مقياس برمجيات تحليل البيانات

المستوى : السنة الثانية ماستر

تخصص : علم اجتماع الجريمة والانحراف

المحاضرة السادسة

عنوان المحاضرة : التحقق من ملائمة البيانات

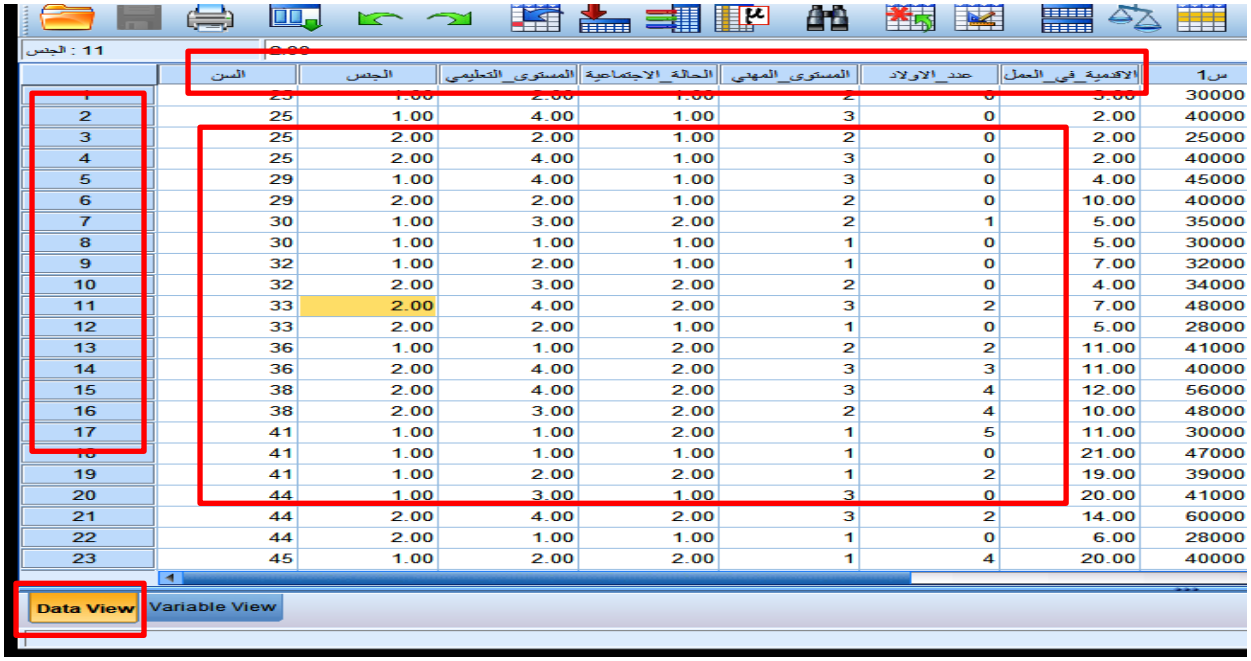
أهداف المحاضرة

- التعرف على طريقة تفريغ البيانات في نافذة Data View
- معرفة كيفية التحقق من ملائمة البيانات واكتمالها
- معرفة طريقة الكشف عن البيانات المكررة
- معرفة طريقة اكتشاف البيانات غير الاعتيادية
-

أولا : تفريغ البيانات في نافذة Data View

بعد الانتهاء من تحديد المتغيرات وضبطها في نافذة Variables View ، نقوم بالانتقال الى النافذة الملاصقة لها والتي تحمل اسم Data View ، بالضغط عليها مباشرة او عن طريق لوحة المفاتيح عن طريق ctrl+T ، حيث تخصص هذه النافذة لتفريغ البيانات التي تم جمعها اضافة الى القيام بالعمليات الاحصائية والتحليلية.

يوضح الشكل التالي نافذة عرض البيانات .



	السن	الجنس	المستوى_التعليمي	الحالة_الاجتماعية	المستوى_المهني	عدد_الاولاد	الاقدمية_في_العمل	س1
2	25	1.00	4.00	1.00	3	0	2.00	40000
3	25	2.00	2.00	1.00	2	0	2.00	25000
4	25	2.00	4.00	1.00	3	0	2.00	40000
5	29	1.00	4.00	1.00	3	0	4.00	45000
6	29	2.00	2.00	1.00	2	0	10.00	40000
7	30	1.00	3.00	2.00	2	1	5.00	35000
8	30	1.00	1.00	1.00	1	0	5.00	30000
9	32	1.00	2.00	1.00	1	0	7.00	32000
10	32	2.00	3.00	2.00	2	0	4.00	34000
11	33	2.00	4.00	2.00	3	2	7.00	48000
12	33	2.00	2.00	1.00	1	0	5.00	28000
13	36	1.00	1.00	2.00	2	2	11.00	41000
14	36	2.00	4.00	2.00	3	3	11.00	40000
15	38	2.00	4.00	2.00	3	4	12.00	56000
16	38	2.00	3.00	2.00	2	4	10.00	48000
17	41	1.00	1.00	2.00	1	5	11.00	30000
18	41	1.00	1.00	1.00	1	0	21.00	47000
19	41	1.00	2.00	2.00	1	2	19.00	39000
20	44	1.00	3.00	1.00	3	0	20.00	41000
21	44	2.00	4.00	2.00	3	2	14.00	60000
22	44	2.00	1.00	1.00	1	0	6.00	28000
23	45	1.00	2.00	2.00	1	4	20.00	40000

Data View | Variable View

الشكل رقم 01 : نافذة عرض البيانات Data View

تمثل المتغيرات ضمن الاطار الافقي (السن ، الجنس ، المستوى التعليمي ...) المتغيرات التي قمنا بضبطها في نافذة Variables View

تمثل الارقام (من 01 الى 23) الموضحة في الاطار العمودي عدد الحالات التي قمنا بتفريغها او ما يمكن ان نعبر عنه بعدد الاستمارات او عدد المبحوثين.

يمثل الاطار الاكبر في الوسط البيانات التي قمنا بتفريغها حسب المتغيرات المحددة مسبقا.

ثانيا : التحقق من ملائمة البيانات

بعد الانتهاء من تفريغ البيانات وقبل البدء في العمليات الاحصائية والتحليلية ، لابد من القيام أولا من التأكد من صحة البيانات ومدى صدقها ، حيث يتم هذا الاثبات عبر ثلاث خطوات ، تتمثل الاولى في اكتمال البيانات المدخلة (لا توجد قيم مفقودة لانها تؤثر على عملية التحليل) ، عدم وجود حالات مكررة عدم وجود قيم غير اعتيادية ضمن البيانات .

1.اثبات عدم وجود قيم مفقودة (تم ادخال كل البيانات)

نقوم بعرض المثال التالي الذي يمثل مجموعة من التلاميذ تم اخذ قياساتهم (الطول والوزن) ، حيث كانت النتائج كالتالي :

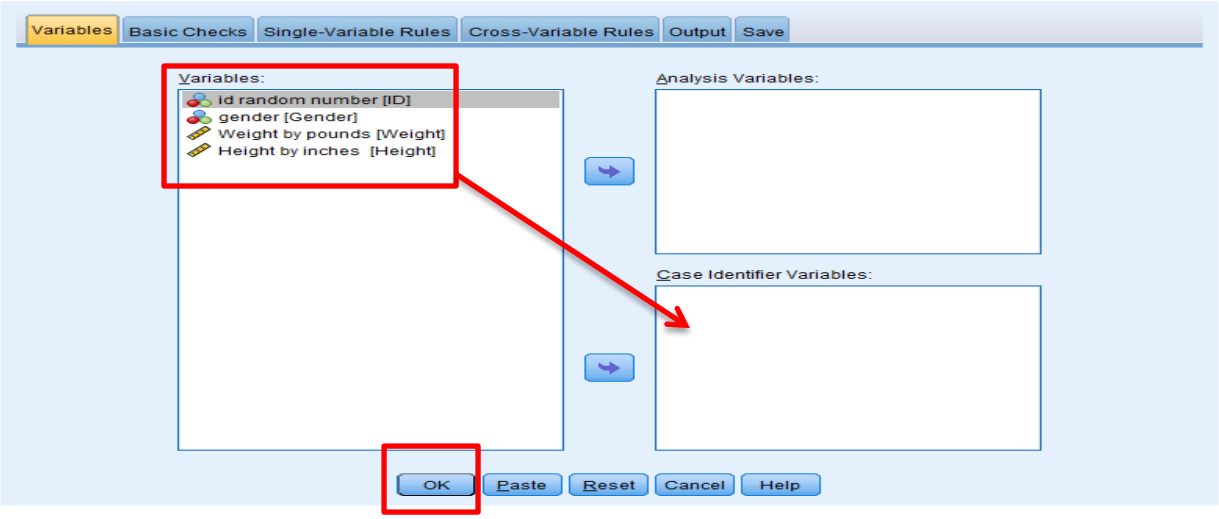
1 : Height 66.00

	ID	Gender	Weight	Height	var	var	var	var
1	1.00	Male	140.00	66.00				
2	2.00	Male	140.00	65.00				
3	3.00	Female	156.00	60.00				
4	4.00	Male	144.00	63.00				
5	5.00	Female	122.00	65.00				
6	6.00	Female	145.00	66.00				
7	7.00	Male	156.00	55.00				
8	8.00	Male	140.00	58.00				
9	9.00	Female	144.00	68.00				
10	10.00	Male	120.00	70.00				
11	11.00	Female	145.00	62.00				
12	12.00	Male	140.00	60.00				
13	13.00	.	4500.00	62.00				
14	14.00	Male	140.00	.				
15	15.00	Female	150.00	65.00				
16	16.00	Male	156.00	68.00				
17	17.00	Female	140.00	70.00				
18	18.00	Male	144.00	66.00				
19								
20								
21								
22								
23								

Data View | Variable View

الشكل رقم 02 : نتائج المثال

تتمثل أول خطوة في الضغط على نافذة Data، ثم اختيار Validation، ثم اختيار Validate Data، ثم نجد الجدول التالي



ثم نقوم بنقل المتغيرات **Variables** الى خانة **Case Identifier Variables** ثم نضغط على **ok**

لنتحصل في الاخير على جدول التالي :

Identifier Checks

Incomplete Identifiers

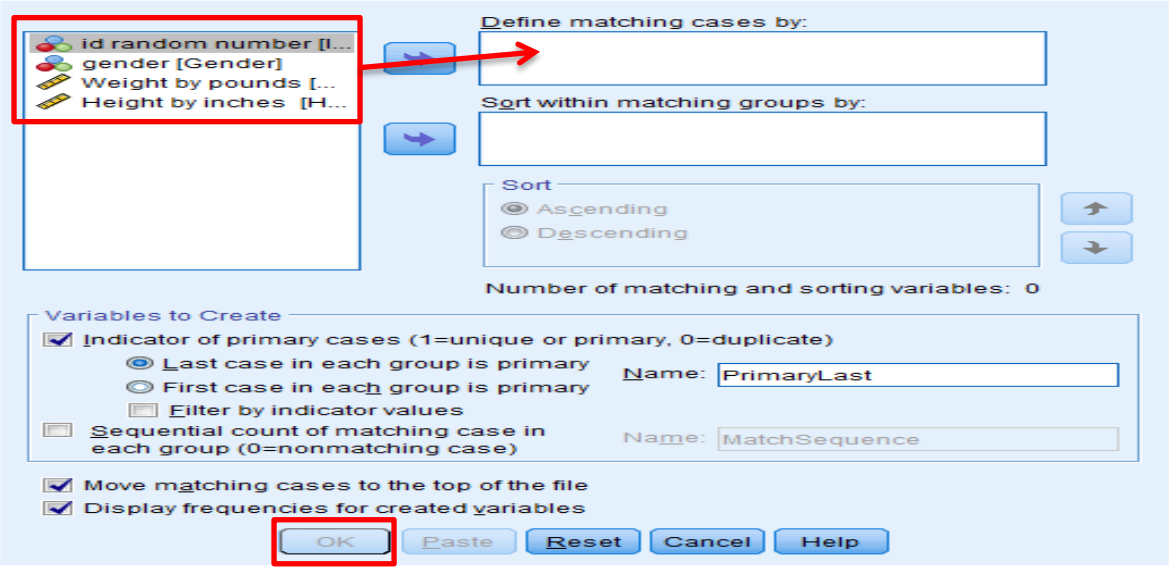
Case	Identifier			
	id random number	gender	Weight by pounds	Height by inches
13	13.00	.	4500.00	62.00
14	14.00	Male	140.00	.

الشكل 03 : جدول يوضح البيانات غير المكتملة Incomplete Identifiers

من خلال الجدول نلاحظ أن Case(الحالات او الاستمارات او المبحوثين) 13 و 14 وهي الحالات التي تم تسجيل وجود بينات غير مكتملة فيها . في حين تشير identifier الى المتغيرات التي تناولناها في المثال وعددها أربع متغيرات .

حيث تتم قراءة الجدول على النحو التالي : في الحالة رقم 13 لم يتم إدخال البيانات المتعلقة بالجنس Gender، أما الحالة رقم 14 لم يتم إدخال البيانات التي تتعلق بالطول height ، كما يمكن التحقق من ذلك بالرجوع الى البيانات التي تم ادخالها في نافذة Data View .

2. عدم وجود حالات مكررة نقوم أولا بالضغط على اختيار Data ضمن شريط القوائم ثم نختار Identify Duplicated Caes، فنتحصل على الشكل التالي:



نقوم بنقل المتغيرات الى اختيار **Define matching cases by** ، ثم نضغط على OK ، لنتحصل على الجدول التالي :

Frequencies

Statistics

Indicator of each last matching case as Primary

N	Valid	20
	Missing	0

Indicator of each last matching case as Primary

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Duplicate Case	2	10.0	10.0	10.0
	Primary Case	18	90.0	90.0	100.0
	Total	20	100.0	100.0	

جدول يوضح الحالات المكررة

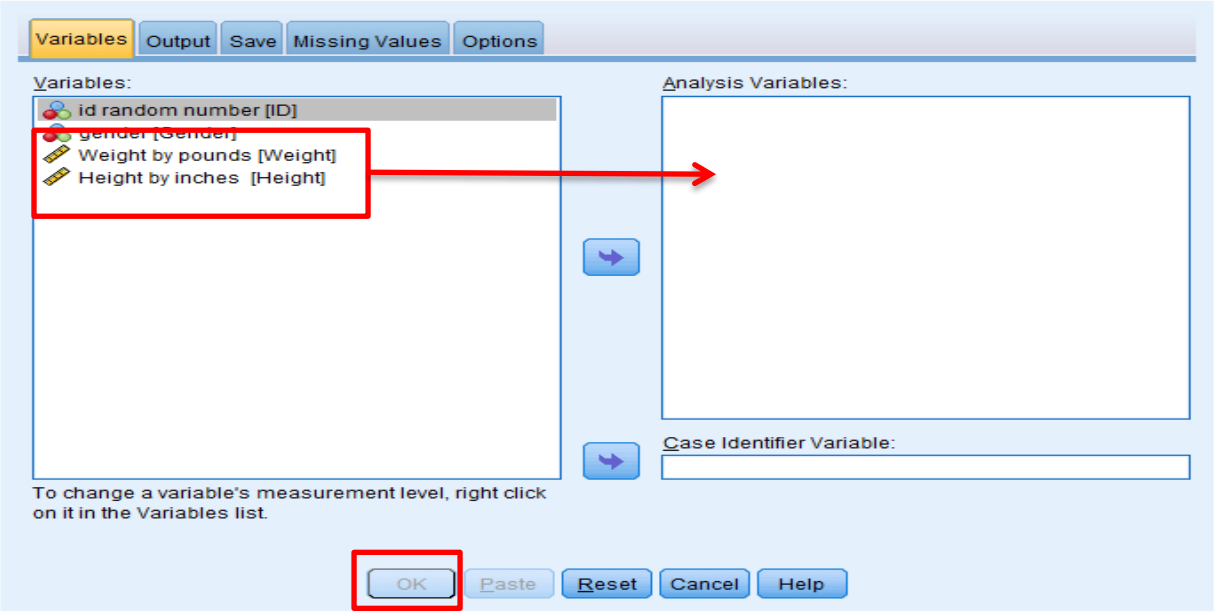
يمثل الجدول الاول تكرار بسيطا لعدد الحالات التي تم ادخالها ، بينما يوضح الجدول الثاني الذي يحمل عنوان Indicator of each last matching case as Primary توضيحا للعدد الحالات التي تم تكرارها ، حيث ان الحالات المكررة يشار اليها بالاسم Duplicated Case وعددها في هذا المثال هو حالتين ، بينما الحالات الاصلية يشار اليها بالاسم Primary Case وعددها 18 .

بعد القيام بهذا الاثبات يقوم البرنامج بإضافة متغير جديد في قائمة Data View يحمل اسم Primary List ، حيث يتم المقارنة بين الحالات الاصلية والحالات المكررة ، كما يظهر في الشكل التالي :

	ID	Gender	Weight	Height	Primary1 ast	var	var	var	var	var	var	var	var	var
1	6.00	Female	145.00	66.00	Duplicate Case									
2	6.00	Female	145.00	66.00	Primary Case									
3	12.00	Male	140.00	60.00	Duplicate Case									
4	12.00	Male	140.00	60.00	Primary Case									
5	1.00	Male	140.00	66.00	Primary Case									
6	2.00	Male	140.00	65.00	Primary Case									
7	3.00	Female	156.00	60.00	Primary Case									
8	4.00	Male	144.00	63.00	Primary Case									
9	5.00	Female	122.00	65.00	Primary Case									
10	7.00	Male	156.00	55.00	Primary Case									
11	8.00	Male	140.00	58.00	Primary Case									
12	9.00	Female	144.00	68.00	Primary Case									
13	10.00	Male	120.00	70.00	Primary Case									
14	11.00	Female	145.00	62.00	Primary Case									
15	13.00	Male	4500.00	62.00	Primary Case									
16	14.00	Male	140.00	60.00	Primary Case									
17	15.00	Female	150.00	65.00	Primary Case									
18	16.00	Male	156.00	68.00	Primary Case									
19	17.00	Female	140.00	70.00	Primary Case									
20	18.00	Male	144.00	66.00	Primary Case									
21														

3.معرفة الكشف عن الحالات غير الاعتيادية

نقوم باختيار **Data** من شريط القوائم ثم اختيار **Identify Unusual cases** .



بعد ذلك نقوم بنقل المتغيرات (في هذه الحالة نقوم بنقل متغير الوزن) الى مربع تحليل المتغيرات Analysis Variables ثم نضغط على OK ، لنتحصل على الجدول التالي :

Anomaly Case Peer ID List

Case	Peer ID	Peer Size	Peer Size Percent
15	1	20	100.0%

Anomaly Case Reason List

Reason: 1

Case	Reason Variable	Variable Impact	Variable Value	Variable Norm
15	Weight	1.000	4500.00	360.3500

يوضح الجدول الذي يحمل عنوان Anomaly Case Reason list ، البيانات غير الطبيعية التي قمنا بادخالها في نافذة عرض البيانات ، حيث يوضح الجدول كل من :

Case: وهي الحالة التي تقع القيمة الطبيعية ضمنها ، في هذا المثال هي الحالة رقم 15.

Reason Variable: نوع المتغير الغير طبيعي في مثالنا هذا هو الوزن weight

Variable Value: القيمة الغير طبيعية والتي تمثل هنا فيهذا المثال 4500.

كما يمكن التحقق منها بالرجوع الى نافذة عرض البيانات Data View ، والتحقق من الحالة رقم 15.