

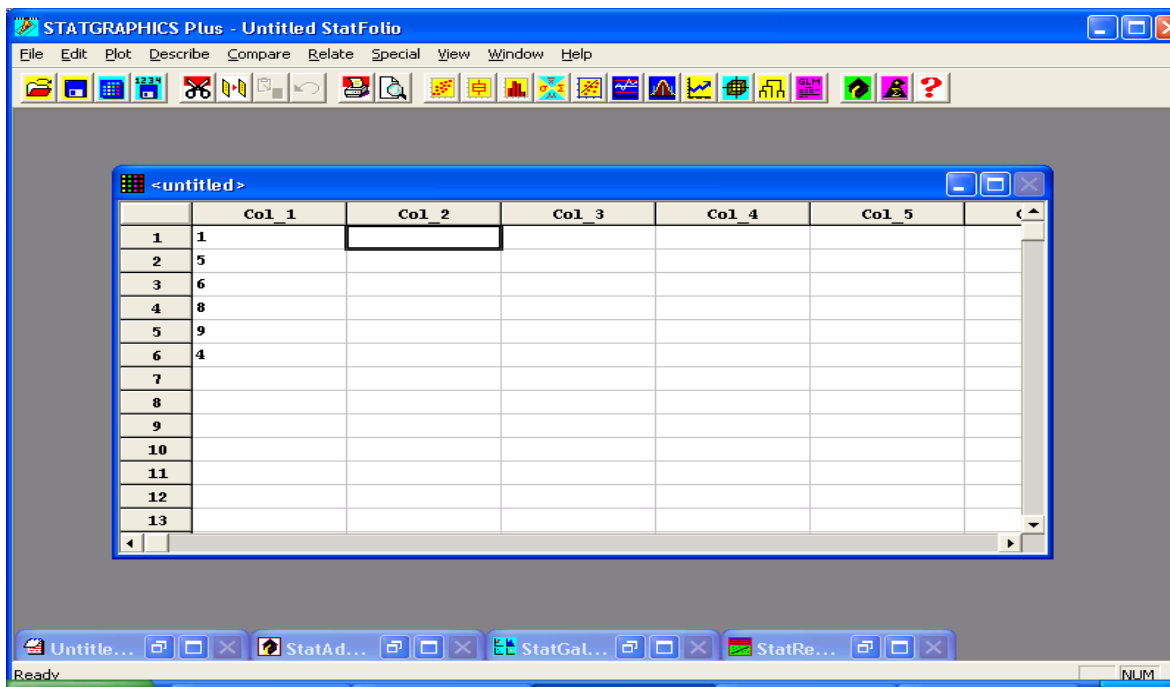
**508: ESTADÍSTICA EN DISEÑO INDUSTRIAL.
PRÁCTICAS DE STATGRAPHICS. CURSO 2005/06
PRÁCTICA 1**

INTRODUCCION AL MANEJO DEL STATGRAPHICS

PROCEDIMIENTOS GENERALES

La versión Windows del Statgraphics muestra una ventana (Statfolio) con una barra de opciones dentro de las cuales encontramos varios submenús agrupados en bloques temáticos:

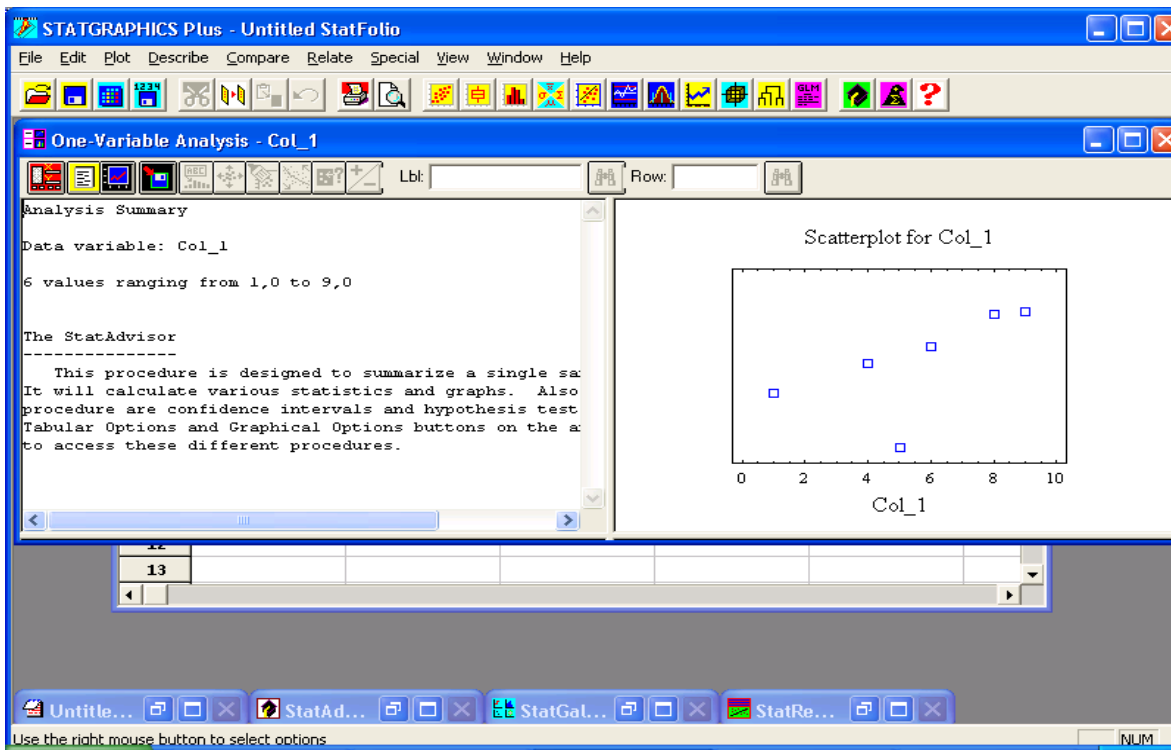
- File**: Permite la creación y manipulación de archivos, entre las opciones destaca: Open (Abre), Close (Cierra) Statfolios, un archivo de datos o una Statgallery, Save, etc.
- Edit**: Permite labores relacionadas con la edición de archivos.
- Plot**: Permite acceder a los diferentes tipos de gráficos disponibles en el programa.
- Describe**: Permite un análisis descriptivo de los datos.
- Compare**: Permite acceder a tipos de análisis que generan y comparan estadísticos descriptivos.
- Relate**: Permite acceder a análisis que modelizan la relación entre variables dependientes e independientes.
- Special**: Módulos más especializados y relacionados con técnicas estadísticas avanzadas.
- View**: Visualización de opciones de trabajo.
- Window**: Visualización de ventanas de trabajo.
- Help**: Ayuda.



USO DE DIFERENTES TIPOS DE VENTANAS

Al entrar por primera vez en el programa os encontraréis con diversos tipos de ventana. Nosotros básicamente utilizaremos sólo:

- Ventana de cálculo (*tipo hoja de cálculo de Excel*). Aparece cuando se selecciona “untitled” (si no le hemos dado otro nombre). Esta ventana se usa para crear nuevas hojas de cálculo, modificar las ya existentes o modificar datos usando el editor. La usaremos para crear nuestros archivos de datos.
- Ventana de análisis (Analysis window). Aparece después de que hallamos seleccionado un gráfico o un análisis estadístico de la barra de menús. La ventana de análisis contiene varios iconos, que por orden de aparición serían:



- Input dialog (en rojo) : muestra el icono para el análisis (por si desea modificarse)
- Tabular options (libreta sobre fondo amarillo): muestra las distintas opciones de cálculo
- Graphical options (gráfica azul): muestra los distintos gráficos disponibles para representar
- Save results (disket azul): permite guardar resultados en la hoja de cálculo

Además, con el botón derecho del ratón (pulsando en el panel que queramos modificar) podremos acceder a diversas opciones (*Pane options* o *Análisis options*, *Graphics options*), que nos permitirán cambiar los valores por defecto. Los análisis (gráficos y tablas) los podremos pegar en el Word para hacer la memoria.

Pulsando dos veces con el botón izquierdo del ratón en el panel que queramos, conseguiremos ampliarlo o bien dejarlo en su tamaño original.

CREACIÓN Y MANEJO DE FICHEROS DE DATOS

Creación de un fichero y manejo de las variables contenidas en él.

La manera más cómoda de trabajar con Statgraphics es tener los datos guardados en ficheros de datos. Para crear un archivo de datos:

1. Seleccionamos en la barra de tareas principal dentro de la opción **Window** el archivo **untitled**.

2. Seleccionamos con el botón izquierdo del ratón **Col_1**, el nombre por defecto de la variable.

3. Con el botón de la derecha del ratón seleccionamos dentro del menú que aparece la opción **Modify Column** y rellenamos los campos que aparecen:

- Escribimos el nombre de la variable (Name).
- Podemos añadir comentarios sobre la variable (Comment).
- Especificamos la anchura que queremos para la columna (Width).
- Especificamos el tipo de variable:
 - Numeric: variables numéricas (tantos decimales como queramos).
 - Character: variables cualitativas o números tratados como valores cualitativos.
 - Integer: Enteros
 - Fixed Decimal: números con una cantidad de decimales prefijada.
 - Formula: permite generar variables mediante formulas a partir de otras existentes.

- Volvemos a la columna de la variable que queremos generar y vamos introduciendo los valores.

Repetimos el proceso para cada una de las variables que queramos introducir y para guardar los datos seleccionamos en la opción **FILE** el submenú **SAVE AS** y **SAVE DATA FILE AS**.

Manejo de Ficheros. File Utilities: Las operaciones con ficheros de datos las podemos hacer con la opción **FILE**, como ya se ha dicho.

Guardar los resultados y recuperarlos: Se hace fácilmente con el menú **FILE** (con **OPEN**, **SAVE** y **PRINT**), como ya se dijo al principio.

Nota: en la versión 4 del Statgraphics, al comenzar la sesión nos preguntará el tipo de análisis y datos que deseamos tratar. Para comenzar en nuestra primera sesión, le daremos a *Cancel* y *Sí* (que queremos salir del Statwizard).

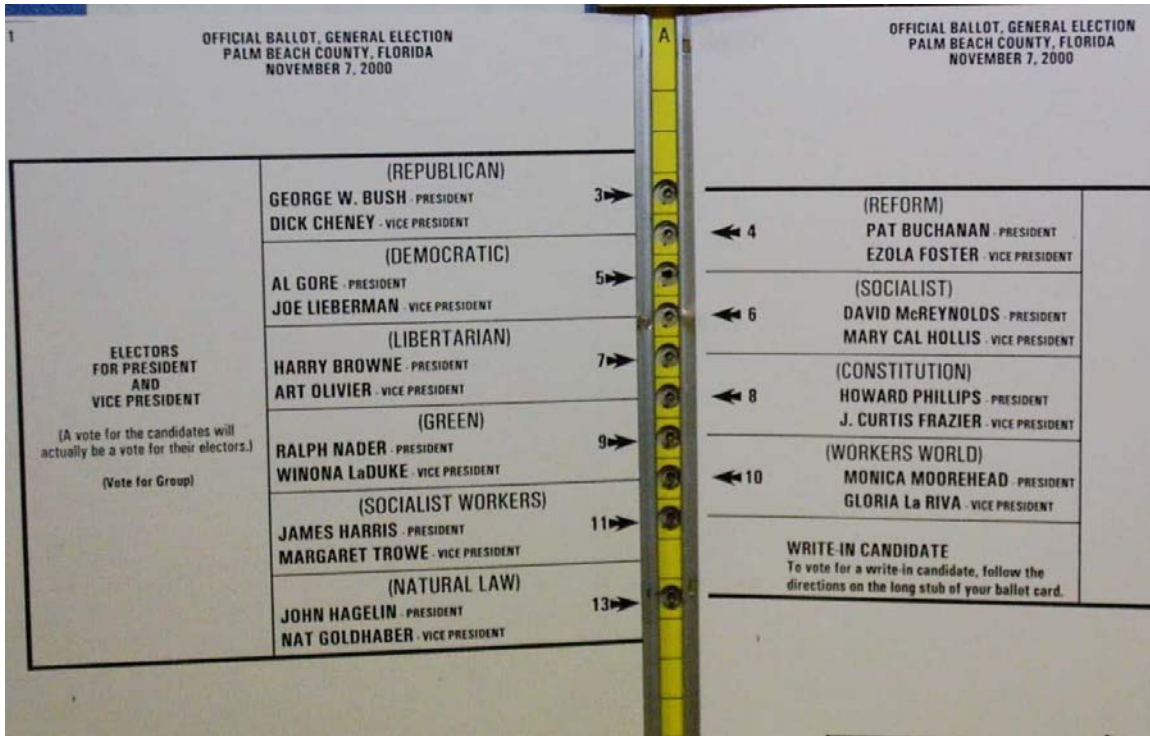
PRÁCTICA 1: LAS GRAVÍSIMAS CONSECUENCIAS QUE PUEDE TENER UN MAL DISEÑO

Esta primera sesión será completamente guiada pues en ella introduciremos la metodología del Statgraphics, para familiarizarnos con ella, a través de un ejemplo sencillo, pero desgraciadamente real: ¿El diseño de la papeleta electoral costó a Al Gore las elecciones presidenciales del 2000?

En las elecciones presidenciales del 2000 en los Estados Unidos, el candidato democrático era Al Gore, mientras que el candidato republicano era George W. Bush. En el condado de Palm Beach, en Florida, en las elecciones iniciales, 3407 votos fueron para el candidato del partido de la Reforma (Reform party), Pat Buchanan. Algunos analistas políticos pensaron que este total parecía asombrosamente grande. Creían que la mayoría de estos votos iban realmente para Al Gore (cuyo nombre estaba al lado del Buchanan en la papeleta), pero erróneamente señalaron a Buchanan debido al diseño de la "papeleta--mariposa" que se utilizó en ese condado, y con el que algunos votantes encontraron confusión.

La papeleta-mariposa (butterfly ballot), que se encuentra seguidamente, fue una innovación de Theresa LePore, Supervisora de Elecciones en el condado de Palm Beach. Este formato fue únicamente usado en este condado y en esta ocasión. Se le denomina "mariposa" porque tiene dos columnas de nombres de candidatos (las alas de la mariposa), con una columna central con los agujeros a perforar. Estos agujeros corresponden alternativamente a los candidatos de la parte izquierda y derecha de la papeleta. De esta forma, el primer agujero corresponde a Bush, y el segundo a Buchanan, el primer candidato de la parte derecha. El segundo candidato de la parte izquierda era Gore, al que le corresponde el tercer agujero. Si se lee sólo la parte izquierda, sin fijarse en la parte derecha de la papeleta, se podría erróneamente concluir que el segundo agujero correspondía a Gore, cuando en realidad era para Buchanan.

En la página <http://www3.uji.es/~epifanio> en el curso correspondiente, y para esta sesión, se encuentra un fichero, obtenido de <http://www.stat.unc.edu/faculty/rs/palmbeach.html>. Guarda este fichero y ábrelo: (**File>Open>Open data file**, Tipo all files *.* , *blank delimited*).



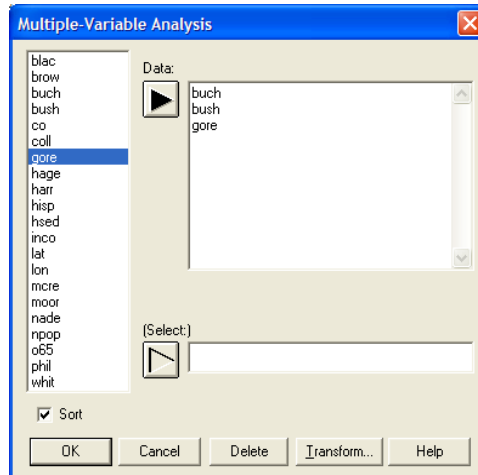
En este fichero encontramos los siguientes datos, resaltados en negrita los que usaremos:

<i>co</i>	County Number (1-67)
<i>lat</i>	Latitude (degrees north)
<i>lon</i>	Longitude (degress west)
<i>npop</i>	Population (1997)
<i>whit</i>	Percentage of whites (1996)
<i>blac</i>	Percentage of blacks (1996)
<i>hisp</i>	Percentage of Hispanics (includes other races) (1996)
<i>o65</i>	Percentage of population over 65 years (actually this is the 1996 population of those age 65 and over, expressed as a percentage of the 1997 total population)
<i>hshed</i>	Percentage completing high school education (1990)
<i>coll</i>	Percentage completing college education (1990)
<i>inco</i>	Mean income per person (1994)
<i>bush</i>	Votes for Bush / Cheney (REP)
<i>gore</i>	Votes for Gore / Lieberman (DEM)
<i>brow</i>	Votes for Browne / Olivier (LIB)
<i>nade</i>	Votes for Nader / LaDuke (GRE)
<i>harr</i>	Votes for Harris / Trowe (SWP)
<i>hage</i>	Votes for Hagelin / Goldhaber (LAW)
<i>buch</i>	Votes for Buchanan / Foster (REF)
<i>mcre</i>	Votes for McReynolds / Hollis (SPF)
<i>phil</i>	Votes for Phillips / Frazier (CPF)
<i>moor</i>	Votes for Moorehead / La Riva (WWP)

El nombre de los condados de Florida junto con sus números son:

1 Alachua	18 Franklin	35 Lee	52 Pinellas
2 Baker	19 Gadsden	36 Leon	53 Polk
3 Bay	20 Gilchrist	37 Levy	54 Putnam
4 Bradford	21 Glades	38 Liberty	55 Santa Rosa
5 Brevard	22 Gulf	39 Madison	56 Sarasota
6 Broward	23 Hamilton	40 Manatee	57 Seminole
7 Calhoun	24 Hardee	41 Marion	58 St. Johns
8 Charlotte	25 Hendry	42 Martin	59 St. Lucie
9 Citrus	26 Hernando	43 Miami-Dade	60 Sumter
10 Clay	27 Highlands	44 Monroe	61 Suwannee
11 Collier	28 Hillsborough	45 Nassau	62 Taylor
12 Columbia	29 Holmes	46 Okaloosa	63 Union
13 Desoto	30 Indian River	47 Okeechobee	64 Volusia
14 Dixie	31 Jackson	48 Orange	65 Wakulla
15 Duval	32 Jefferson	49 Osceola	66 Walton
16 Escambia	33 Lafayette	50 Palm Beach	67 Washington
17 Flagler	34 Lake	51 Pasco	

Para ir calentando motores, obtengamos cuántos (*sum*) votaron a Buchanan, Bush y Gore en el condado de Florida, y el número medio de votos (*average*) por condado para estos tres candidatos. Podemos usar: **Describe>Numeric data>Multiple variable análisis** e incluir las variables buch (número de votos para Buchanan), bush (número de votos para Bush), gore (número de votos para Gore).



En la libreta sobre fondo amarillo (Tabular options), señalaremos *Summary statistics*. Si no aparecen los resultados que nos interesan, con el botón derecho del ratón, en *pane options*, podemos seleccionarlos. Incluye los resultados en la memoria.

Vamos a representar conjuntamente los datos para las variables gore (número de votos para Gore) y buch (número de votos para Buchanan), usando **Plot>scatterplot>X-Y plot**. En la X pondremos gore, y en la Y buch, cada punto corresponde a un condado. Incluye este dibujo en la memoria.

¿Te parece que haya algún punto inusual en este gráfico?

Para identificarlo, puedes pinchar sobre él con el ratón sostenidamente. ¿A qué condado corresponde?

Repite los ejercicios anteriores pero usando en la X la variable bush (número de votos para Bush), para ello puedes usar el botón rojo del *input dialog*.

Estudios más elaborados que éstos, llevados a cabo por estadísticos, predijeron que menos de 900 votos iban realmente para Buchanan en Palm Beach, comparados con los 3407 que recibió. Bush ganó el estado sólo por menos de mil votos (y con él las elecciones), así que los votos en Palm Beach, con el diseño mariposa de la papeleta, fue un factor crítico en el resultado de esas elecciones. Otros factores que también habrían jugado un papel relevante fueron las 110.000 papeletas descalificadas por “sobrevoto”, en las cuales la gente votó equivocadamente por más de un candidato presidencial (con Gore marcado en 84.197 y Bush en 37.731), a menudo debido a la confusión de los nombres que eran enumerados en más de una página de la papeleta, y también serían importantes las 61.000 papeletas rechazadas por “infravoto” debido en algunos condados a las máquinas perforadoras manuales.

Ahora la ley de Florida ha cambiado, no usándose más en futuras elecciones tarjetas perforadas sino electrónicas. En cualquier caso, como Richard L. Smith señala en su artículo “STATISTICAL ASSESSMENT OF BUCHANAN’S VOTE”: <<Although there were many other contentious issues in this election—the failure of machines to record votes correctly, the alleged harassment of racial minority voters, the disputed legality of some of the absentee ballots that were counted and the whole question of recounting the ballots by hand—most commentators agree that there is no proof that any of these factors were responsible for Bush winning the election. The most convincing evidence that Gore should have won the election is based on the outcome in Palm Beach county>>.