

# ١ الإحصاء: تعريف ومبادئ

- ١- مقدمة
- ٢- طبيعة البيانات
- ٣- العشييرة (المجتمع) والعينة
- ٤- برامج الحزم الجاهزة للتحليل الإحصائي



يمكن تعريف علم الإحصاء بأنه "العلم الذى ينمى ويستخدم أكفأ الطرق فى تجميع وتبويب وترجمة البيانات الكمية بطريقة تمكن من معرفة حجم الخطأ فى الاستنتاجات والتقديرات واتخاذ القرارات فى حالات عدم التأكد وذلك بواسطة الفهم العلمى المبني على أسس الاحتمالات"، وعلى هذا فإن الإحصاء هو أحد فروع الرياضيات التطبيقية.

وتعتبر الإحصاء فى نفس الوقت تكنولوجياً الطريق العلمى، فهى بذلك تمنح الباحث السبيل والتقنيات التى يستعملها فى الوصول إلى استنتاجات معينة وتصميمات محددة من التجارب التى يجريها. وقد أوضح R. A. Fisher أن الإحصاء تشتمل على ثلاثة مجالات للدراسة وهى :

(١) دراسة العشائر (المجموعات)

(٢) دراسة الاختلافات أو التباينات

(٣) دراسة طرق تلخيص البيانات

(١) دراسة العشائر (المجموعات): فكرة العشائر لا تنطبق فقط على الأفراد أو الأشياء، فأى قياس أو مشاهدة تتكرر لا نهائياً فإن مجموع هذه النتائج المتحصل عليها تكون عشيرة (مجتمع) population من هذه القياسات. هذه العشائر هى مجال الدراسة الخاصة بنظرية الإحصاء، وأن حساب المتوسطات والخطأ القياسى هى محاولات لدراسة هذه العشائر.

(٢) دراسة الاختلافات أو التباينات: قيام الإحصاء بدراسة العشائر يودى مباشرة إلى دراسة التباين. فالعشيرة التى تتكون من أفراد متطابقة فى كل النواحي يمكن وصفها عن طريق وصف أحد هذه الأفراد فقط وعددها. والعشائر التى تقوم الإحصاء بدراستها دائماً ما تظهر تبايناً فى إحدى النواحي أو أكثر، والإحصاء يدرس هذا التباين بدلاً من أن يعتبره مصدراً لعدم الدقة يراد التخلص منه.

(٣) دراسة طرق تلخيص البيانات: وهذه عبارة عن دراسة طرق تلخيص البيانات المتوفرة عن العشائر فى عدد قليل من القيم ذات معنى تمكن من توصيف هذه العشائر، فمن المستحيل أن يحتفظ عقل أى باحث بجميع البيانات عن العشيرة أو حتى العينة التى يدرسها. وفى كثير من الأحيان فإنه يمكن توصيف العشيرة التى تنتسب إليها العينة.

## ٢-١ طبيعة البيانات

عند دراسة أى موضوع فإنه من الضروري الحصول على معلومات خاصة به وذلك من خلال إجراء مشاهدات عليه وبالتالي الحصول على قدر من البيانات data (والمفرد datum) عن هذا الموضوع. وعادة ما تكون مجموعات البيانات المتاحة للتحليل الإحصائي مختلفة الأحجام sizes والأشكال shapes وبالتالي فإن طريقة تحليل هذه البيانات تختلف من حالة إلى أخرى حسب طبيعة هذه البيانات. ويمكن أن تقسم البيانات بعدة طرق منها أن تقسم حسب نوعيتها type إلى بيانات وصفية qualitative أو بيانات كمية quantitative. أو تقسم البيانات حسب طريقة قياسها scale of measurement إلى بيانات اسمية nominal أو بيانات رتيبة ordinal أو بيانات فترية interval أو بيانات نسبية ratio. وهذه التقسيمات تعتمد فى النهاية على طبيعة المتغيرات محل الدراسة.

## ١-٢-١ تقسيم البيانات حسب نوعيتها Type of data

## أ - البيانات الوصفية Qualitative data

وهى تلك البيانات التى تقع فى أقسام categories كما لو أنه قد قسمت المشاهدات إلى نوعين (مثلاً حيوانات بيضاء أو سوداء) وفى هذه الحالة فإنه سوف يعطى عدداً لأعداد الحيوانات من هذين النوعين (وقد يكون أكثر من نوعين).

## ب - البيانات الكمية Quantitative data

وهى تلك البيانات الناتجة من قياسات أو تقديرات عددية. فالقياسات تعطى المتغيرات variables. وهنا يظهر نوعان من المتغيرات: الأول قد يمثل عدد الأفراد فى كل تركيب وراثي من مجموعة من التراكيب الوراثية وهذه لا تأخذ إلا قيما (أو وحدات) كاملة وهذه تعرف بالبيانات أو التوزيعات المتقطعة discrete أو غير المستمرة discontinuous، والنوع الآخر قد تكون البيانات فيه تمثل الاختلافات الملاحظة فى الأوزان لمجموعة من الحملان عند فطامها أو لإنتاج اللبن من الأبقار أو الدخل أو الدرجة التى يحصل عليها الطالب فى امتحان ما... الخ. الاختلافات هنا تكون مستمرة continuous ويعرف التوزيع للمتغير فى هذه الحالة بأنه مستمر.

## ٢-٢-١ تقسيم البيانات حسب طريقة قياسها Scale of measure

## أ - البيانات الاسمية Nominal (categorical) data

وهي تلك البيانات الوصفية qualitative والتي تكون على هيئة أقسام ليس لها ترتيب (rank) order أو بناء structure معين أو معنى عددي، وبالتالي لا يمكن إجراء أى عمليات حسابية عليها سوى العد counting. ومن أمثلة تلك البيانات: الجنس gender (ذكور و إناث، والتي يمكن أن يرمز لها مجازاً بالحروف M، F أو 0، 1 عند استخدامها فى التحليل الإحصائى)، اللون (مثلاً أحمر، أخضر...). وفى هذه الأمثلة لا يهم أن تكون الذكور أولاً يليها الإناث أو اللون الأحمر يليه الأخضر... وهكذا.

ويمكن تحليل هذه النوعية من البيانات إحصائياً باستخدام المنوال mode أو مربع كاي  $\chi^2$  أو الإحصاءات اللامعلمية non-parametric statistics.

## ب - البيانات الرتبية Ordinal data

وهي البيانات التي تم ترتيبها بحيث إن الرتبة الأعلى تمثل قيمة أعلى من قيمة الرتبة الأقل مع الأخذ فى الاعتبار أن الفرق بين قيم الرتب ليس بالضرورة أن يكون متساوياً بمعنى أنها تعطى مفهوماً أكبر من أو يساوى أو أصغر من فقط. ويمكن أن يطلق عليها أيضاً أنها بيانات وصفية. ومثال ذلك أن تقسم أفراد العشيرة (المجتمع) طبقاً لمستوى المعيشة إلى طبقات وفيها يمكن القول أن الأفراد الذين ينتمون إلى طبقة فوق المتوسط دخولهم أعلى من الذين ينتمون إلى الطبقة المتوسطة الدخل ولكن لا يمكن تقدير ذلك بطريقة كمية (فلا يمكن أن يقال إن الطبقة الأولى أعلى من الثانية بمقدار 20% مثلاً).

ويمكن تحليل هذه النوعية من البيانات إحصائياً باستخدام الوسيط (الوسط) median أو المنوال mode أو ارتباط الرتب rank correlation أو الإحصاءات اللامعلمية non-parametric statistics.

## ج - البيانات الفترية Interval data (فى فترات)

وهي تلك البيانات الكمية quantitative والتي يمكن ترتيبها وإجراء العمليات الحسابية عليها مثل الجمع والطرح والضرب والقسمة. ومن أمثلة هذه النوعية من البيانات التواريخ dates ودرجة الحرارة temperature. الفرق بين

وحدات هذه النوعية من البيانات له معنى ولكن النسبة ليس لها معنى، فمثلاً في بيانات درجات الحرارة temperature فإن  $10^{\circ} - 10^{\circ} = 20^{\circ} - 20^{\circ} = 30^{\circ}$  ، ولكن  $20^{\circ} / 10^{\circ}$  لا تعنى ضعف الحرارة (السخونة).

ويمكن في مثل هذه النوعية من البيانات حساب المتوسط mean، الانحراف القياسي (المعياري) standard deviation، الارتباط correlation البسيط والمتعدد، الانحدار regression البسيط والمتعدد، تحليل التباين analysis of variance، التحليل العاطلي factor analysis. بمعنى أنه يمكن استخدام الإحصاءات المعلمية parametric statistics.

#### د - البيانات النسبية Ratio data

وهي تلك البيانات التي تنتمي للبيانات الكمية والتي يمكن وضعها على هيئة كسر اعتيادي، بسط ومقام بشرط أن المقام لا يساوي الصفر، وليس الترتيب ومقدار المسافة هما العنصرين المهمين فقط ولكن أيضاً عندما تكون النسبة بين أي قياسين هي الهدف الأساسي للقياس. ومن أمثلة ذلك محيط الصدر والوزن فإنه من المنطقي القول إن 50 كج ضعف 25 كج، وهذه المقولة صحيحة بغض النظر عن وحدات الوزن إذا كانت كيلوجرام أو جرام، وهذا بسبب وجود الصفر الطبيعي natural zero لهذه النوعية من البيانات. وتحتوي مجموعة البيانات النسبية على أعداد صحيحة موجبة أو سالبة أو مساوية للصفر أو كسر اعتيادي أو كسر عشري أو نسبة مئوية. ويوجد عدد لا نهائي من الأعداد النسبية المحصورة بين أي عددين نسبيين. وتجرى جميع العمليات الحسابية على هذه النوعية من البيانات. ويمكن عرض وتلخيص وتحليل هذه النوعية من البيانات إحصائياً بنفس الطرق السابق ذكرها في البيانات الفترية.

#### ٣-١ العشييرة (المجتمع) Population والعينة Sample

في بعض الأحيان، ولكن نادراً ما يمكن قياس كل القيم التي تنتمي إلى موضوع معين، وفي هذه الحالة فإنه يتم التعامل مع ما يعرف بالعشييرة (المجتمع) population وهي تشمل جميع القيم التي تنتمي لوضع ما محدد بتوصيفات معينة. وعلى هذا الأساس فإن العشييرة تحتوي على كل القيم الخاصة بالمتغير الذي يمثل صفة ما أو قياساً معيناً. وإذا أمكن قياس جميع هذه القيم فإن العشييرة من الممكن وصفها دون خطأ. وقد تكون العشييرة محددة finite مثال ذلك جميع الأفراد الناشئة من حدوث طفرة

جديدة لم تظهر مسبقاً (مثلاً)، وبالتالي من الممكن عد وتوصيف جميع أفراد هذه العشيرة. أو قد تكون عشيرة غير محددة infinite وهو الأكثر شيوعاً ويمثلها على سبيل المثال عشيرة الجاموس البرى فى أفريقيا أو نباتات القمح المزروعة فى مصر.

أما العينة sample فهي جزء من العشيرة (وقد تشمل العشيرة كلها). والغرض من استخدام العينة هو الحصول على استدلالات عن العشيرة. وعلى ذلك فإنه من المرغوب فيه أن تحدد العشيرة بقدر من الدقة حتى يمكن الحصول منها على العينة الممثلة والتي تعطى تقديرات سليمة. وفى هذا المجال فإن الأرقام التي تصف العشيرة تسمى بمعالم العشيرة population parameters ومثال هذه المعالم المتوسط الحسابى ( $\mu$ ) وهو المتوسط الذى يدخل فى تحديده كل القيم التي تنتمى إلى هذه العشيرة. ومن ناحية أخرى فإن الأرقام المقابلة والتي تصف أو تحسب من العينة تعرف بالإحصاءات أو التقديرات estimates or sample statistics مثل متوسط العينة  $\bar{X}$  وهي تعتبر تقديرات لمعالم العشيرة.

وتبحث الإحصاء فى العلاقة الرياضية بين الإحصاءات والمعالم فى صورة احتمالية بمعنى أنها تحدد مدى الثقة فى أن يكون الإحصاء المحسوب من العينة تقديراً لمعلم العشيرة.

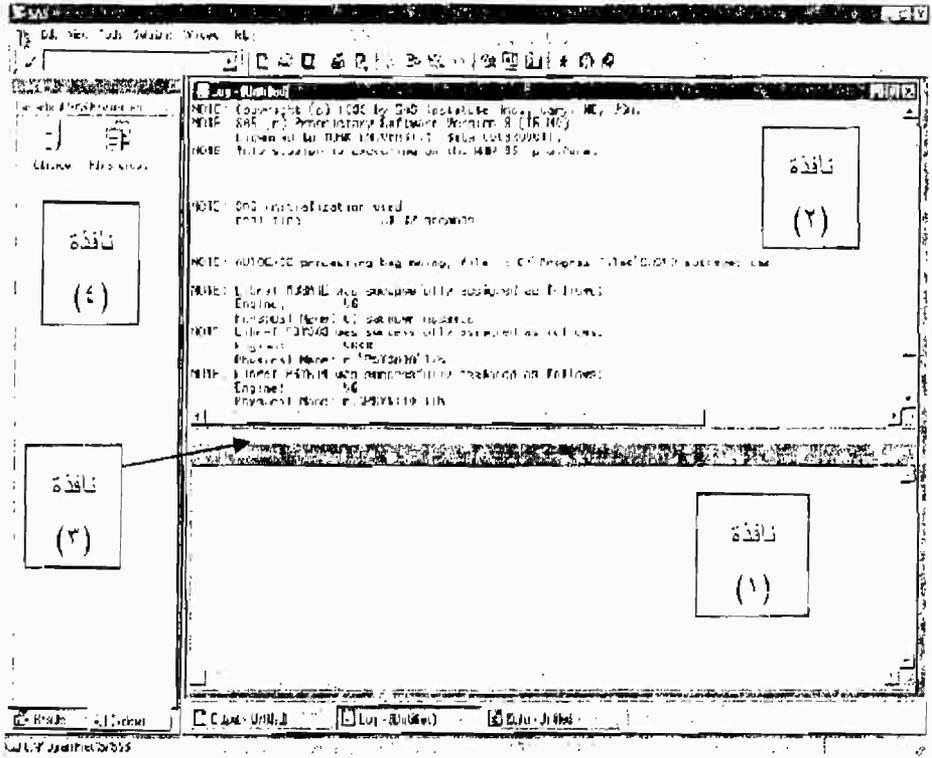
## ١-٤ برامج الحزم الجاهزة للتحليل الإحصائى

### Ready packages for statistical analysis

يوجد عدد كبير من حزم البرامج الجاهزة والتي يمكن استخدامها فى التحليل الإحصائى ومنها برنامج SAS وبرنامج SPSS وهما الأكثر شيوعاً مقارنة بالحزم الأخرى. وهذه الحزم بعضها متاح مجاناً أو بتكلفة بسيطة وبعضها مكلف. وفى هذا الجزء سوف يتم عرض شرح مختصر ومبسط عن كيفية استخدام كل من البرنامجين SAS و SPSS فى التحليل الإحصائى.

### ١-٤-١ برنامج SAS (Statistical Analysis System)

عند بداية تشغيل برنامج SAS يظهر أربع نوافذ رئيسية (وهناك نوافذ أخرى يمكن للمستخدم فتحها). الشكل ١-١ يبين النوافذ الرئيسية لبرنامج SAS



شكل ١-١ النوافذ الرئيسية لبرنامج SAS

نافذة (١) نافذة تحرير البرنامج Program editor window وتستخدم هذه النافذة لكتابة خطوات البرنامج وقد تحتوي أيضا على البيانات data المراد تحليلها.

نافذة (٢) نافذة المسجل Log window ويظهر بهذه النافذة رسائل الملاحظات والأخطاء التي قد تكون في خطوات البرنامج المكتوب بواسطة المستخدم. ويجب أن يتأكد المستخدم من عدم ظهور أى أخطاء بهذه النافذة.

نافذة (٣) نافذة المخرجات Output window ويظهر بها نتائج التحليل بعد تشغيل البرنامج الذي كتبه المستخدم.

نافذة (٤) نافذة المستكشف Explorer window ويظهر بها مكتبة برنامج SAS ويمكن بها عمل تحرير للمخرجات قبل طباعتها.

ويمكن للمستخدم فتح أى من هذه النوافذ (أفقا أو رأسيا) أو إغلاقها أو إعادة فتحها حسب الرغبة.

يتميز برنامج SAS بوجود بعض المفردات vocabulary والجمل المركبة بطريقة معينة syntax خاصة به، يقوم المستخدم لهذا البرنامج بإعداد البيانات المطلوب تحليلها والسؤال المطلوب الإجابة عليه من هذه البيانات بطريقة معينة على هيئة جمل statements وخطوات باستخدام لغة SAS التي تعتمد على كلمات مفتاحية keywords ويتم وضع كل ذلك بترتيب معين. وتقسّم الجمل والخطوات التي توضع في برنامج SAS إلى مجموعتين:

١ - مجموعة خطوات البيانات Data steps

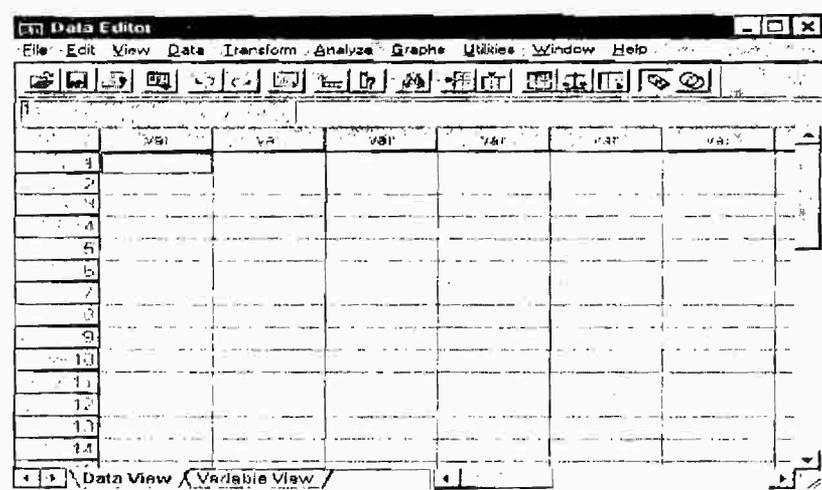
٢ - مجموعة خطوات التحليل Proc steps (Procedure steps)

يوجد داخل كل مجموعة من هاتين المجموعتين جمل لا بد من وضعها لكي يعمل البرنامج بطريقة صحيحة وجمل أخرى اختيارية لا تؤثر في عمل البرنامج.

سوف يتم في الأبواب التالية شرح كيفية استخدام هذا البرنامج تبعا للأمتثلة المعروضة في كل باب.

### ١-٤-٢ برنامج SPSS (Statistical Package for Social Sciences)

يختلف برنامج SPSS عن برنامج SAS بوجود العديد من النوافذ والقفوالم المنسدلة في SPSS والتي تسهل من طريقة الاستخدام حيث إن المستخدم ليس في حاجة لكتابة خطوات البرنامج كما في برنامج SAS. هذا بالإضافة لوجود نافذة لإدخال البيانات والمتغيرات كذلك النافذة الموجودة ببرامج الجداول الإلكترونية (مثلا برنامج EXCEL). عند بداية تشغيل البرنامج تظهر النافذة التالية:



يمكن البدء في إدخال البيانات بنفس طريقة الجداول الإلكترونية. ويمكن التحكم في أسماء المتغيرات ومواصفاتها عن طريق فتح نافذة Variable View أسفل الشاشة. بعد الانتهاء من إدخال البيانات تظهر الشاشة التالية:

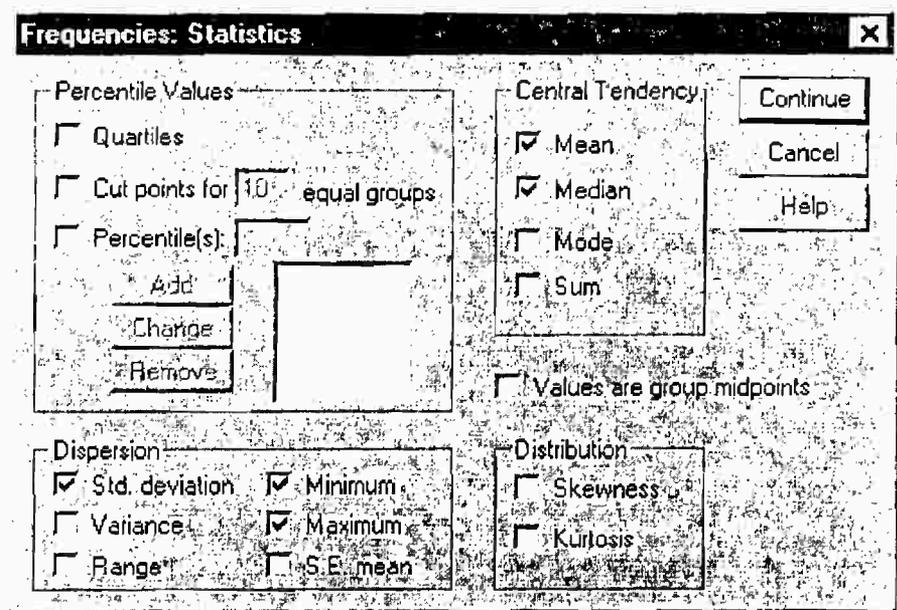
	age	marital	address	income	inccat	car
1	37	Marital status	12	35.00	2.00	1
2	33	1	12	29.00	2.00	1
3	42	1	21	34.00	2.00	1
4	58	0	28	49.00	2.00	3
5	56	0	9	57.00	3.00	1
6	46	0	5	39.00	2.00	1
7	47	1	4	56.00	3.00	2
8	62	1	16	250.00	4.00	7
9	52	1	27	73.00	3.00	3

بعد التأكد من صحة إدخال البيانات يمكن البدء في التحليل الإحصائي عن طريق القوائم المنسدلة من اختيار Analyze فتظهر الشاشة التالية:

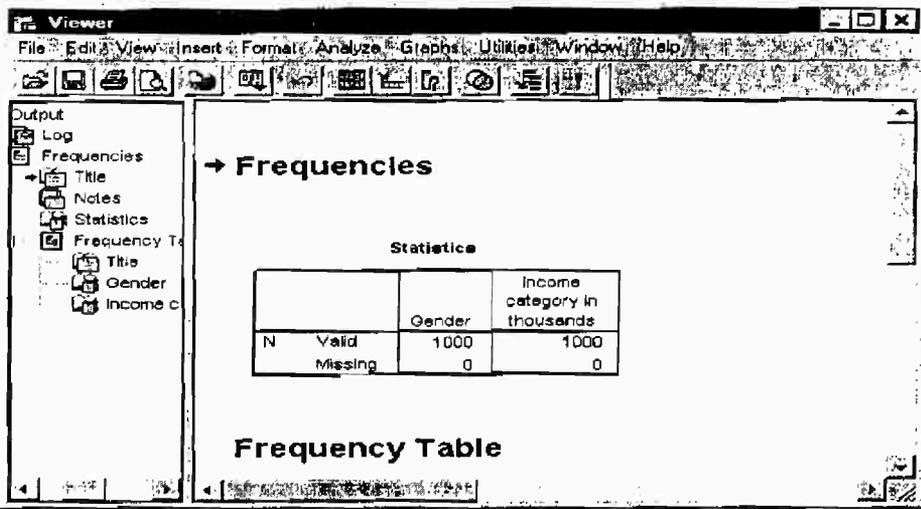
File	Edit	View	Data	Transform	Analyze	Graphs	Utilities	Window	Help
					Reports				
					Descriptive Statistics	Frequencies...			
					Custom Tables	Descriptives...			
					Compare Means	Explore...			
					General Linear Model	Crosstabs...			
					Mixed Models	Ratio...			
					Correlate				
					Regression				
					Loglinear				
					Classify				
					Data Reduction				
					Scale				
					Nonparametric Tests				
					Time Series				
					Survival				
					Multiple Response				
					Missing Value Analysis...				

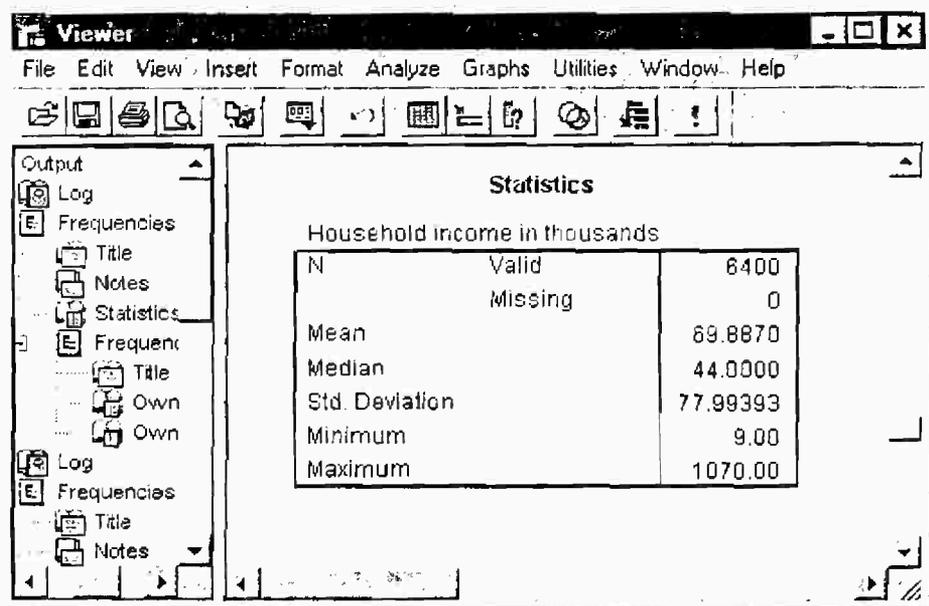


يمكن تحديد الخيارات المطلوبة للتكرارات عن طريق شاشة Statistics فتظهر الشاشة التالية:

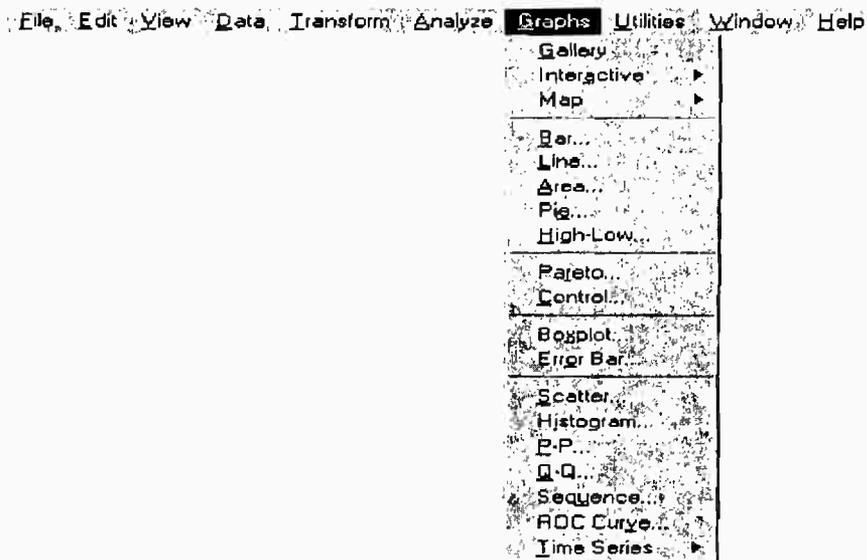


عند الانتهاء من تحديد الخيارات المطلوبة تظهر النتائج كالتالي:

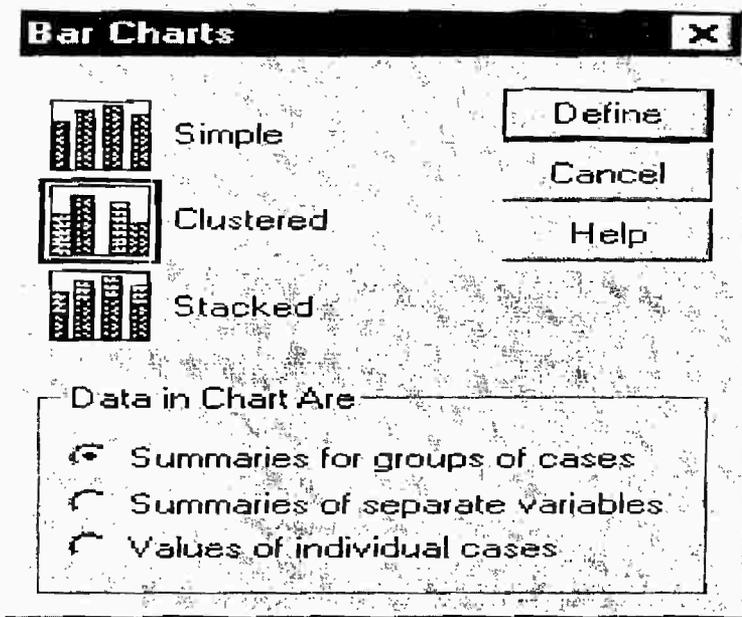




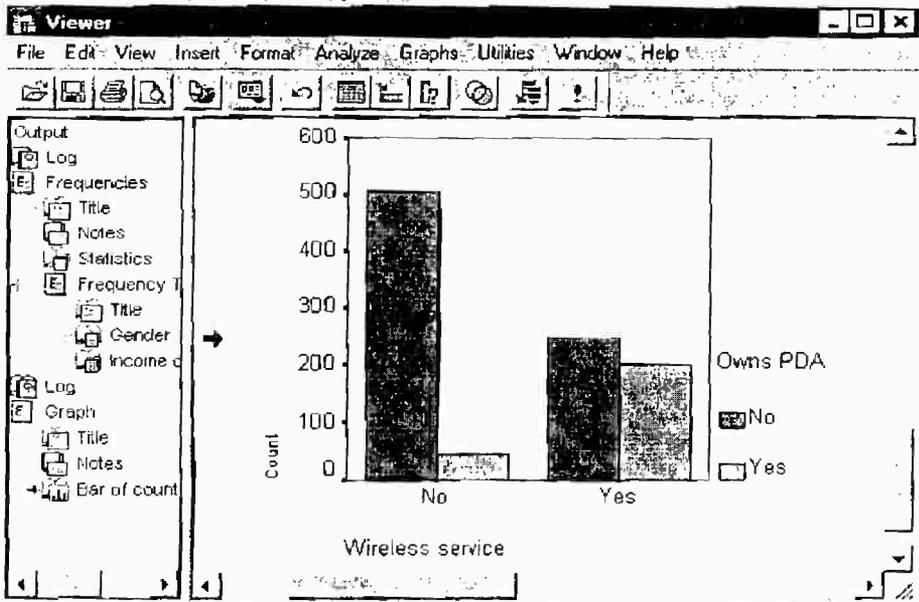
وعند الرغبة في عمل أشكال توضيحية كوسيلة لعرض البيانات يتم اختيار نافذة الأشكال البيانية Graphs كالتالي:



عند اختيار شكل الأعمدة تظهر الشاشة التالية:



وبعد الانتهاء من الخيارات يظهر الشكل التالي:



توضح الأمثلة السابقة أنه يمكن عمل الكثير من التحليلات الإحصائية باستخدام برنامج SPSS بطريقة سهلة للمستخدم.

ومن المفيد ملاحظة أنه يمكن بسهولة استخدام أحد برامج الجداول الإلكترونية مثل EXCEL في تجهيز البيانات المطلوب تحليلها وتخزينها في ملفات مستقلة ثم فتح هذه الملفات بواسطة SAS أو SPSS بدون الحاجة إلى إعادة إدخال هذه البيانات مرة أخرى.