١ الإحصاء: تعريف ومبادئ

١– مقدمة

٢– طبيعة البيانات

٣– العشيرة (المجتمع) والعينة

٤– برامج الحزم الجاهزة للتحليل الإحصائى

obeikandl.com

obeikandl.com

١-١ مقدمة

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 المبنى على أسس الاحتمالات"، وعلى هذا فإن الإحصاء هو أحد فروع الرياضيات التطبيقية.

وتعتبر الإحصاء فى نفس الوقت تكنولوجية 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 (المجتمعات)

(٢) دراسة الاختلافات أو التباينات

(٣) دراسة طرق تلخيص البيانات

(١) **دراسة العشائر (المجتمعات):** 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 يؤدى مباشرة إلى دراسة التباين. فالعشيرة التى تتكون من أفراد متطابقة فى كل النواحى يمكن وصفها عن طريق وصف أحد هذه الأفراد فقط وعددها. والعشائر التى تقوم الإحصاء بدراستها دائما ما تظهر تبايناً فى إحدى النواحى أو أكثر، والإحصاء يدرس هذا التباين بدلا من أن يعتبره مصدرا 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.

١-٢ طبيعة البيانات

عند دراسة أى موضوع فإنه من الضرورى الحصول على معلومات خاصة به وذلك من خلال إجراء مشاهدات عليه وبالتالى الحصول على قدر من البيانات data (والمفرد datum) عن هذا الموضوع. وعادة ما تكون مجموعات البيانات المتاحة للتحليل الإحصائى مختلفة الأحجام sizes والأشكال shapes وبالتالى 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 رتبية 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 (مثلا حيوانات بيضاء أو سوداء) وفى هذه الحالة فإنه سوف يعطى عداً 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 وراثى 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.

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

أ - البيانات الاسمية Nominal (categorical) data

وهى تلك البيانات الوصفية qualitative والتى تكون على هيئة أقسام ليس لها ترتيب order (rank) أو بناء structure معين أو معنى عددى، وبالتالى لا يمكن إجراء أى عمليات حسابية عليها سوى العد counting. ومن أمثلة تلك البيانات: الجنس gender (ذكور و إناث، والتى يمكن أن يرمز لها مجازا بالحروف M، F أو 0، 1 عند استخدامها فى التحليل الإحصائى)، اللون (مثلا 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 إحصائيا باستخدام المنوال mode أو مربع كاى χ^2 أو الإحصاءات اللامعلمية non-parametric statistics.

ب - البيانات الرتبية Ordinal data

وهى البيانات التى 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 متساويا بمعنى أنها تعطى مفهوما أكبر من أو يساوى أو أصغر من فقط. ويمكن أن يطلق عليها أيضا أنها بيانات وصفية. ومثال ذلك أن تقسم أفراد العشيرة (المجتمع) طبقا 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 إحصائيا باستخدام الوسيط (الوسط) median أو المنوال mode أو ارتباط الرتب rank correlation أو الإحصاءات اللامعلمية non-parametric statistics.

ج - البيانات الفترية (فى فترات) Interval data

وهى تلك البيانات الكمية quantitative والتى 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، فمثلا فى بيانات درجات الحرارة temperature فإن
$10^{\circ} = 20^{\circ} - 10^{\circ} = 30^{\circ} - 20^{\circ}$ ، ولكن $20^{\circ}/10^{\circ}$ لا تعنى ضعف الحرارة (السخونة).

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

د – البيانات النسبية Ratio data

وهى تلك البيانات التى تنتمى للبيانات الكمية والتى يمكن وضعها على هيئة كسر اعتيادى، بسط ومقام بشرط أن المقام لا يساوى الصفر. وليس الترتيب ومقدار المسافة هما العنصرين المهمين فقط ولكن أيضا عندما تكون النسبة بين أى قياسين هى الهدف الأساسى للقياس. ومن أمثلة ذلك محيط الصدر والوزن فإنه من المنطقى القول إن 50 كج ضعف 25 كج، وهذه المقولة صحيحة بغض النظر عن وحدات الوزن إذا كانت كيلوجرام أو جرام، وهذا بسب وجود الصفر الطبيعى natural zero لهذه النوعية من البيانات. وتحتوى مجموعة البيانات النسبية على أعداد صحيحة موجبة أو سالبة أو مساوية للصفر أو كسر اعتيادى أو كسر عشرى أو نسبة مئوية. ويوجد عدد لا نهائى من الأعداد النسبية المحصورة بين أى عددين نسبيين. وتجرى جميع العمليات الحسابية على هذه النوعية من البيانات.
ويمكن عرض وتلخيص وتحليل هذه النوعية من البيانات إحصائيا بنفس الطرق السابق ذكرها فى البيانات الفترية.

١-٣ العشيرة (المجتمع) Population والعينة Sample

فى بعض الأحيان، ولكن نادراً ما يمكن قياس كل القيم التى تنتمى إلى موضوع معين، وفى هذه الحالة فإنه يتم التعامل مع ما يعرف بالعشيرة (المجتمع) population وهى تشمل جميع القيم التى تنتمى لوضع ما محدد بتوصيفات معينة. وعلى هذا الأساس فإن العشيرة تحتوى على كل القيم الخاصة بالمتغير الذى يمثل صفة ما أو قياسا معينا. وإذا أمكن قياس جميع هذه القيم فإن العشيرة من الممكن وصفها دون خطأ. وقد تكون العشيرة محددة finite مثال ذلك جميع الأفراد الناشئة من حدوث طفرة

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

أما العينة sample فهى 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 والتى تعطى تقديرات سليمة. وفى هذا المجال فإن الأرقام التى تصف العشيرة تسمى بمعالم العشيرة population parameters ومثال هذه المعالم المتوسط الحسابى (μ) وهو المتوسط الذى يدخل فى تحديده كل القيم التى تنتمى إلى هذه العشيرة. ومن ناحية أخرى فإن الأرقام المقابلة والتى تصف أو تحسب من العينة تعرف بالإحصاءات أو التقديرات estimates or sample statistics مثل متوسط العينة $\overline{X}$ وهى 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 تقديرا لمعلم العشيرة.

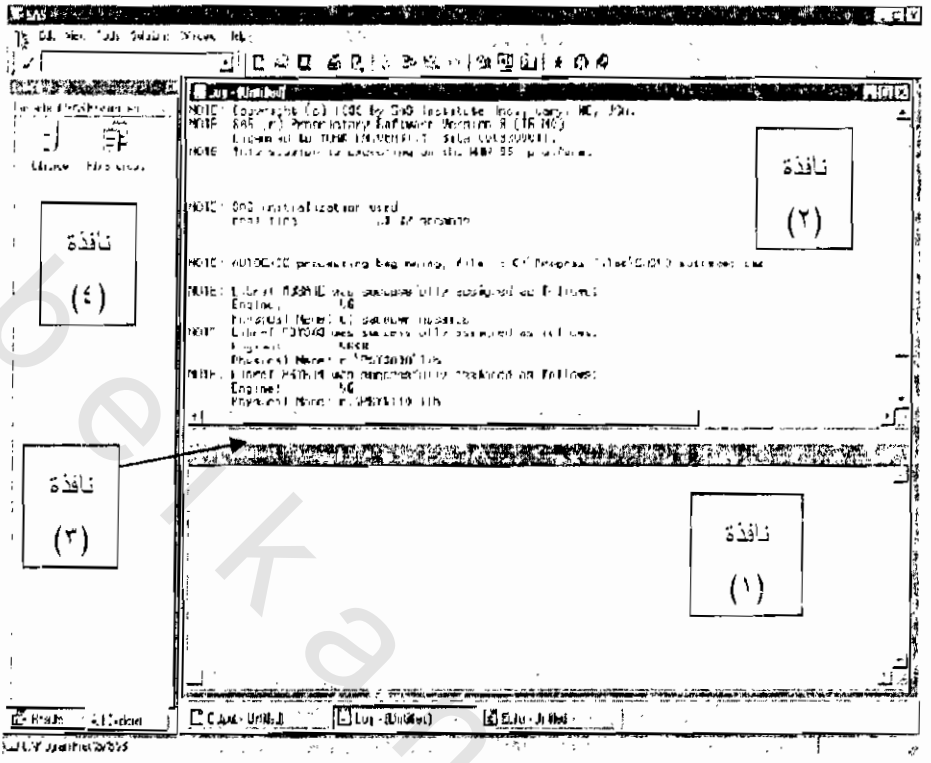
١-٤ برامج الحزم الجاهزة للتحليل الإحصائى

Ready packages for statistical analysis

يوجد عدد كبير من حزم البرامج الجاهزة والتى يمكن استخدامها فى التحليل الإحصائى ومنها برنامج SAS وبرنامج SPSS وهما الأكثر شيوعا مقارنة بالحزم الأخرى. وهذه الحزم بعضها متاح مجانا 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 تحتوى أيضا على البيانات data المراد تحليلها.

نافذة (٢) نافذة المسجل Log window ويظهر بهذه النافذة رسائل الملاحظات والأخطاء التى قد تكون فى 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 الذى 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 (أفقيا أو رأسيا) أو غلقها أو إعادة فتحها حسب الرغبة.

obeikandi.com

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 التى تعتمد على كلمات مفتاحية keywords ويتم وضع كل ذلك بترتيب معين. وتقسم الجمل والخطوات التى توضع فى برنامج SAS إلى مجموعتين:

١ – مجموعة خطوات البيانات Data steps

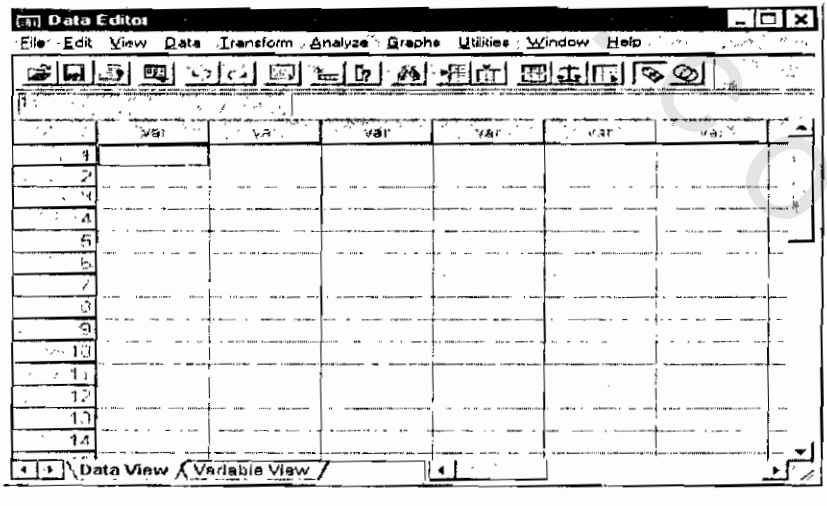
٢ – مجموعة خطوات التحليل Proc steps (Procedure steps)

يوجد داخل كل مجموعة من هاتين المجموعتين جمل لابد من وضعها لكى يعمل البرنامج بطريقة صحيحة وجمل أخرى اختيارية لا تؤثر فى عمل البرنامج.

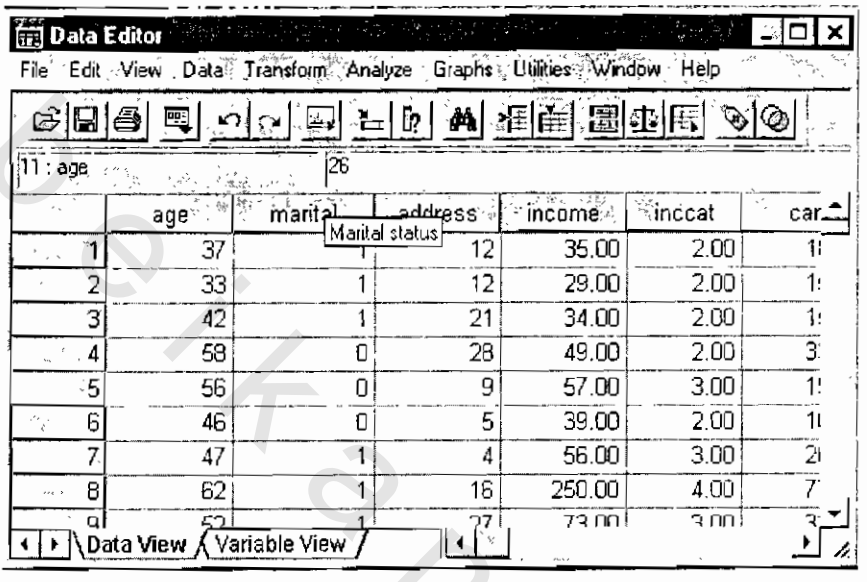
سوف يتم فى الأبواب التالية شرح كيفية استخدام هذا البرنامج تبعا للأمثلة المعروضة فى كل باب.

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

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

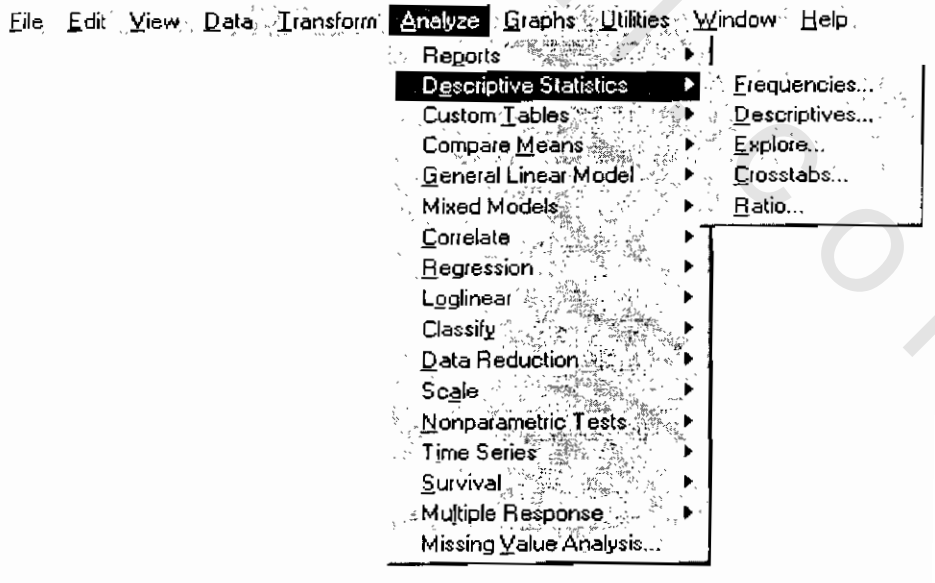


يمكن البدء فى إدخال البيانات بنفس طريقة الجداول الإلكترونية. ويمكن التحكم فى أسماء المتغيرات ومواصفاتها عن طريق فتح نافذة Variable View أسفل الشاشة. بعد الانتهاء من إدخال البيانات تظهر الشاشة التالية:



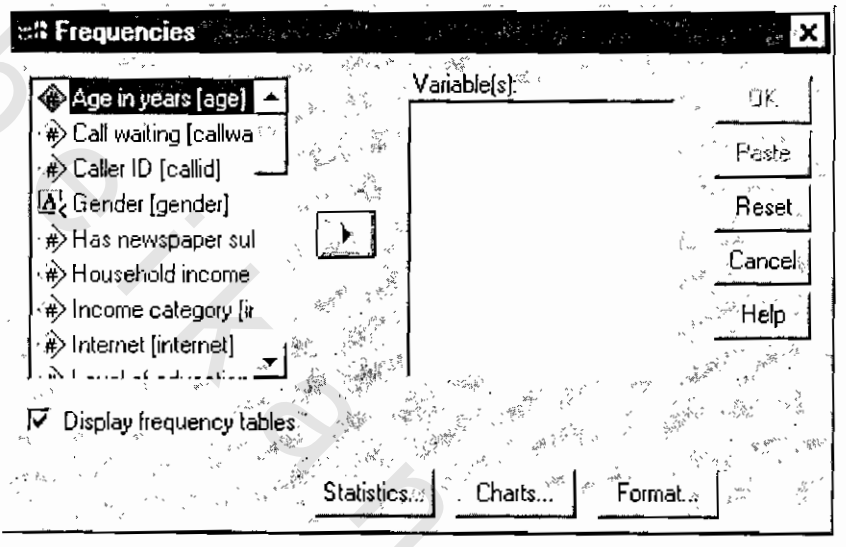
	age	marital	address	income	inccat	car
1	37	1	12	35.00	2.00	1
2	33	1	12	29.00	2.00	1
3	42	1	21	34.00	2.00	1
4	58	0	28	49.00	2.00	3
5	56	0	9	57.00	3.00	1
6	46	0	5	39.00	2.00	1
7	47	1	4	56.00	3.00	2
8	62	1	16	250.00	4.00	7
9	52	1	27	73.00	3.00	3

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

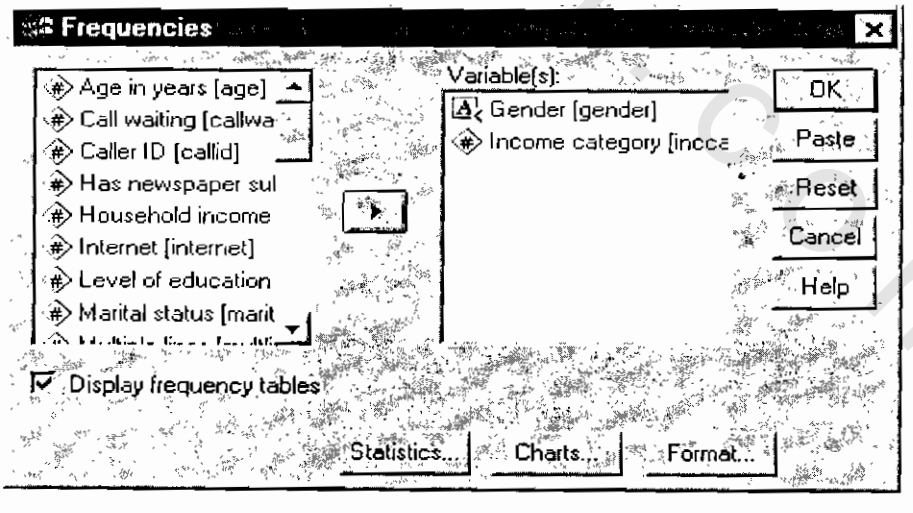


هذه الشاشة تحتوى على جميع التحليلات الإحصائية التى يمكن عملها بواسطة هذا البرنامج

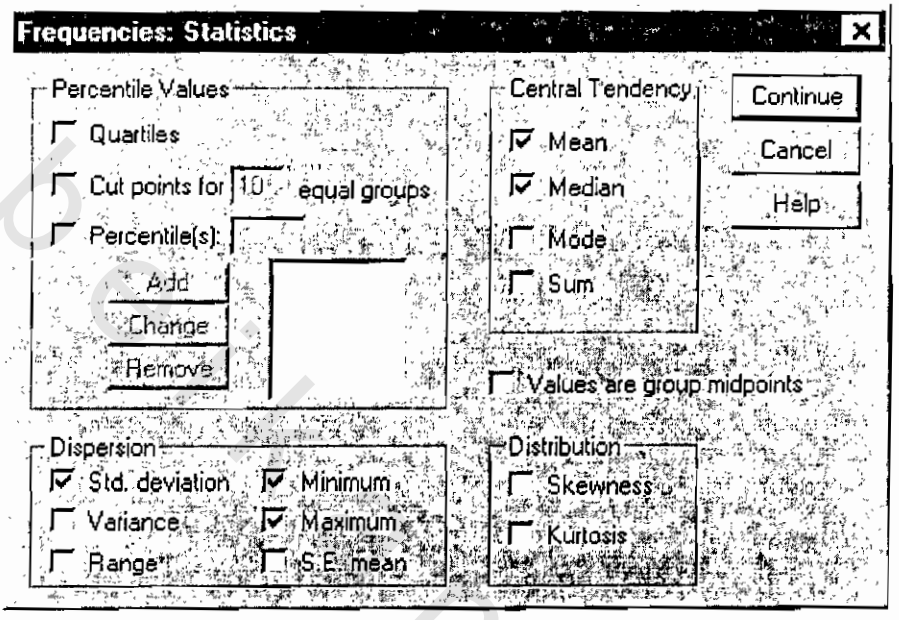
على سبيل المثال عند اختيار Descriptive Statistics ثم Frequencies تظهر شاشة أخرى كالتالى:



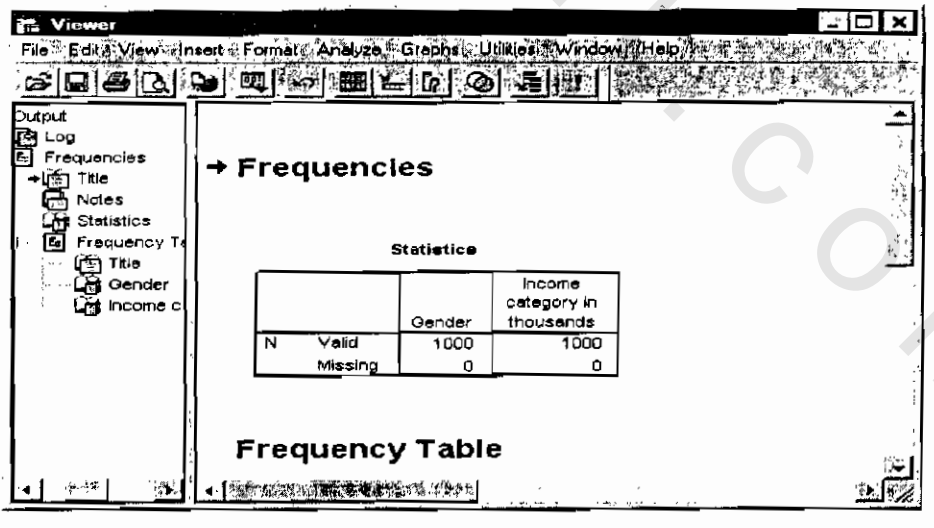
من قائمة المتغيرات يمكن اختيار المتغيرات المرغوب فى عمل تكرارات لها كالتالى:

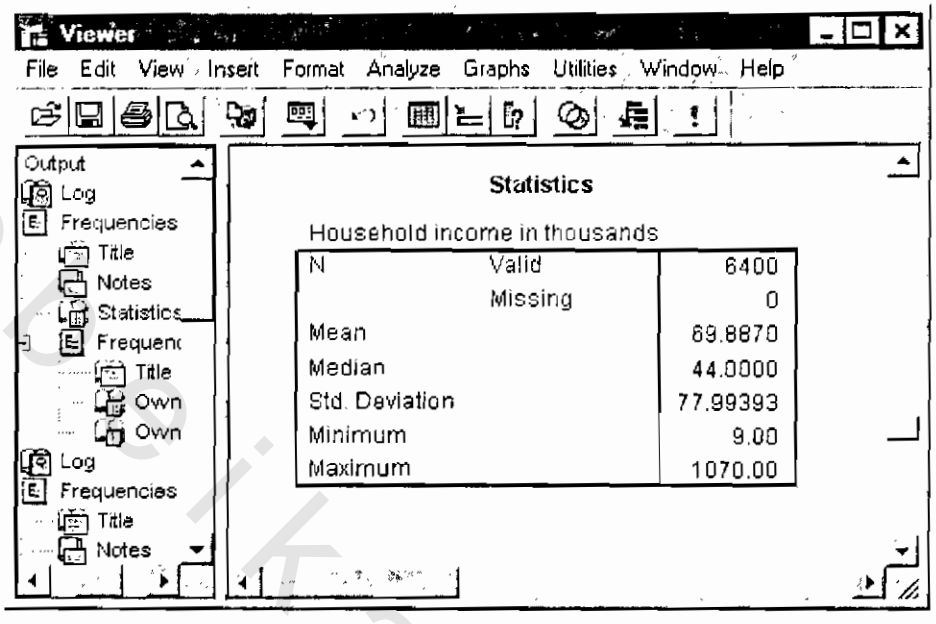


يمكن تحديد الخيارات المطلوبة للتكرارات عن طريق شاشة Statistics فتظهر الشاشة التالية:

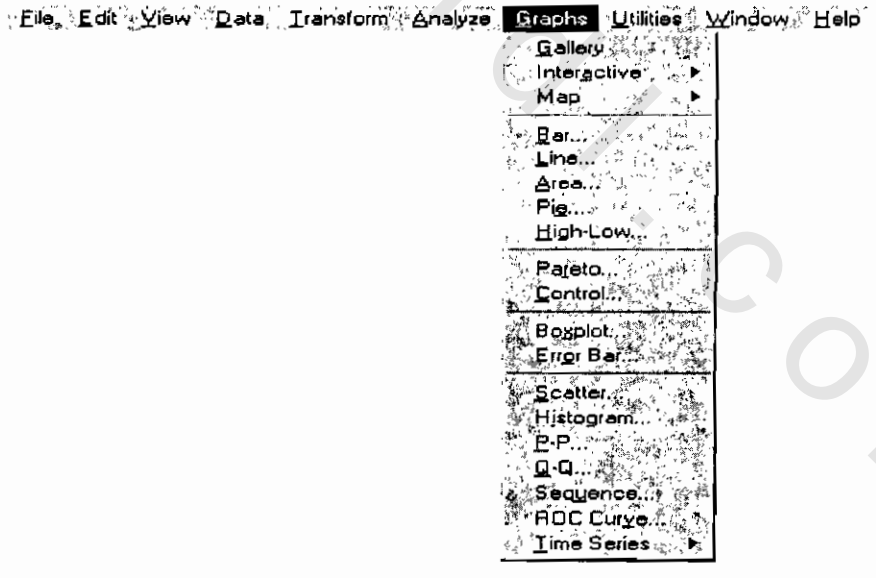


عند الانتهاء من تحديد الخيارات المطلوبة تظهر النتائج كالتالى:

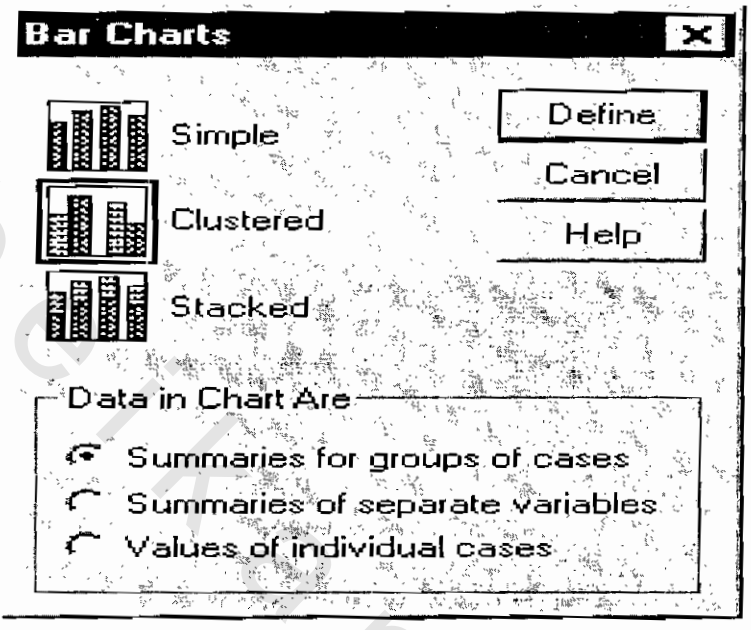




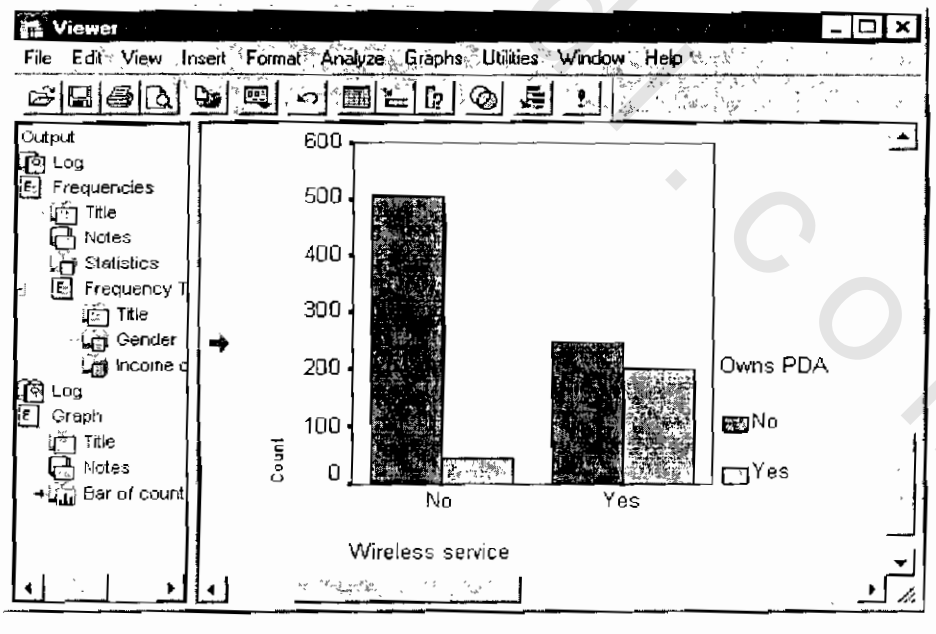
وعند الرغبة فى عمل أشكال توضيحية كوسيلة لعرض البيانات يتم اختيار نافذة الأشكال البيانية Graphs كالتالى:



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 فى تجهيز البيانات المطلوب تحليلها وتخزينها فى ملفات مستقلة ثم فتح هذه الملفات بواسطة SAS أو SPSS بدون الحاجة إلى إعادة إدخال هذه البيانات مرة أخرى.

obeikandl.com