

*. Correlation is significant at the 0.05 level (2-tailed).

ونجـد أن معامـل الارتباط بين درجة الطالب في الإحصاء ودرجتـه في الرياضيات مقدارها 0.9، وهذا يعني أن الارتباط طردي وقوي جدا ولاختبار الفرض الصفري:

$$H_0: \rho = 0$$

ضد الفرض البديل:

$$H_A: \rho \neq 0$$

ونلاحظ أن الفرض البديل لا يساوي صفرًا، وهذا ما دل عليه اختيارنا السابق، حيث إن الاختبار ذو اتجاهين، فإذا افترضنا أننا نريد اختبار ذلك بمستوى معنوية مقدارها α :0.05 فإن قيمة: $\text{p-value} = 0.019 < 0.05 = \alpha$ دل ذلك على أن الارتباط بين درجة الرياضيات والإحصاء ارتباط يختلف عن الصفر بمستوى معنوي مقداره 0.05، أي إننا نقبل الفرض البديل القائل: إن الارتباط لا يساوي صفرًا.

معامل ارتباط الرتب لـ(سبيرمان) Spearman

إذا كانت الظاهرة محل الدراسة تحتوي على متغيرين وصفيين ترتيبيين، ومثال على ذلك قياس العلاقة بين تقديرات الطلبة في مادتين، أو العلاقة بين درجة تفضيل المستهلك لسلعة معينة، ومستوى الدخل، فإنه يمكن استخدام طريقة أخرى في حساب معامل ارتباط يعتمد على رتب مستويات المتغيرين كبديل للقيم الأصلية، ويطلق على هذا المعامل "معامل ارتباط سبيرمان" Spearman.

مثال (8-2):

فيما يلي بيانات درجة تفضيل ومستوى دخل لعشرة مستهلكين لسلعة ما كعينة، والمطلوب معرفة الارتباط بين مستوى الدخل ودرجة تفضيل المستهلكين للسلعة، وما مدلول ذلك بالنسبة لجميع المستهلكين:

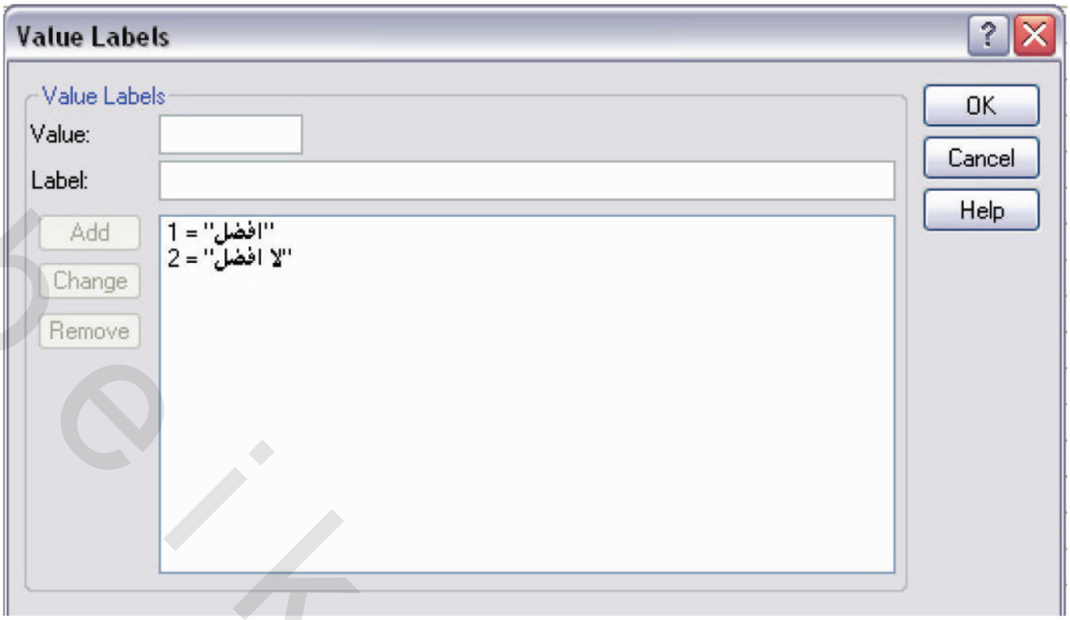
جدول (8-3)

مستوى التفضيل	أفضل	لا أفضل	أفضل	لا أفضل	لا أفضل	لا أفضل	أفضل	لا أفضل	أفضل	أفضل
مقدار الدخل الشهري	100	200	250	400	130	270	700	500	330	200

الحل:

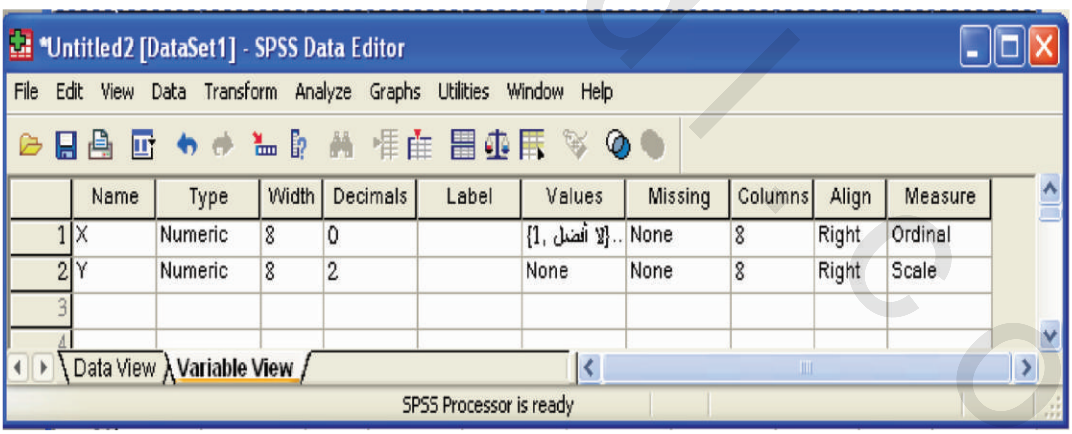
خلال البرنامج نقوم بإدخال البيانات على فرض أن "أفضل" نرمز لها بالرمز 1 و"لا أفضل" برمز 2.

ونقوم بتعريف المتغير الترتيبي x كما يلي في محرر المتغيرات نقوم بتعريف قيم المتغير في عمود values كما يلي:



شكل (8-3)

ثم في عمود measure نعرف المتغير X بأنه متغير ترتيبي ordinal، ونعرف المتغير Y بأنه متغير كمي عادي فيظهر المحرر كالآتي:



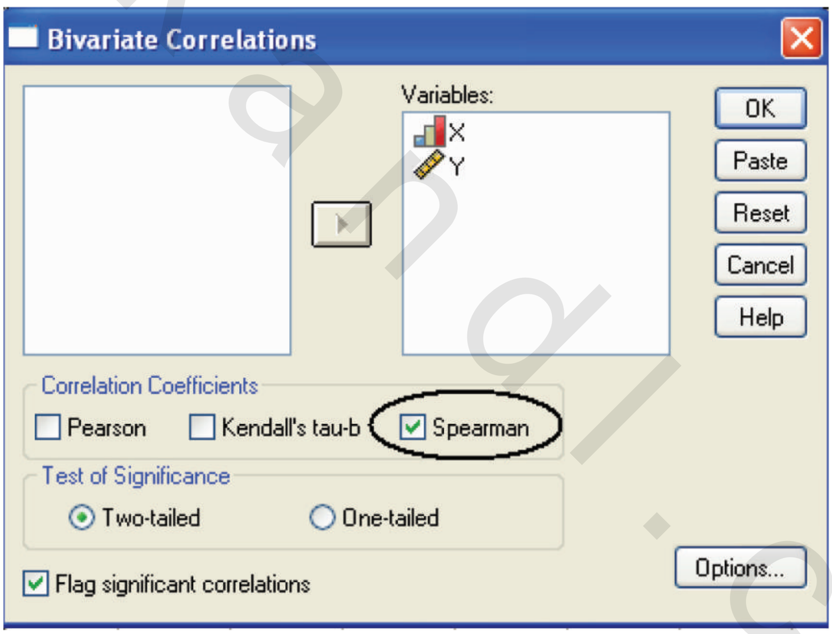
شكل (8-4)

ثم شاشة محرر البيانات نقوم بإدخال البيانات السابقة فتظهر في الشكل الآتي:

	X	Y
1	1	100.00
2	2	200.00
3	1	250.00
4	2	400.00
5	2	130.00
6	2	270.00
7	1	700.00
8	2	500.00
9	1	330.00
10	1	200.00

شكل (5-8)

من شريط قوائم نختار Analyze ثم نختار Correlate ثم نختار Bivariate فيظهر الشكل الآتي:



شكل (6-8)

ونلاحظ أن المتغير X معرف في عمود المتغير المستخدم في التحليل (العمود الأول من اليمين)، وأننا اخترنا Spearman كمعامل للارتباط، والاختيارات السابقة نفسها في مقياس Pearson ثم ننقر على ok فيظهر الجدول الآتي:

جدول (4-8)

Correlations

Control Variables			x	z
y	x	Correlation	1.000	.693
		Significance (2-tailed)	.	.057
		df	0	6
	z	Correlation	.693	1.000
		Significance (2-tailed)	.057	.
		df	6	0

ونجد أن معامل الارتباط بين درجة تفضيل المستهلكين لسلعة ومستوى دخولهم مقداره 0.070، وهذا يعني أن الارتباط طردي وضعيف ولاختبار الفرض القائل إن:

$$H_0: \rho = 0$$

ضد الفرض البديل القائل إن:

$$H_A: \rho \neq 0.$$

ونلاحظ أن الفرض البديل معامل الارتباط لا يساوي صفرًا، وهذا ما دل عليه اختيارنا السابق أن الاختبار ذو اتجاهين، فإذا افترضنا أننا نريد اختبار ذلك بمستوى معنوية مقدارها $\alpha = 0.05$، فإن قيمة P-eulav $= 0.05 = \alpha$ 0.424 < دل ذلك على أن الارتباط بين تفضيل المستهلكين لسلعة ومستوى دخولهم ارتباط يساوي الصفر بمستوى معنوي مقداره 0.05، أي إننا نقبل فرض العدم القائل: إنَّ الارتباط يساوي صفرًا في كامل بيانات المجتمع.

(3 - 8) الارتباط الجزئي Partial Correlation

يقيس معامل الارتباط الجزئي قوة العلاقة بين متغيرين بثبوت متغير ثالث أو أكثر. مثلا قد نحصل على قيمة عالية لمعامل الارتباط البسيط للعلاقة بين أسعار اللحوم البيضاء واللحوم الحمراء، فقد لا توجد علاقة فعلية بين المتغيرين، ولكن كلا المتغيرين يتأثر بمعامل ثالث هو المستوى العام للأسعار، فإذا استبعدنا المستوى العام للأسعار (أو تثبيته) عند قياس العلاقة بين أسعار اللحوم البيضاء والحمراء، فسيتم الحصول على قيمة أقل لمعامل الارتباط، وهذا يعرف بالارتباط الجزئي. علما بأنه يمكن استبعاد أي عدد من المتغيرات عند قياس العلاقة بين ظاهرتين.

والآتي مثال يوضح كيفية تطبيق ذلك من خلال البرنامج:

مثـال (8-3):

البيانــات الآتية تعطي الدخل الشهري لمجموعـة أسر (X) والإنفاق الشهري على الطعام (Y) وحجم مدخراتها الشهرية (Z) وذلك في عينة من تسع أسر:

جدول (8-5)

الأسرة	1	2	3	4	5	6	7	8	9
x	40	65	90	60	75	80	120	45	65
Y	22	28	40	20	32	32	50	20	35
Z	6	10	8	25	12	7	10	5	7

المطلوب:

1 - حساب معامل الارتباط الجزئي لـ x و y بثوت z.

2 - حساب معامل الارتباط الجزئي لـ x و z بثوت y.

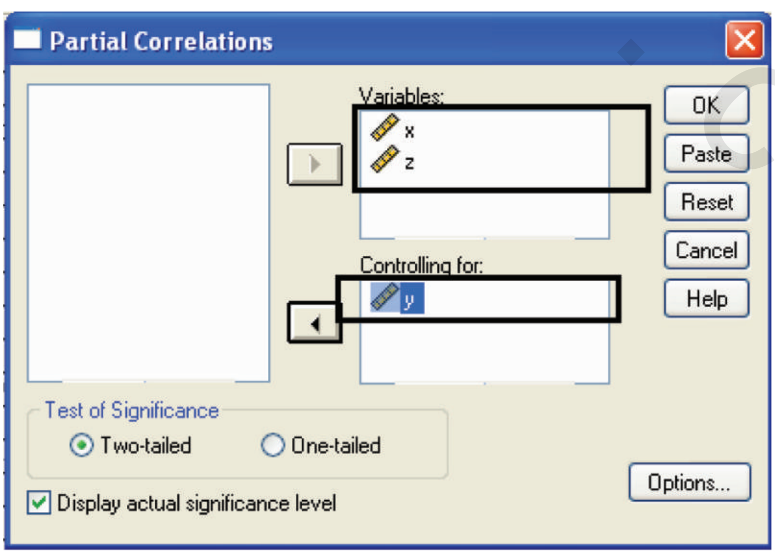
3 - حساب معامل الارتباط الجزئي لـ z و y بثوت x.

4 - اختبار معنوية (دلالة) الارتباط بمستوى معنوية قدره 0,05.

الحل:

أولا: حساب معامل الارتباط الجزئي لـ x و y بثوت z.

مـن شريــط قوائم نختار Analyze ثم نختــار Correlate ثم نختار Partial فيظهــر مربع حوار معامل الارتباط الجزئي الآتي في الشكل الآتي:



شكل (8-7)

في خانة variables يتم إدخال المتغيرات التي يراد حساب معامل الارتباط الجزئي لها (x، y)، وفي خانة Controlling for يتم إدخال المتغير (المتغيرات) الذي يراد استبعاد أثره (z).

وعند النقر على زر ok نحصل على النتيجة الآتية:

جدول (8-6)

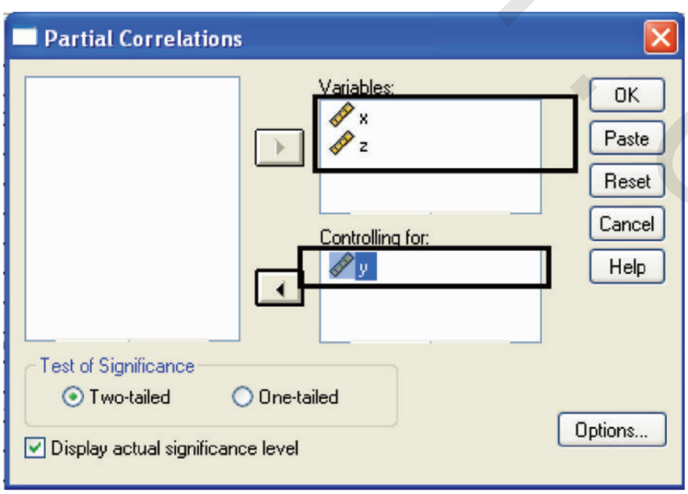
Correlations

Control Variables			z	y
x	z	Correlation	1.000	-.710
		Significance (2-tailed)	.	.048
		df	0	6
	y	Correlation	-.710	1.000
		Significance (2-tailed)	.048	.
		df	6	0

نلاحظ أن قيمة معامل الارتباط من الشكل السابق تساوي 0,965، أي إنه توجد علاقة طردية وقوية بين الدخل والإنفاق، كما أن قيمة p-value=0 وهي أقل من قيمة مستوى المعنوية مقسومًا على 2=0,025، حيث إننا نختبر فرض عدم ذا اتجاهين، ولذلك فإننا نقبل الفرض البديل القائل إن الارتباط بين المتغيرات في المجتمع لا يساوي صفرًا.

ثانياً: حساب معامل الارتباط الجزئي لـ x و z بثوت y.

من شريط قوائم نختار Analyze ثم نختار Correlate ثم نختار Partial فيظهر مربع حوار معامل الارتباط الجزئي الآتي في الشكل الآتي:



شكل (8-8)

يتم إدخال المتغيرات (x،z) في خانة variables، وفي خانة Controlling for يتم إدخال المتغير (y).

عند النقر على زر ok نحصل على النتيجة الآتية:

جدول (8-7)

Correlations

Control Variables			x	z
y	x	Correlation	1.000	.693
		Significance (2-tailed)	.	.057
		df	0	6
	z	Correlation	.693	1.000
		Significance (2-tailed)	.057	.
		df	6	0

أن قيمة معامل الارتباط من الشكل السابق تساوي 0,693 أي إنه توجد علاقة طردية بين الدخل الادخار، كما أن قيمة p-value=0,057 وهي أكبر من قيمة مستوى المعنوية مقسومًا على 2، حيث إننا نختبر فرض عدم ذا اتجاهين، ولذلك فإننا نقبل الفرض العدم القائل إن الارتباط بين المتغيرات في المجتمع يساوي صفرًا.

ثالثاً: حساب معامل الارتباط الجزئي لـ z و y بثوت x.

كما فعلنا في أولا وثانيًا من خلال إدخال وتعريف المتغيرات نفعل في تلك الخطوة وتظهر نتيجة معامل الارتباط كالآتي:

جدول (8-8)

Correlations

Control Variables			z	y
x	z	Correlation	1.000	-.710
		Significance (2-tailed)	.	.048
		df	0	6
	y	Correlation	-.710	1.000
		Significance (2-tailed)	.048	.
		df	6	0

فنجد أن قيمة معامل الارتباط بين الاستهلاك والادخار تساوي 0,710 -، وهذا يعني أنه توجد علاقة بين الاستهلاك والادخار عكسية وقوية، ونجد أيضا أن قيمة p-value=0,048 وهي أكبر من قيمة مستوى المعنوية مقسومًا على 2=0,025، حيث إننا نختبر فرض عدم ذا اتجاهين، ولذلك فإننا نقبل الفرض العدم القائل إن الارتباط بين المتغيرات في المجتمع يساوي صفرًا في ضوء بيانات العينة.