



Ejercicios con ordenador

Conglomerados

Prof. Esperanza Ayuga Téllez

Tareas a realizar en la práctica



• *Análisis de los Conglomerados*

En este caso concreto:

Preparación de los datos

Selección del método, algoritmo y distancia de conglomeración

Selección del número de conglomerados

Asignar objetos a conglomerados y cálculo de centroides

Validar la agrupación

Ejemplo con datos del sector forestal

Información



Información sobre los datos y resultados de las pruebas realizadas

De cada una

Comentario explicativo

Preparación de Datos

Los datos son de diferente magnitud. Se han transformado con... y eliminado las observaciones.....

Selección de métodos y número de grupos

Se ha utilizado el método de las k-medias con distancia euclídea y el cálculo de F se determinan 3 grupos

Información



Preparación de Datos

STATGRAPHICS Plus - StatFolio sin Nombre - [Análisis Multivariable]

Archivo Edición Gráficos Descripción Comparación Dependencia Avanzado SnapStats!! Ver Ventana Ayuda

Etiquetas: Filas: 2

Resumen del Procedimiento

	Muestra 1:	Muestra 2:	n
Frecuencia	17	17	1
Media	2,97596E6	439694,0	1
Desviación típica	3,04819E6	464620,0	1
Coef. de variación	102,427%	105,669%	9

	Muestra 5:	Muestra 6:	n
Frecuencia	17	17	1
Media	27,6229	0,412941	1
Desviación típica	22,5815	0,387052	1
Coef. de variación	81,7491%	93,7306%	1

Covarianzas

	Superficie	SF pública	S
Superficie	9,29148E12 (17)	1,2775E12 (17)	3
SF pública	1,2775E12 (17)	2,15871E11 (17)	3
SF privada	3,09178E12 (17)	3,77542E11 (17)	1
Población total	9,13996E11 (17)	8,22582E10 (17)	4
Inversiones por ha F-1	-1,21788E7	-2,64125E6	-

Utilice el botón alternativo para seleccionar opciones

NUM

Inicio Bandeja de entrada ... Urban studies (refer... Prácticas Microsoft PowerPoint... STATGRAPHICS Plus ... ES 12:59

Información



Análisis Cluster

The screenshot displays the STATGRAPHICS Plus software interface. The title bar reads "STATGRAPHICS Plus - StatFolio sin Nombre". The menu bar includes "Archivo", "Edición", "Gráficos", "Descripción", "Comparación", "Dependencia", "Avanzado", "SnapStats!!", "Ver", "Ventana", and "Ayuda". The "Avanzado" menu is open, showing options: "Control de Calidad", "Diseño Experimental", "Análisis Series Temporales", "Métodos Multivariables", and "Regresión Avanzada". The "Métodos Multivariables" sub-menu is also open, listing "Componentes Principales...", "Análisis Factorial...", "Análisis Cluster...", "Análisis Discriminante...", and "Correlaciones Canónicas...". The "Análisis Cluster..." option is highlighted. The status bar at the bottom shows "sector forestal.sf3".

Información



Elección de métodos y distancias



STATGRAPHICS Plus - StatFolio sin Nombre

Archivo Edición Gráficos Descripción Comparación Dependencia Avanzado SnapStats! Ver Ventana Ayuda

Análisis Cluster

Etq.: Fil:

Resumen del Análisis

Datos/Variables:
Superficie
SF pública
SF privada

Inversiones por ha F
Contratos F
E aserrado
E tableros
E carpintería
Embalajes
fábrica de muebles
Otros

Número de casos completos: 17
Método Cluster: Vecino más cercano (Una sola Unión)
Distancia Métrica: Euclidean Cuadrado

Cluster	Miembros	Porcentaje
1	17	100,00

Centroides

Cluster	Superficie	SF pública	SF privada	Población to Inversiones	Contratos F	E aserrado	E tableros	
1	2,97596E6	439694,0	1,08878E6	2,14035E6	27,6229	0,412941	106,412	30,7059

Opciones Análisis Cluster

Método

- Vecino más próximo
- Vecino más lejano
- Centroide
- Mediana
- Media de Grupo
- Ward
- k-Medias

Distancia Métrica

- Euclidiano Cuadrado
- Euclidiano
- Bloque de Población

Cluster

- Observaciones
- Variables

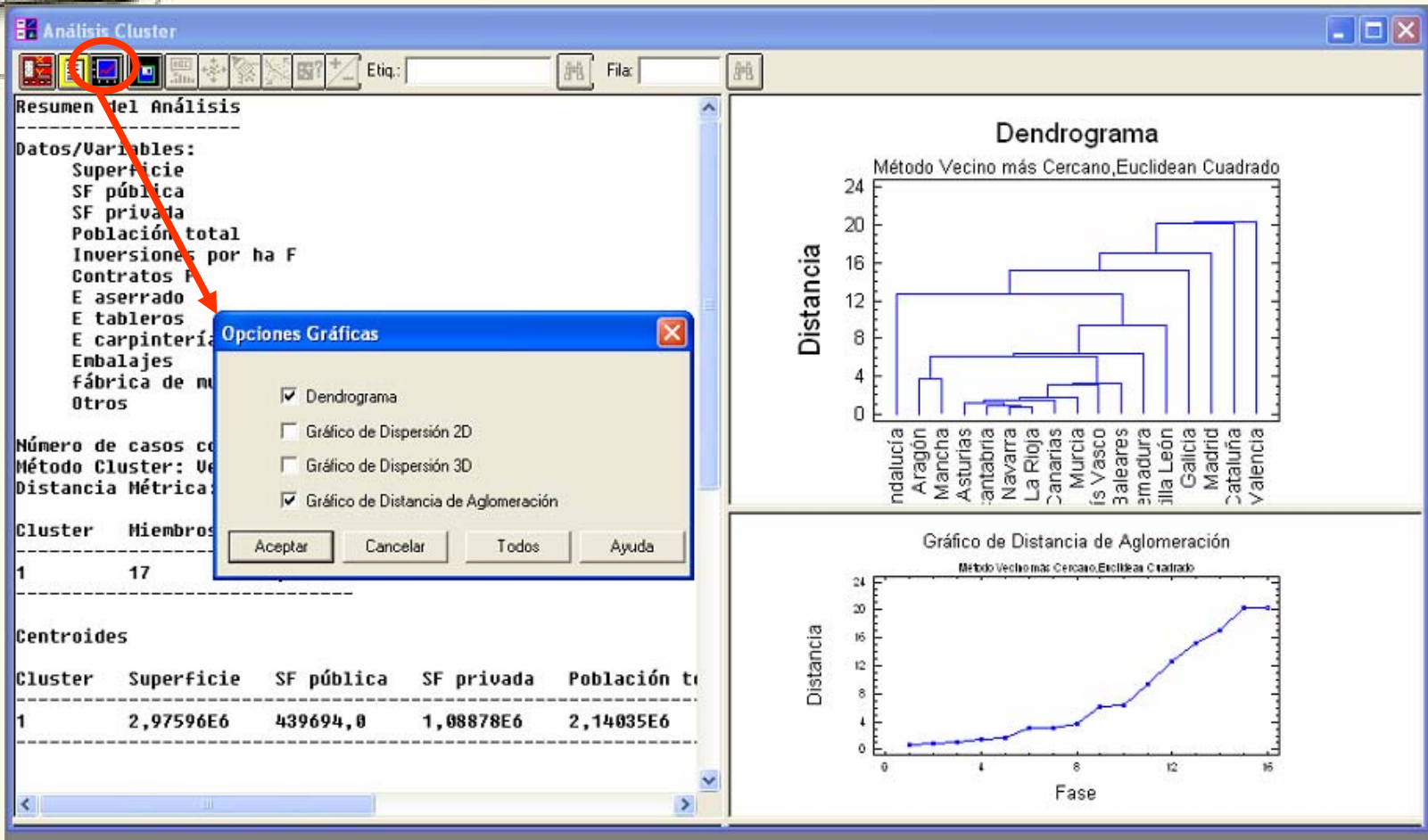
Número de Clusters: 1

Estandarizado

Aceptar
Cancelar
Particiones...
Ayuda

Información

Elección del número de grupos



The screenshot shows the 'Análisis Cluster' software interface. The 'Resumen del Análisis' panel on the left lists variables such as 'Superficie', 'SF pública', 'SF privada', 'Población total', 'Inversiones por ha F', 'Contratos P', 'E aserrado', 'E tableros', 'E carpintería', 'Embalajes', 'fábrica de m', and 'Otros'. The 'Número de casos co', 'Método Cluster: Ue', and 'Distancia Métrica' are also visible. The 'Opciones Gráficas' dialog box is open, showing the following options:

- Dendrograma
- Gráfico de Dispersión 2D
- Gráfico de Dispersión 3D
- Gráfico de Distancia de Aglomeración

The 'Aceptar', 'Cancelar', 'Todos', and 'Ayuda' buttons are at the bottom of the dialog. The 'Dendrograma' graph on the right shows the hierarchical clustering process using the 'Método Vecino más Cercano, Euclidean Cuadrado' method. The y-axis is 'Distancia' (0 to 24) and the x-axis lists regions: Andalucía, Aragón, Mancha, Asturias, Cantabria, Navarra, La Rioja, Canarias, Murcia, País Vasco, Baleares, Comunidad de Madrid, Galicia, Cataluña, and Valencia. The 'Gráfico de Distancia de Aglomeración' graph shows the distance between clusters as a function of the number of clusters (Fase). The y-axis is 'Distancia' (0 to 24) and the x-axis is 'Fase' (0 to 16).

Resultados gráficos



Incluir en análisis de cluster

Gráfico de aglomeración

Dendrograma

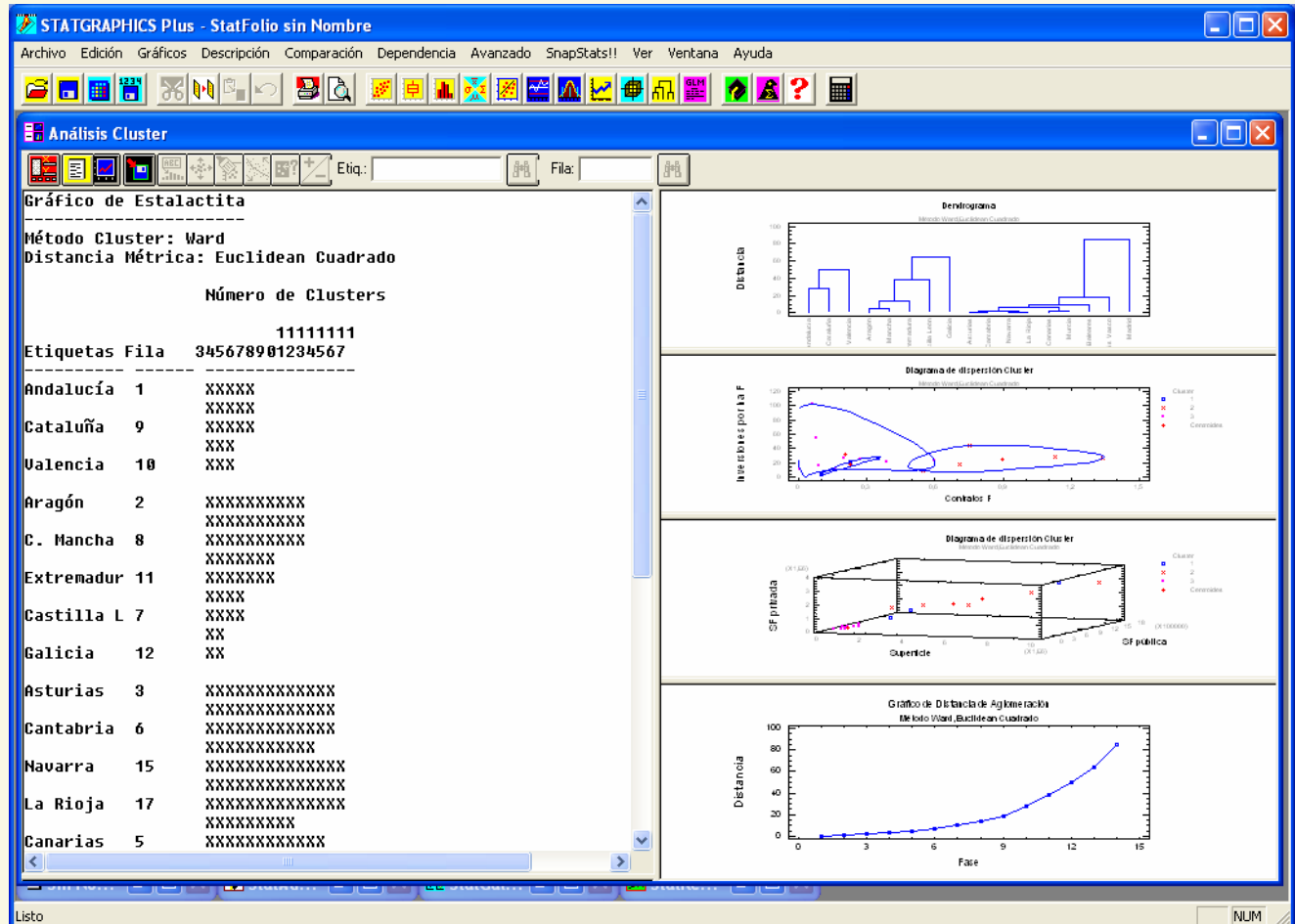
Observaciones agrupadas frente a variables

Gráfico de contornos de cluster

Resultados gráficos



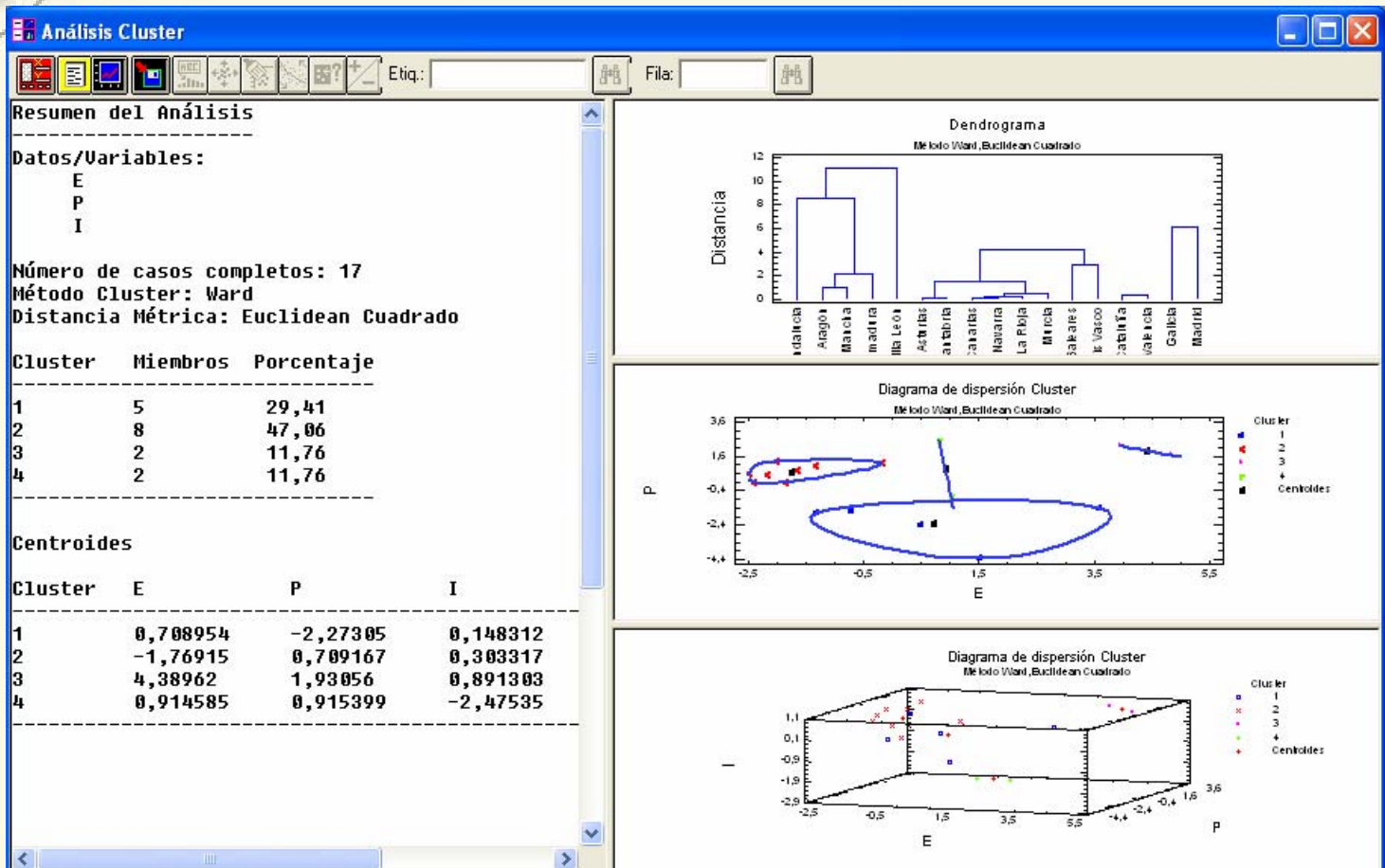
- Se obtienen en opciones gráficas y el gráfico de estalactita en opciones tabulares



Resultados



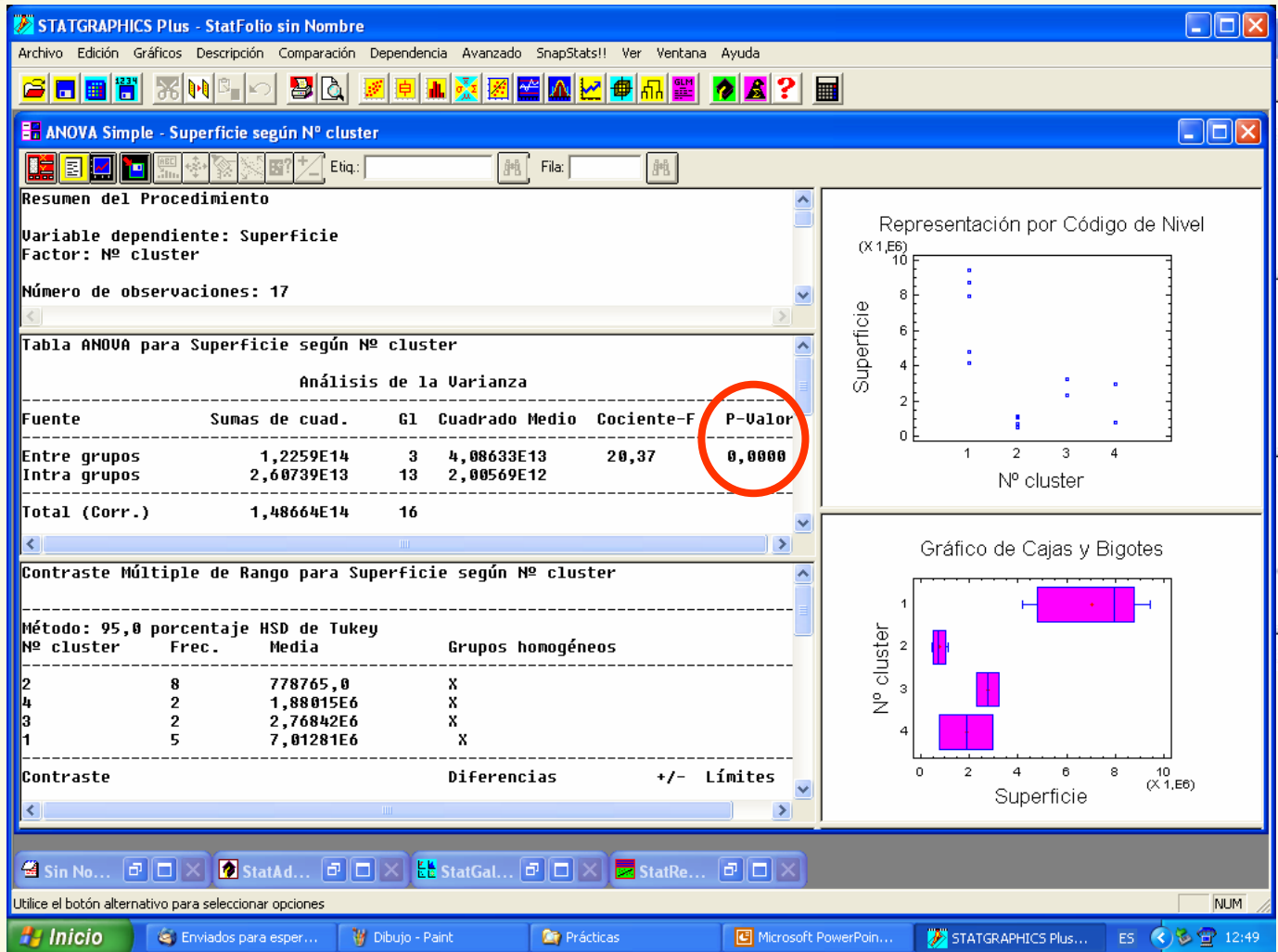
- Validación de resultados: Centroides y Contorno de Cluster



Resultados



- Validación de resultados ANOVA



Resultados



- Interpretación de resultados

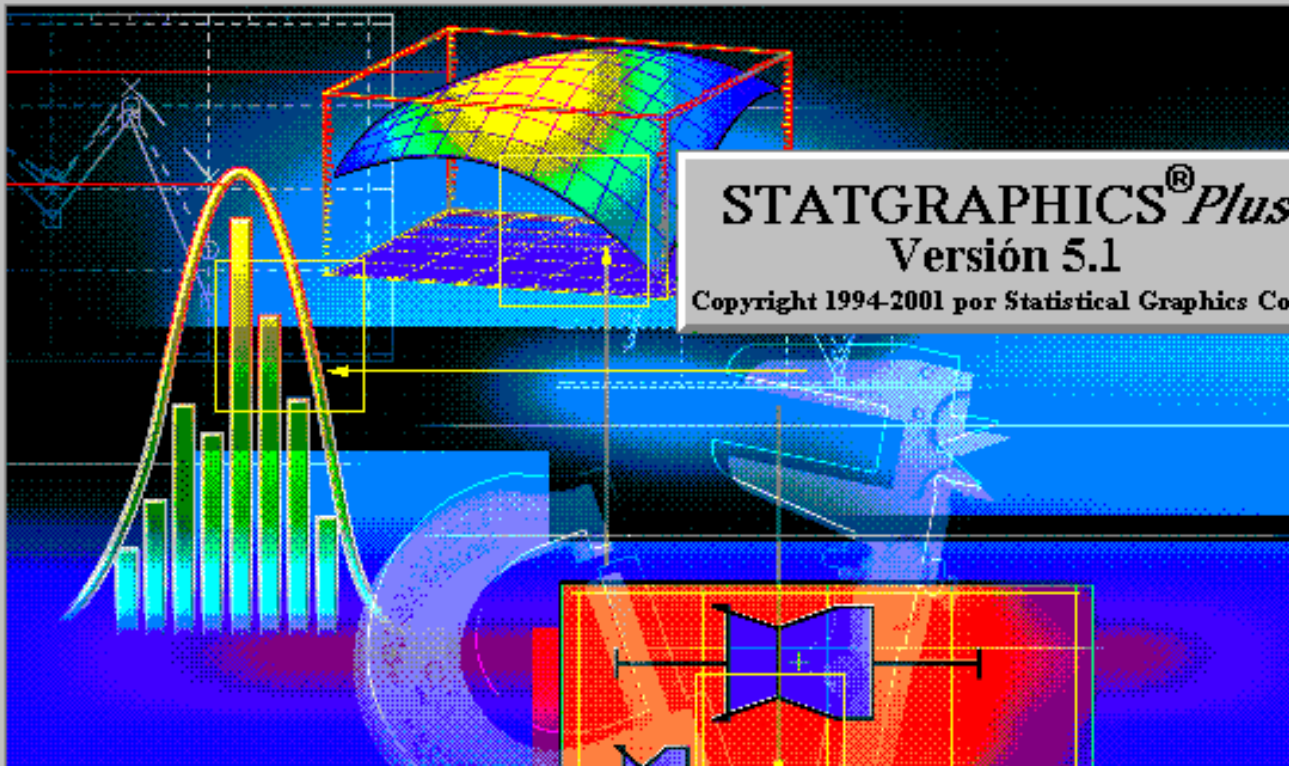
Los 3 grupos lo constituyen las observaciones...
La característica del primer grupo es un valor
alto de la variable X y medio de Y.....

.....

- Consecuencias prácticas

Se pueden diferenciar tres grupos distintos
caracterizados por..... que muestran las
diferentes zonas en que el sector forestal tiene
diferentes grados de desarrollo.

EJERCICIO



The image displays the STATGRAPHICS Plus software interface. It features a central 3D surface plot with a rainbow color gradient. To the left, there is a bar chart with green bars and a yellow normal distribution curve overlaid. Below the bar chart, a pie chart is partially visible. In the bottom right corner, there is a box plot with a red background. The background of the interface is dark blue with a grid pattern. A white text box in the upper right corner contains the following text:

STATGRAPHICS[®] Plus
Versión 5.1
Copyright 1994-2001 por Statistical Graphics Corp.

Este programa está protegido por las leyes internacionales y de Estados Unidos de América referentes a los derechos de propiedad según se indica en la caja.