الفصل الخامس

العرض والتحليل الإحصائي لمتغيرين

View and Statistical Analysis of Bivariate Random Variable

(5-1) مقدمة :

بعد عرض وصف متغير واحد في الفصل السابق يكون من الضروري هنا أن نوضح كيفية وصف متغيرين أو أكثر، حيث نجد أن العلاقة بين متغيرين تكون من أهم الأشياء التي لا غنى عنها في التحليل الإحصائي. ولذلك كان من المهم التعرف على كيفية وصف متغيرين وتوضيح العلاقة بينهما باستخدام أشكال الانتشار والجداول المزدوجة، بالإضافة إلى المقاييس الإحصائية المناسبة لهم.

توجد ثلاث طرق للعرض في حالة متغيرين:

- العرض الجدولي.
- العرض البياني.
- العرض باستخدام أحد المقاييس الإحصائية.

(5-2) الجداول المزدوجة:

الجدول المزدوج هو جدول ثنائي يلخص العلاقة بين متغيرين جدوليًا فيبرز الخصائص المشتركة لهما.

مثال (5-1):

الجدول الآتي يمثل توزيع الطلاب والطالبات المقبولين في كلية الآداب لعام 2007 / 2008م.

جدول (5-1)

التخصص	اللغة العربية	اللغة الإنجليزية	التاريخ	الجغرافيا	الأعلام	الاجتماع	المجموع
عدد الطلاب	74	115	254	109	201	138	891

عدد الطالبات	129	134	155	57	-	485	950
المجموع	203	249	409	155	201	523	1851

يمكن تحويل التكرارات الموجودة في الجدول إلى نسب (لكل قيم الجدول) كالآتي:

- ✓ إذا تمت قسمة جميع أرقام الجدول على مجموع التكرارات، وهو 1851 نحصل على نسبة كل خلية للمجموع الكلي.
- ✓ إذا تمت قسمة تكرارات أي صف على مجموع الصف فسنحصل على نسبة الخلية للصف.
- ✓ إذا تمت قسمة التكرار في كل خلية على مجموع العمود فسنحصل على نسبة الخلية لمجموع العمود.

مثال (5-2):

فيما يلي قائمة ببعض البيانات التي استخرجت من ملفات بعض العاملين في إحدى المصالح الحكومية، الذين يمثلون عينة عشوائية تم اختيارها من بين العاملين في تلك المصلحة الحكومية، التي تمثل مدة الخدمة للعامل والمرتب الشهري.

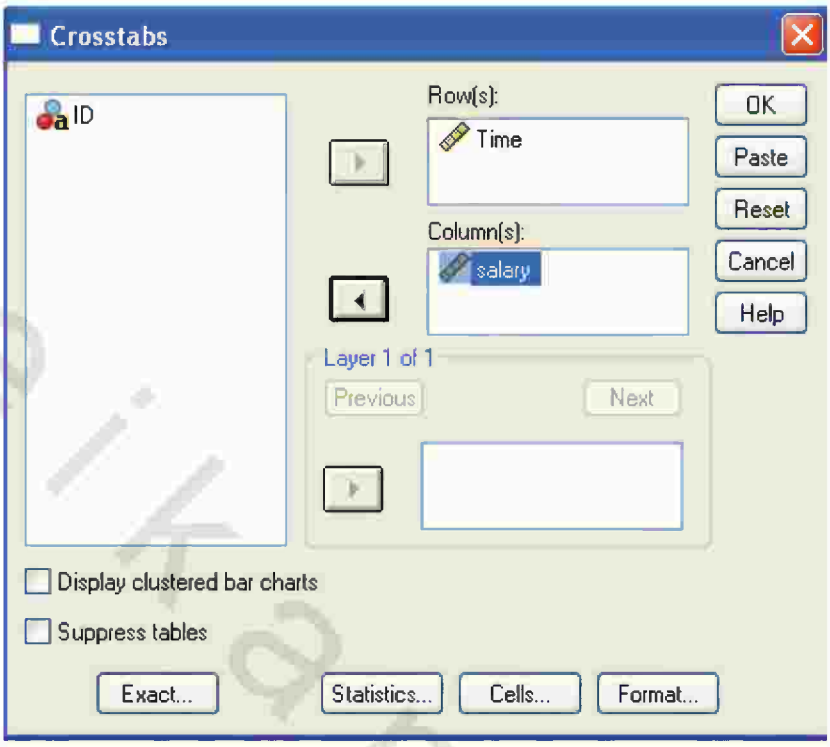
جدول (5-2)

رقم الملف	مدة الخدمة	المرتب الشهري
15/1	0.4	200
2/33	2	240
2/23	1	200
85/1	5	390
5/45	18	520
52/12	5	400
2/125	11	490
5/12	18	490
5/123	9	390
55/12	12	500
12/125	4	400
55/123	0.8	200
89/44	20	390
55/23	19	200

لإنشاء الجدول المزدوج باستخدام الحزمة SPSS لبيانات المثال نمر بالخطوات الآتية:

- ⊙ من القائمة Analyze نختار Descriptive Statistics وننقر الأمر Cross tabs.

فيظهر الصندوق انظر الشكل(5-1):

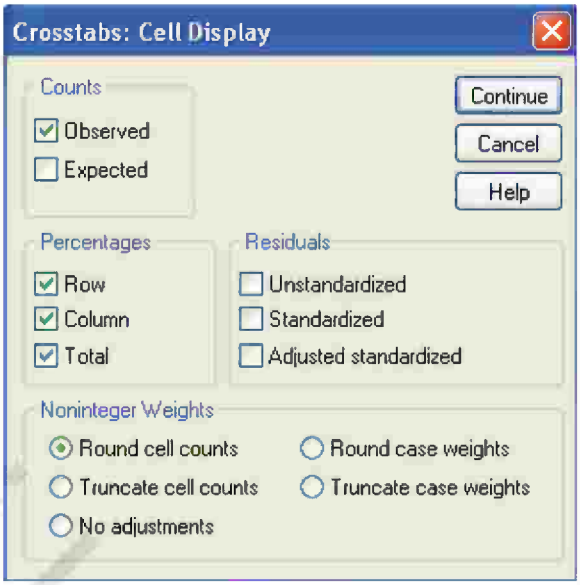


شكل (5-1)

حيث إن:

- Row(s): يتضمن المتغير الذي نرغب في جعل فئاته صفوفًا في الجدول.
- Column(s): يتضمن المتغير الذي نرغب في جعل فئاته أعمدة في الجدول.
- Display clustered bar charts: لتمثيل بيانات الجدول بالأعمدة البيانية المزدوجة.
- Suppress tables: لإخفاء الجداول المزدوجة عند التأشير.

⊙ انقر على Cells فنحصل على الصندوق انظر الشكل (5-2):



شكل (5-2)

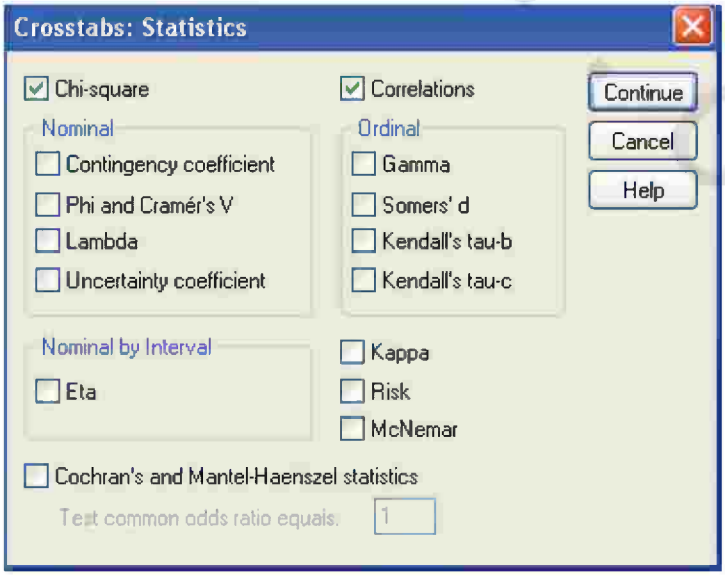
والذي نلاحظ احتواءه على العناصر الآتية:

- Observed: التكرار المشاهد للخلية.
- Expected: التكرار المتوقع للخلية (يستخدم لحساب اختبار الاستقلال).
- Residuals: للحصول على البواقي.

ننقر على continue للعودة للصندوق الحواري الأصلي.

ثم ننقر على الأمر Statistics يظهر الصندوق Cross tabs: Statistics.

انظر الشكل(5-3):



شكل (5-3)

والذي نلاحظ احتواءه على العناصر الآتية:

- Chi-square: لحساب استقلالية الصفوف عن الأعمدة.
- Correlation: يستخدم لحساب معاملي الارتباط Spearman وPearson يحسب Spearman عندما تكون قيم المتغيرات الداخلة عددية Numeric، وباعتبار أن كلا من الصفوف والأعمدة عبارة عن قيم ترتيبية.
- ويحسب Pearson باعتبار أن كلا من الصفوف والأعمدة متغيرات كمية ويعبر عنها بالخيار Scale.
- Nominal: تتيح هذه القائمة حساب أربعة معاملات للارتباط في حالة كون كل من الصفوف والأعمدة عوامل غير كمية، بالإضافة إلى عدم إمكانية ترتيب البيانات مثل متغير الجنس (ذكور/ إناث).
- Ordinal: تتيح هذه القائمة حساب أربعة معاملات للارتباط في حالة كون كل من الصفوف والأعمدة متغيرات ترتيبية.
- Nominal by Interval: لحساب إحصائية Eta للاقتران بين متغيرين أحدهما المعتمد يقاس ضمن فترة interval scale مثل متغير الدخل والآخر مستقل له عدد محدود من الفئات مثل الجنس.

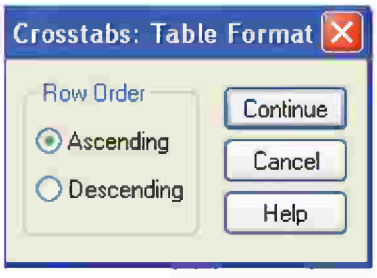
نظرا لكون الصفوف والأعمدة في هذا المثال متغيرات كمية Numeric فقد قمنا بحساب إحصائية Chi-square لاختبار الاستقلالية مع حساب قيمة معامل الارتباط بين المتغيرين.

ننقر على continue للعودة للصندوق الأصلي.

- ثم ننقر على Ok للتنفيذ.

ملاحظة:

- يمكن ترتيب صفوف الجدول تصاعديًا أو تنازليًا بنقر الزر Format في صندوق حوار Cross tabs كما في الشكل الآتي:



شكل (4-5)

بعد تحديد نوعية الترتيب، هنا تم اختيار الترتيب التصاعدي، نقوم بالنقر على Continuous لنرجع إلى القائمة الظاهرة في الشكل (5-1) وننقر على OK.

النتائج التي تم الحصول عليها في شاشة النتائج:

جدول (5-3)

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Time *salary	14	100.0%	0	.0%	14	100.0%

الجدول السابق به البيانات المتاحة والمفقودة.

جدول (5-4)

Time * salary Crosstabulation

			salary							Total
			200.00	240.00	390.00	400.00	490.00	500.00	520.00	
Time	.40	Count	1	0	0	0	0	0	0	1
		% within Time	100.0%	.0%	.0%	.0%	.0%	.0%	.0%	100.0%
		% within salary	25.0%	.0%	.0%	.0%	.0%	.0%	.0%	7.1%
		% of Total	7.1%	.0%	.0%	.0%	.0%	.0%	.0%	7.1%
	.80	Count	1	0	0	0	0	0	0	1
		% within Time	100.0%	.0%	.0%	.0%	.0%	.0%	.0%	100.0%
		% within salary	25.0%	.0%	.0%	.0%	.0%	.0%	.0%	7.1%
		% of Total	7.1%	.0%	.0%	.0%	.0%	.0%	.0%	7.1%
	1.00	Count	1	0	0	0	0	0	0	1
		% within Time	100.0%	.0%	.0%	.0%	.0%	.0%	.0%	100.0%
		% within salary	25.0%	.0%	.0%	.0%	.0%	.0%	.0%	7.1%
		% of Total	7.1%	.0%	.0%	.0%	.0%	.0%	.0%	7.1%
	2.00	Count	0	1	0	0	0	0	0	1
		% within Time	.0%	100.0%	.0%	.0%	.0%	.0%	.0%	100.0%
		% within salary	.0%	100.0%	.0%	.0%	.0%	.0%	.0%	7.1%
		% of Total	.0%	7.1%	.0%	.0%	.0%	.0%	.0%	7.1%
	4.00	Count	0	0	0	1	0	0	0	1
		% within Time	.0%	.0%	.0%	100.0%	.0%	.0%	.0%	100.0%
		% within salary	.0%	.0%	.0%	50.0%	.0%	.0%	.0%	7.1%
		% of Total	.0%	.0%	.0%	7.1%	.0%	.0%	.0%	7.1%
	5.00	Count	0	0	1	0	0	0	0	1
		% within Time	.0%	.0%	100.0%	.0%	.0%	.0%	.0%	100.0%
		% within salary	.0%	.0%	33.3%	.0%	.0%	.0%	.0%	7.1%
		% of Total	.0%	.0%	7.1%	.0%	.0%	.0%	.0%	7.1%
	6.00	Count	0	0	0	1	0	0	0	1
		% within Time	.0%	.0%	.0%	100.0%	.0%	.0%	.0%	100.0%
		% within salary	.0%	.0%	.0%	50.0%	.0%	.0%	.0%	7.1%
		% of Total	.0%	.0%	.0%	7.1%	.0%	.0%	.0%	7.1%

	9.00	Count	0	0	1	0	0	0	0	1
		% within Time	.0%	.0%	100.0%	.0%	.0%	.0%	.0%	100.0%
		% within salary	.0%	.0%	33.3%	.0%	.0%	.0%	.0%	7.1%
		% of Total	.0%	.0%	7.1%	.0%	.0%	.0%	.0%	7.1%
	11.00	Count	0	0	0	0	1	0	0	1
		% within Time	.0%	.0%	.0%	.0%	100.0%	.0%	.0%	100.0%
		% within salary	.0%	.0%	.0%	.0%	50.0%	.0%	.0%	7.1%
		% of Total	.0%	.0%	.0%	.0%	7.1%	.0%	.0%	7.1%
	12.00	Count	0	0	0	0	0	1	0	1
		% within Time	.0%	.0%	.0%	.0%	.0%	100.0%	.0%	100.0%
		% within salary	.0%	.0%	.0%	.0%	.0%	100.0%	.0%	7.1%
		% of Total	.0%	.0%	.0%	.0%	.0%	7.1%	.0%	7.1%
	18.00	Count	0	0	0	0	1	0	1	2
		% within Time	.0%	.0%	.0%	.0%	50.0%	.0%	50.0%	100.0%
		% within salary	.0%	.0%	.0%	.0%	50.0%	.0%	100.0%	14.3%
		% of Total	.0%	.0%	.0%	.0%	7.1%	.0%	7.1%	14.3%
	19.00	Count	1	0	0	0	0	0	0	1
		% within Time	100.0%	.0%	.0%	.0%	.0%	.0%	.0%	100.0%
		% within salary	25.0%	.0%	.0%	.0%	.0%	.0%	.0%	7.1%
		% of Total	7.1%	.0%	.0%	.0%	.0%	.0%	.0%	7.1%
	20.00	Count	0	0	1	0	0	0	0	1
		% within Time	.0%	.0%	100.0%	.0%	.0%	.0%	.0%	100.0%
		% within salary	.0%	.0%	33.3%	.0%	.0%	.0%	.0%	7.1%
		% of Total	.0%	.0%	7.1%	.0%	.0%	.0%	.0%	7.1%
Total		Count	4	1	3	2	2	1	1	14
		% within Time	28.6%	7.1%	21.4%	14.3%	14.3%	7.1%	7.1%	100.0%
		% within salary	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	28.6%	7.1%	21.4%	14.3%	14.3%	7.1%	7.1%	100.0%

الجدول السابق هو الجدول التكراري المزدوج للمتغيرين الأعمدة لـ Recover والصفوف لـ .Treat

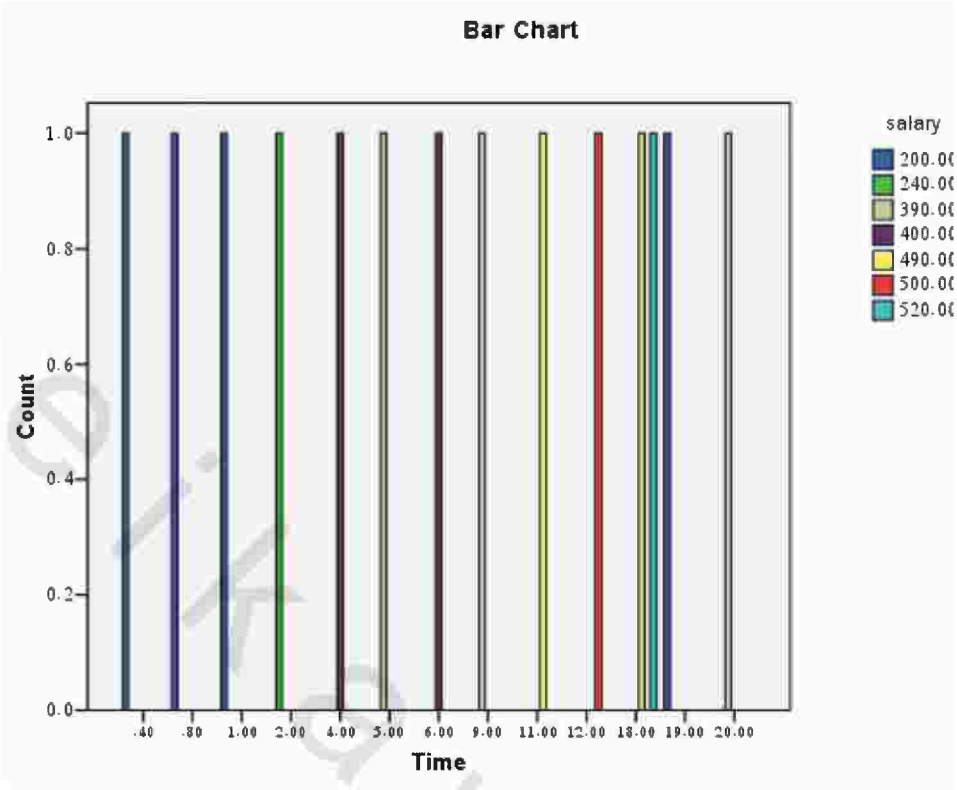
- Count: هو التكرار المشاهد.
- % within treat: النسبة للصف.
- % Within recover: النسبة للعمود. ثم للكلي % total of.

جدول (5-5)

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	73.500[a]	72	.429
Likelihood Ratio	47.894	72	.987
Linear-by-Linear Association	3.758	1	.053
N of Valid Cases	14		

a. 91 cells (100.0%) have expected count less than 5. The minimum expected count is .07.



شكل (5-5)

✓ تم الحصول على رسمة الأعمدة البيانية المزدوجة للمتغيرين معا، وهي أعمدة متلاصقة كل عمود مخصص لبيانات متغير معين وله لون مختلف، حيث إن تكرار مدة الخدمة واحد.

(5-3) شكل الانتشار:

شكل الانتشار ما هو إلا تمثيل للبيانات المزدوجة بنقاط على محورين أحدهما المحور السيني ويمثل عليه المتغير المستقل Independent والآخر المحور الصادي ويمثل عليه المتغير التابع Dependent. وبالنظر إلى شكل الانتشار نستطيع تحديد نوع العلاقة ودرجتها تقريبًا طبقًا للضوابط الآتية:

- إذا تجمعت النقاط حول خط يصنع زاوية حادة أو منفرجة مع المحور الأفقي تكون العلاقة خطية، وإذا تجمعت البيانات حول أي شكل آخر غير الخط تكون العلاقة غير خطية.
- إذا كانت العلاقة خطية واقتربت النقاط حول خط وهمي يمكن تخيله وتحديد ما إذا كانت العلاقة قوية والعكس صحيح.
- إذا كانت العلاقة خطية (أم غير خطية) ووقعت النقاط في شكل الانتشار جميعها فوق الخط كانت العلاقة تامة، وهذه أشد صورة من صور الارتباط.

مثال (5-3):

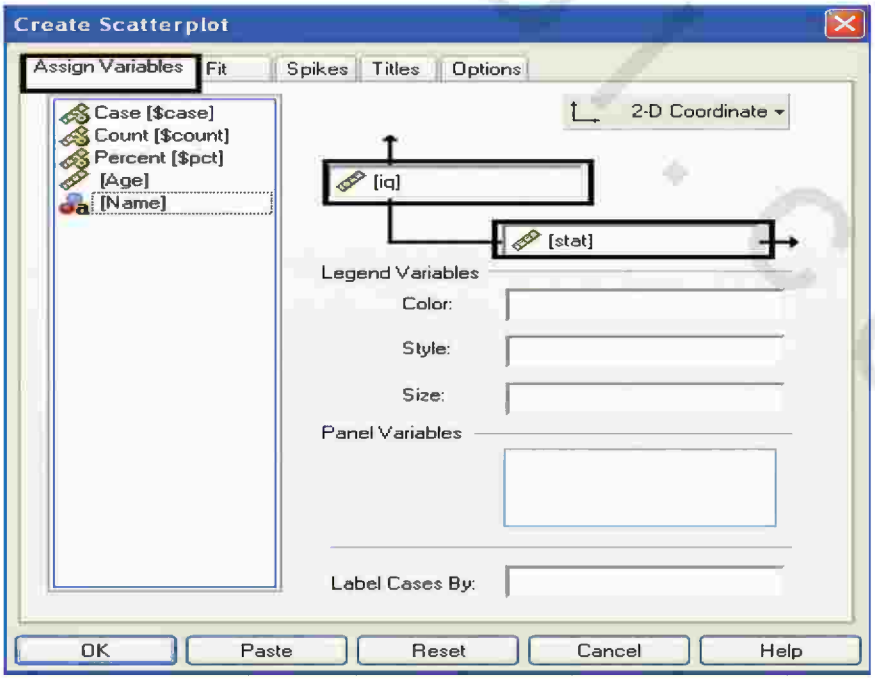
نفرض أن لدينا بيانات خاصة بنتائج 25 طالبًا في اختبار للذكاء، ودرجة مقرر الإحصاء مقررين يرمز لهما بالرمزين IQ ، Stat

جدول (5-6)

IQ	STAT	العمر	الاسم
50 ,00	40 ,00	21 ,00	Ahmad
70 ,00	59 ,00	30 ,00	Hanna
50 ,00	55 ,00	21 ,00	Heba
90 ,00	85 ,00	22 ,00	Kamala
80 ,00	77 ,00	27 ,00	Mervat
83 ,00	80 ,00	23 ,00	Mohamed
50 ,00	35 ,00	22 ,00	Samer
50 ,00	55 ,00	25 ,00	Satar
80 ,00	50 ,00	21 ,00	Shimaa

باستخدام البرنامج يمكن رسم شكل الانتشار كالآتي:

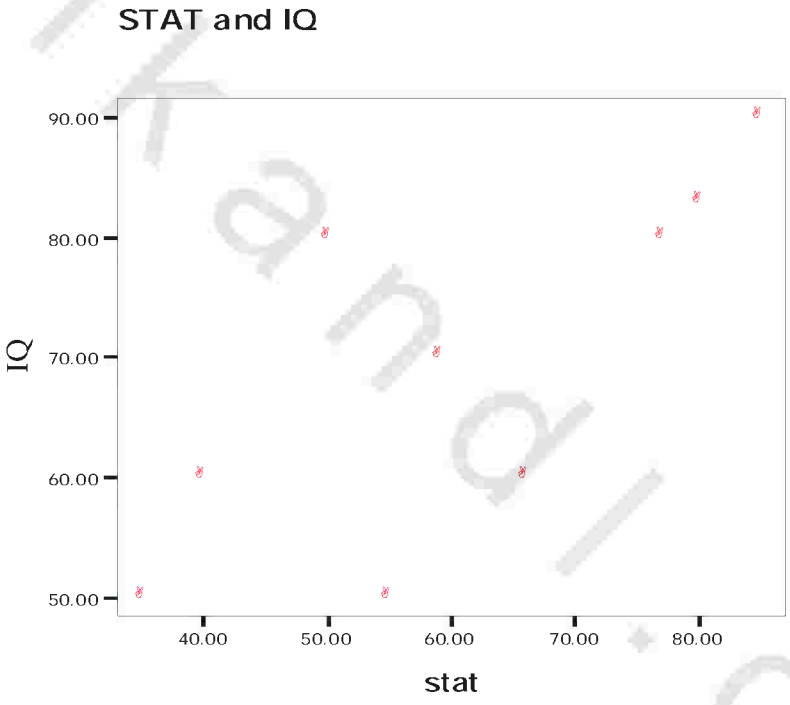
✓ من القائمة Graphs انقر الأمر interactive ثم انقر Scatter plot يفتح الصندوق Scatter plot انظر الشكل (5-6).



شكل (5-6)

انقـر علـى الأمر Assign Variables وذلك لإدخال المتغيرات التـي نرغب في إيجاد شكل الانتشار لها، وليكن المتغير iq المستطيل الخاص بالمحور الرأسي Y Axis، والمتغير stat إلى مستطيل المحور الأفقي X Axis كما هو واضح في الشكل السابق.

- انقـر علـى الأمر Title لكتابـة عنوان الشكل المرسوم ثـم انقر continue للعـودة للصندوق الأصلي وليكن نريده أن يكون STAT and IQ.
- انقر على الأمر options لتحدد موقفك عن كيفية معالجة القيم المفقودة.
- انقر على الأمر continue لتعود للصندوق الحواري الأصلي.
- انقر على Ok لتنفيذ الأمر.



شكل (5-7)

يتضـح من شكل الانتشـار أن البيانات متجمعة حول خط يصنع زاوية حـادة مع المحور الأفقي، أي أن العلاقة بين المتغيرين يمكن أن تكون خطية طردية، وكذلك العلاقة غير تامة لأن النقاط جميعها لم تتجمع فوق خط معين، لا يمكن معرفة قيمة معامل الارتباط من الشكل، ولكن يمكن تصور اقترابه من رقم معين. وسوف نتعرض في الفصول القادمة لكيفية حساب معامل الارتباط للبيانات المختلفة.