

BOALOX INFORMATICA

Genl. Franco, 87

Tlf. (988)221647

ORENSE - 32003

S T A T

PAQUETE DE PROGRAMAS PARA CALCULO ESTADISTICO

I.- CONTENIDO

Es un conjunto de programas destinado a cubrir la mayor parte del trabajo de tratamiento estadístico de datos, desde las aplicaciones más sencillas (tratamiento de pequeños conjuntos de datos) hasta programas que ofrecen un grado muy elevado de automatización en el tratamiento de la información numérica. Ofrece amplias posibilidades gráficas (histogramas en tres dimensiones, gráficos de nube de puntos, representación gráfica de puntuaciones) tanto en pantalla como por impresora.

El núcleo del paquete de programas lo constituyen los que están dedicados a tratar datos pertenecientes a dos dimensiones (dos variables) distintas, bien sea tomados aisladamente o a partir de una matriz de datos que contiene más de dos variables (es decir, más de dos valores para cada sujeto o caso). Estos programas satisfacen la mayor parte de las necesidades habituales de cálculo.

En programas aparte, se incluyen también otros instrumentos de cálculo de uso menos frecuente (cálculo de diferencias entre medias, cálculo de puntuaciones típicas y percentiles, su representación gráfica, correlación parcial, correlación múltiple).

II.- USO DE LOS PROGRAMAS

1.-Programas generales de cálculo

Cuando se dispone de datos que deben ser sometidos a tratamiento estadístico (puntuaciones de alumnos en una asignatura, o número de días

que cada enfermo tardó en curar) la primera tarea suele consistir en obtener los parámetros fundamentales de la población (valores máximo y mínimo, media aritmética, medida de la dispersión en torno a la media -desviación típica- etc).

Los programas del paquete STAT están configurados para tratar, según es habitual, los datos por parejas: por ejemplo, para cada alumno del grupo, puntuación en matemáticas y puntuación en idioma; o, en un grupo de enfermos, días que cada paciente tardó en curar y dosis de medicación recibida. (Si en algún caso excepcional el usuario dispone de una sola serie de datos, deberá introducir como segundo dato un número ficticio -por ejemplo, número de orden en la lista, o número de expediente- y despreciar después los resultados correspondientes a esa variable ficticia y los que se refieren a la relación entre las dos variables).

El problema crucial en el tratamiento estadístico de la información mediante ordenador suele radicar en el modo de introducción de los datos. Por ello, los programas contenidos en STAT están diversificados inicialmente según el modo de introducir los datos en el ordenador.

Si Vd. desea tratar estadísticamente sólo los datos de unos pocos sujetos, y cada sujeto tiene pocos datos (por ejemplo, quiere Vd. ver la relación entre edad y número de días de enfermedad, en una muestra de 15 pacientes), la solución más rápida consiste en introducir los datos directamente a través del teclado, y STAT proporciona esta posibilidad. Para evitar tener que teclear de nuevo estos datos, le da opción a grabarlos en cartucho, y así podrá usarlos de nuevo, añadir nuevos cálculos, nuevos casos, etc.

Pero si estos datos son de muchos sujetos, o si cada sujeto tiene muchos datos, entonces no resulta cómodo introducir cada vez los datos por el teclado, porque, como lo que suele pretenderse es relacionar cada serie de datos con todas las demás, sería preciso teclear varias veces cada una

de las series de datos. Para estos casos, se parte de un modo de entrada de los datos que evite la necesidad de escribirlos dos veces, y que ofrece indudables ventajas. Consiste en la creación de programas independientes del principal, en que la información está almacenada en líneas DATA, fáciles de comprobar y de corregir, y que, además, permiten introducir -mediante un programa auxiliar- nuevas variables derivadas (ratios), eliminar variables, etc.

2.- Programas para cálculos complementarios

Una vez obtenidos los principales datos estadísticos de una muestra mediante la aplicación de esos programas generales a que se acaba de aludir, el usuario puede querer efectuar nuevos cálculos a partir de los resultados iniciales. En un programa independiente ("COMPLEM") se han reunido algunas herramientas estadísticas de uso frecuente:

- Cálculo de la significación de diferencias entre medias de muestras independientes.

- Cálculo de la significación de diferencias entre medias de muestras relacionadas.

- Correlación parcial $r_{12.3}$ (correlación entre dos variables, excluyendo los efectos de una tercera).

- Correlación parcial $r_{12.34}$ (excluyendo los datos de otras dos variables).

- Correlación múltiple con tres variables.

- Correlación múltiple con cinco variables.

- Cálculo de puntuaciones típicas.

- Cálculo de percentiles.

- Representación gráfica de puntuaciones típicas y percentiles.

El programa está dirigido mediante un menú

--- X ---

Para adquirir soltura en el manejo del paquete de programas, conviene comenzar por familiarizarse con los de cálculo estadístico

general, cuyo modo de carga se indica a continuación.

III.- CARGA DE LOS PROGRAMAS

1.- Programas de cálculo estadístico general

Proceda del modo siguiente:

a).- Con la impresora -si dispone de ella- ya conectada, y sin ningún cartucho colocado en los microdrives, pulse el botón de "reset" (en el lateral derecho del aparato).

b).- Si su copia de STAT viene en cartucho de microdrive, coloque en el cartucho original -o la copia de seguridad que haya hecho- en mdv1 (el de más a la izquierda), y pulse la tecla F2 (a la izquierda del teclado). Esto hará que se cargue un programa-llave, que, según la alternativas que Vd. elija respecto a la forma de introducir los datos, cargará los programas que sean precisos.- Si su copia de STAT viene en diskette, colóquelo en la unidad de discos número 1 y teclee "lrun fdk1_boot" (ENTER). Esto carga el programa-llave.

Si el programa se carga sin hacer previamente "reset", conviene asegurarse de que otros programas anteriores no hayan modificado variables del sistema o hayan efectuado otras operaciones que puedan interferir con STAT (por ejemplo, cargando bytes en código máquina que reduzcan la memoria libre, o cerrando canales que STAT necesita para funcionar, etc.).

2.- Programas para cálculos complementarios

a).- Con la impresora -si la tiene- ya conectada, pulse "reset" y F2 (modo TV).

b).- Coloque en mdv1 la copia de STAT (o, si la copia está en diskette, colóquela en el drive 1).

c).- Teclee "LRUN MDV1_COMPLEM" (o LRUN FDK1_COMPLEM para la versión en diskette).

Como en el caso anterior, si se carga el programa sin haber hecho "reset" hay que asegurarse de que la memoria viva del QL no está alterada debido al uso de otros programas.

IV.- EMPLEO DE LOS PROGRAMAS DE CALCULO ESTADISTICO GENERAL

Una vez cargado el programa-llave, se le ofrece la opción de introducir los datos a través del teclado, a través de un fichero de microdrive o a través de un programa DATA especialmente confeccionado.

Cada una de estas tres alternativas tiene su propio campo de aplicación:

1).- La forma de introducción de datos tecleándolos directamente debe ser elegida cuando se pretende tratar estadísticamente unos cuantos datos, (por ejemplo, un máximo de 50, por señalar una cifra indicativa). Lo habitual es que disponga Vd. de dos datos por cada sujeto o caso, por ejemplo, edad y estatura de cada uno de los componentes de un grupo de 15 alumnos, o días que tardó en curar cada paciente de un grupo de 32 pacientes y miligramos de medicación recibida en el período que duró la enfermedad.- Si tiene Vd. más de dos datos por sujeto o caso; y quiere incluirlos en los cálculos a fin de conocer la inter-relación entre esas diversas (tres o más) series de datos, es mejor que utilice la opción 3 (uso de un programa DATA especialmente confeccionado) que le permitirá evitar repeticiones innecesarias al teclear los datos. Por el contrario, si tiene Vd. solamente un dato por cada sujeto o caso y desea tratar estadísticamente esa información para conocer algunos datos sobre ella (la media y la desviación típica o medida de la dispersión), es necesario que introduzca Vd., además, para cada caso, un número ficticio (que bien puede ser el número de orden que el dato ocupa en la serie: 1,2,3,...; o el número de expediente, o un número al azar).

Supongamos que, como Vd. no dispone en este momento de datos reales, decide copiar los que aparecen en un ejemplo tomado de un libro de estadística:

Altura (X) del padre (pulg.) 65 63 67 64 68 62 70 66 68 67
69 71

Altura (Y) del hijo (pulg.) 68 66 68 65 69 66 68 65 71 67
68 70

Para tratar estadísticamente esos datos, elija la opción 1 "1.-Va a introducir los datos a través del teclado?").

El programa-llave carga entonces el programa necesario para utilizar este modo de introducción de los datos.

Cuando el programa haya terminado de cargarse, el ordenador le pedirá el nombre de la muestra. Supongamos que esos 12 pares de datos de estaturas de padres e hijos corresponden a la sección B del curso de COU; un nombre adecuado sería, por ejemplo. "COU-SECCION B". Se permite un máximo de 20 caracteres.

El ordenador le pedirá también que indique el nombre de los datos situados en primer lugar en cada par. Debe usted teclear "Estatura padres". Esta variable recibe el nombre de variable independiente, porque se supone que es la que mejor se conoce y no interesa predecirla.

(Una de las finalidades de la estadística es predecir; si sabemos que a padres más altos corresponden hijos más altos y si conocemos la expresión matemática de esa inter-correlación en un número suficiente de casos, podríamos llegar a predecir para un recién nacido -sin necesidad de esperar a que crezca- qué estatura alcanzará cuando sea adulto, sólo con conocer la estatura de sus padres. En cambio, la predicción inversa -predecir la estatura de los padres conociendo la de los hijos- carece, por lo general, de interés práctico).

A continuación, teclee el nombre de la variable situada en segundo lugar en cada par (es decir, la variable dependiente o variable que interesa predecir): "Estatura hijos".

Cuente el número de pares de datos que va a tratar (en el caso del ejemplo, 12 pares), y tecléelo cuando el ordenador lo solicite mediante la

petición "Escriba el número de pares de datos que va a introducir (y ENTER)".

El ordenador le indica que ya puede teclear los datos, seguido, cada uno de ellos, de ENTER.

Para facilitar la entrada de datos la pantalla cambia de color según que se trate del primer dato o del segundo, y en todo momento se indica cuál es el número de orden del dato que se está introduciendo.

Si se equivoca al teclear algún dato, no se preocupe; tendrá ocasión de corregirlo más tarde. Límitese a tomar nota de cuál es el dato erróneo (en qué par está, y si es el primer dato de ese par o el segundo).

Cuando termine de introducir los datos, el ordenador le ofrece un listado por pantalla, que conviene que Vd. confronte con sus datos.

Si todo está bien, elija la opción 1 ("SEGUIR"), o la 2 ("IMPRIMIR"), en el caso de que tenga conectada la impresora y desee una copia de los datos en papel.

Una advertencia importante: no elija nunca opciones de impresión o copiado de pantalla si no tiene conectada la impresora; el QL podría caer en un bloqueo del que a veces sólo se sale retirando los cartuchos de los microdrives y haciendo "reset" -con lo que, naturalmente, se pierde todo el contenido de la memoria-. De todos modos, intente primero detener el programa pulsando CTRL y la barra espaciadora.

En cambio, si encuentra algún error en los datos, elija la opción 3 ("CORREGIR"), que le permitirá introducir las modificaciones necesarias; obtendrá Vd. siempre, después de efectuarlas, un nuevo listado de comprobación con la mismas opciones de seguir, imprimir el listado o modificarlo de nuevo.

Cuando Vd. manifieste su conformidad con los datos que tiene en memoria (eligiendo la opción 1, "SEGUIR") el programa le ofrece la posibilidad de grabar los datos en microdrive (o en diskette). Para eso debe Vd. teclear el nombre completo del fichero que quiere crear para contener esos datos,

es decir, incluyendo la mención del periférico donde ha de hacerse la grabación (ejemplos: mdv1_datos, mdv2_cou, fdk1_griposos etc). Esta opción es útil cuando son bastantes datos, cuando se tiene la sospecha de que habrá de modificar en el futuro algún dato, cuando se prevé que será necesario usar de nuevo esos datos para otros cálculos, o ampliarlos añadiéndoles otros nuevos. Cuando grabe los datos, tome nota de cuál es el nombre del fichero bajo el cual se hizo la grabación. (De todos modos siempre podrá comprobar el nombre del fichero mediante el comando "DIR" usado desde Superbasic).

El ordenador se dispone ya a someter a tratamiento estadístico los datos; pero antes ofrece al usuario la posibilidad de sacar los resultados del cálculo a través de la pantalla o a través de la impresora.

Los resultados del cálculo incluyen el nombre de la variable, los casos incluidos en el cálculo, la media aritmética, los intervalos de confianza de la media al 5% y al 1%, la desviación típica y la suma de valores, así como el valor máximo y el mínimo. Estos valores se calculan para cada una de las dos variables. Se calcula también el coeficiente de correlación de Pearson entre ambas variables, así como si resulta o no significativo los niveles de confianza del 5% y del 1%.

Al final, en la última línea, se ofrecen también los resultados intermedios que han servido para calcular los estadísticos (suma de valores de cada variable, id. al cuadrado, etc.). Estos valores se utilizan en la mayor parte de las fórmulas estadísticas, por lo que pueden servir para aplicar otras fórmulas distintas.

- 1.- CALCULO DE INTERVALOS E HISTOGRAMA DE LA VARIABLE INDEPENDIENTE
- 2.- ID ID DE LA VARIABLE DEPENDIENTE
- 3.- SOLO GRAFICO DE NUBE DE PUNTOS
- 4.- VOLVER AL LISTADO DE DATOS DIRECTOS
- 5.- VOLVER A LOS RESULTADOS
- 6.- TRATAR OTRO PAR DE VARIABLES

Ya se elija un gráfico de uno u otro tipo, el programa proporciona la posibilidad de modificar los límites de la distribución (por ejemplo, para facilitar la comparación entre gráficos de diversas muestras o para estandarizar la presentación de resultados de una serie de muestras). Como paso previo al trazado de histogramas, el programa efectúa un agrupamiento de los datos en intervalos. El número de intervalos puede ser elegido por el usuario, aunque el ordenador sugiere el número óptimo de intervalos según el tamaño de la muestra (tabla de Kelly). La distribución de frecuencias contiene a veces intervalos en que no coincide ningún valor entero, con lo que la representación gráfica saldrá distorsionada. Usando la opción que permite variar el número de intervalos y la que permite modificar los extremos de la distribución, puede eliminarse esta fuente de error en la representación gráfica.

Si se ha elegido la opción "histograma" de alguna de las dos variables, se puede escoger entre un histograma con efecto de tres dimensiones, que se representa en pantalla y puede ser copiado en papel, o un tipo de histograma que se imprime directamente en la pantalla.

El histograma en tres dimensiones se copia en papel mediante el programa copiador de EASEL (`gprint-prt`); este programa funciona con impresoras tipo EPSON (Newprint Admate, BMC, que es la que comercializa Investrónica con el QL). Si su impresora es capaz de copiar gráficos de EASEL también será capaz de copiar histogramas tridimensionales en STAT.

El otro tipo de histograma está basado en las posibilidades gráficas de las impresoras matriciales antes indicadas. No será difícil efectuar la adaptación para cualquier otra clase de impresoras gráficas.

El gráfico de nube de puntos puede obtenerse también bajo dos variedades: en pantalla (con posibilidad de copiado en papel mediante el programa copiador de EASEL, o directamente en pantalla utilizando las posibilidades de las impresoras matriciales tipo EPSON (Admate Newprint BMC y similares).

2.- La forma de introducir los datos desde un fichero de microdrive (o de diskette) exige que en el menú inicial (inmediatamente después de cargar el programa) se elija la opción 2 ("Introducir los datos a partir de un fichero de microdrive").

Naturalmente, este fichero de datos tiene que haber sido creado previamente, mediante la opción anterior ("Introducción de datos a través de teclado"), efectuando la grabación en microdrive cuando el programa ofrece esa posibilidad.

El uso de este procedimiento de entrada de datos exige que Vd. recuerde el nombre bajo el cual hizo la grabación; antes de ejecutar el programa ha de tenerlo localizado, así como el cartucho o diskette en que está grabado.

A partir de este momento inicial de la entrada de datos, el tratamiento estadístico es igual que en el caso del apartado anterior ("Introducción de los datos a través del teclado").

Si Vd. introduce correcciones en los datos y quiere hacer una nueva grabación de los mismos, no olvide que en un mismo cartucho o diskette no puede haber dos ficheros con nombres iguales. Para obviar este inconveniente, puede Vd. modificar el nombre de la segunda grabación, o hacerla en otro cartucho distinto.

3.- La forma de introducción de los datos a través de un programa de líneas DATA (opción 3 del menú) exige que los datos se hayan pasado con anterioridad a líneas DATA.

Como ya hemos dicho, esta modalidad de introducción de la información numérica es útil cuando se dispone de datos de bastantes sujetos y se tienen varias observaciones o datos por cada sujeto.

Esta es realmente la fórmula de introducción de datos que es preciso usar para tratamiento de grandes masas de información numérica; evita la necesidad de teclear varias veces las mismas cifras, y permite el análisis estadístico "en serie" de extensas matrices de datos. Esta forma de

disponer los datos hace que puedan ser fácilmente visualizados, revisados, modificados o ampliados, pues para ello sólo es preciso usar los comandos más conocidos del lenguaje BASIC (LIST, EDIT, RENUM, MERGE, etc).

Para utilizar con sus propios datos esta modalidad del programa debe Ud. construir un programa DATA. Para construirlo es conveniente no tener en el ordenador ningún otro programa. Es preciso situar en una línea DATA bajo el número 999 el nombre de la muestra, entrecomillado (máximo, 20 caracteres) y en la línea 1000 y siguientes, con intervalos de 10 entre ellas, los DATA correspondientes a los nombres de las variables. A continuación, desde la línea 2000 y siguientes, también espaciadas de 10 en 10, los datos correspondientes a cada sujeto, siempre en el mismo orden que las variables de las líneas 1000 y siguientes, y separados entre sí por comas. Si falta algún dato, se indica con "-" (signo menos entrecomillado); es conveniente poder, entrecomillado, al final la línea DATA de cada caso o sujeto, el nombre o dato de identificación del caso. Después del último dato numérico de la línea 2000 es necesario intercalar el signo "+" (el signo de sumar, entrecomillado). Una vez que se ha acabado de introducir datos, se escribe una última línea de DATA que sólo contenga la orden DATA y la contraseña "99999", que indica al ordenador que no debe seguir leyendo datos.

El programa de DATA ha de quedar así:

```
999 DATA "GRIPE PABELL 3 SULFM": REM se refiere esta muestra a
pacientes hospitalizados en el III pabellón del centro sanitario, que han
sido diagnosticados de gripe y tratados con sulfamidas.
```

```
1000 DATA "EDAD"
```

```
1010 DATA "DIAS HOSPITALIZADO"
```

```
1020 DATA "MEDICACION (MG/KG)"
```

```
1030 DATA "PRESION ARTERIAL INIC"
```

```
1040 DATA "PESO"
```

```
1050 DATA "TEMPER. MAX."
```

```
1060 DATA "INDICE UREMIA"  
2000 DATA 28,5,7,16,75,41.5,7,"+","LEUCEMIO LOPEZ PEREZ"  
2010 DATA 75,16,12,22,55,42,9,"DETERIORO GOMEZ GARCIA"  
2020 DATA 46,2,1,12,70,38,1,"ROBUSTIANO VAZQUEZ FERNANDEZ"  
  
.  
.  
.  
  
2090 DATA "99999"
```

Los nombres de los sujetos que figuran al final de cada línea (o las referencias de cada caso concreto) pueden ser eliminados.

Una vez confeccionado el programa DATA conteniendo sus propios datos debe Vd. grabarlo en un cartucho (o diskette) independiente, con la orden habitual "save mdv1_nombre" (donde "nombre" es la denominación que Vd. quiera darle a la grabación, como "misdatos" -máximo ocho letras-); y tome nota del nombre con que hace la grabación para recuperar después los datos cuando haga funcionar el programa STAT.

Para usar esta modalidad del programa, elija en el menú inicial la opción 3 (los datos han de leerlos el ordenador de un programa DATA); el programa le recuerda entonces que es necesario haber formado antes un programa DATA, y resume brevemente sus características.

Para hacer funcionar el programa, los datos (el programa DATA) han de estar integrado dentro del programa STAT, ocupando las líneas 999 y siguientes; la pantalla ofrece en este momento la posibilidad de incorporar ese programa DATA, o de seguir adelante si ya está incorporado. Como quiera que esta incorporación se efectúa mediante la orden MERGE, de ejecución bastante lenta, es probable que la operación tarde algunos minutos si sus datos son algo copiosos.

A continuación, el programa se diversifica: puede Vd. decidir que la matriz de datos sea tratada de modo automático, calculando los estadísticos de cada variable y los coeficientes de correlación y valores

de la recta de regresión correspondientes a cada par de datos; o bien tratar los datos eligiendo las series dos a dos, y con posibilidad de acceder a las posibilidades gráficas (histogramas y nube de puntos). Aludiremos a continuación a cada una de estas dos opciones:

A).- Cálculo seriado de los principales estadísticos de cada variable y coeficientes de correlación de la matriz de datos.

El programa que se carga al elegir esta opción es el más breve de todos, por lo que la capacidad utilizable para la matriz de datos es mayor que si se utiliza la otra opción.

Dada la gran cantidad de resultados numéricos que se producen, esta opción está pensada para ser utilizada imprimiendo en papel los resultados. No obstante, se ha previsto una salida de los resultados a través de la pantalla para el caso de que solo interese copiar algunos de ellos.

Es necesario poner especial cuidado en colocar el papel de la impresora al comienzo de una nueva hoja, a fin de que los resultados se impriman en forma presentable.

Dada la gran cantidad de cálculos que es preciso realizar para analizar sistemáticamente una matriz de datos de moderadas dimensiones, el programa suele tardar bastante en completar los resultados. (Para una matriz de 100 casos con 20 observaciones cada uno, el tiempo total de trabajo puede ser de unas 8 ó 9 horas, imprimiéndose aproximadamente cada minuto un resultado de tratar un par de variables).

Si el número de variables es menor de 13, el programa imprime al final la matriz de correlaciones.

B).- Si se elige la modalidad de cálculo seleccionando manualmente las variables a tratar, se ofrece primero la posibilidad de cargar el programa copiador de pantalla de EASEL, y, a continuación se indica el número máximo de casos que es posible tratar; el usuario puede decidir tratar menos casos que los contenidos en el programa DATA.

(Otra forma de limitar el número de datos a tratar es colocando en el programa DATA la cifra "99999" al principio de la línea siguiente a aquella que se desea tratar en último lugar).

Las variables elegidas pueden estar en cualquier orden: p. ej., la séptima con la segunda, la tercera con la quinta, etc). La elegida en primer lugar es tomada como variable independiente a efectos de cálculo de los valores de la recta de regresión y de gráfico de nube de puntos.

El cálculo y las posibilidades gráficas son iguales que en las modalidades de introducción de datos a través del teclado y de ficheros de microdrive. El menú de gráficos incluye la posibilidad de tratar otro par de variables de la misma matriz sin tener que volver a cargar el programa.

V.- Programa COMPLEM

Se han reunido en ese programa una serie de subrutinas de cálculo de uso frecuente en estadística:

- Significación de diferencias entre medias de muestras independientes: para usar este programa se parte habitualmente de los datos obtenidos mediante el programa STAT. El resultado es la llamada razón crítica; si resulta mayor de 1.96 se dice que la diferencia entre medias es significativa al nivel de confianza del 5%, y si es superior a 2.56, al del 1%.

- Significación de diferencias entre medias de muestras relacionadas: el modo de interpretación es similar al del caso anterior, pero se usa esta opción cuando las muestras no son independientes, sino que están relacionadas (p. e., parejas de hermanos).

- Correlación parcial $r_{12.3}$: dadas tres variables interrelacionadas permite conocer la verdadera relación entre dos de ellas, eliminando la influencia de una tercera.

- Correlación parcial $r_{12.34}$: similar a la del caso anterior, pero eliminando la influencia de otras dos variables.

- Correlación múltiple con tres variables: permite conocer la verdadera relación entre una variable y otras dos más, eliminando la influencia de la correlación de estas otras dos entre sí.

- Correlación múltiple con cinco variables: el procedimiento de introducción de datos y de cálculo es el método llamado de Wherry-Doolittle. Se elige la variable dependiente, y como independientes se escogen aquellas variables que correlacionen en más alto grado con la dependiente y menos entre sí. Antes de introducir los datos conviene formar un cuadro que contenga las correlaciones entre las variables tomadas en consideración, colocando en último lugar la variable dependiente. El programa proporciona a través de la impresora los valores de los coeficientes beta, los valores de la ecuación de regresión múltiple y el coeficiente de correlación múltiple.

El programa está dirigido a través de un menú principal.

VII.- USO DE IMPRESORAS

El programa está concebido de modo que las salidas se produzcan por impresora. Está preparado para ser usado con impresoras tipo EPSON, como la Newprint Admate o la BMC que se comercializa con el QL. Ambas funcionan a 9600 baudios. Si su impresora no funciona a pesar de estar correctamente conectada al QL a través del port SER1 y de una interface RS232 en la impresora, consulte el manual de ésta para comprobar el número de baudios. Si es distinto de 9600, cargue el programa "boot" colocando la copia de trabajo en mdv1 y escribiendo "load mdv1_boot"; edite la línea 32176 y cámbiela indicando el nuevo número de baudios; a continuación grabe la copia modificada, escribiendo: delete mdv1_boot:save mdv1_boot, (y pulse ENTER).

SIGNIFICADO DE LOS TERMINOS ESTADISTICOS

MEDIA ARITMETICA: es el estadístico que mejor representa al conjunto de la serie. Como se halla dividiendo la suma de los valores de la

observaciones por el número de éstas, no puede calcularse sobre números ordinales (por ejemplo, no podemos hallar la media entre el primero de la clase, el cuarto y el quinto porque no tiene sentido sumar los números ordinales).

La media viene dada en valores de la distribución (por ejemplo puntuaciones académicas, días que un paciente tardó en curar o número de comprimidos que ingirió); pero puede ocurrir que en ningún individuo se de realmente esa puntuación (ocurre por ejemplo, cuando decimos que la media de número de hijos en las familias españolas es de 3,7).

DESVIACION TIPICA: es una medida de dispersión. Dos conjuntos de datos pueden tener una media similar, y ser en realidad muy diferentes.

(Un país en que se coma un pollo por habitante y día -media- y en que los miembros de la clase alta coman un pollo y medio y los de la baja coman medio pollo, ofrecerá en esta variable una dispersión menor que otro país en que, disponiendo también de un pollo por habitante y día, cada individuo de la clase alta se como dos y los de la clase baja ninguno).

Cuanto mayor es la desviación típica, mayor es la dispersión de la distribución. Viene dada en unidades de medida de la distribución; y su cálculo (raíz cuadrada de la media de los cuadrados de las desviaciones respecto a la media) tampoco es posible respecto de números ordinales.

Si la forma de la distribución se aproxima a la de la curva normal de probabilidad (la llamada curva de Gauss, que tiene forma de campana), entre una desviación típica por encima de la media y otra por debajo de ella se encuentra aproximadamente el 68% de la población; entre dos d.t. positivas y dos d.t. negativas en torno a la media está el 95% de los casos de la distribución; y entre dos y media d.t. por encima y dos y media d.t. por debajo de la media está el 99% de los casos.

La d.t. permite comparar entre sí puntuaciones pertenecientes a distribuciones distintas.

COEFICIENTE DE CORRELACION DE PEARSON. Indica el grado en que dos

medidas varían concomitantemente. Puede oscilar entre 1 y -1; en ambos casos extremos, la correlación será perfecta: en el primero, positiva, en el segundo negativa. La ausencia de correlación viene indicada por el cero.

Ejemplos de correlación positiva serían la estatura y el peso (a mayor estatura el peso suele ser mayor); de correlación negativa, el nivel personal de renta y morbilidad -a menores ingresos económicos, mayor número de días de enfermedad al año-.

Un ejemplo de ausencia de correlación sería la falta de concomitancia que existe entre la estatura e inteligencia (no por ser más alto ni más bajo se es más o menos inteligente).

NIVEL DE CONFIANZA DEL COEFICIENTE DE CORRELACION. Indica si el coeficiente de correlación hallado en el caso de una muestra determinada puede ser atribuible o no al azar, y a qué nivel de confianza puede afirmarse que ello es así.

Los niveles utilizados son del 5% y del 1%. Si se indica que el valor del coeficiente de correlación hallado no es significativo, quiere decir que ese valor puede ser debido al azar, (es decir, que es probable que tomando un número más amplio de casos el coeficiente de correlación resulte ser cero). Si se indica que es significativo, quiere decir que podemos afirmar (con un 95% o con un 99% de probabilidades de acertar según el nivel de confianza indicado) que la covarianza hallada no se debe al azar.

RECTA DE REGRESION. Permite predecir los valores más probables en la variable Y de cualquier caso del que conozcamos su puntuación en la variable X. La fórmula de esa recta es: $Y=A+BX$; y los valores A y B son los que el programa proporciona.

HISTOGRAMA: Es el gráfico que indica la forma aproximada de la distribución. Como paso previo, se efectúa una operación de agrupación de frecuencias en intervalos.

NUBE DE PUNTOS: Expresa gráficamente si las dos variables son o no concomitantes, y en qué grado. (En una de las modalidades de esta gráfica en el programa se indican además las medias aritméticas de las dos variables, y la recta de regresión).

Si la correlación es alta, los datos tienden a agruparse formando un huso; si es baja, tienden a formar un círculo o se distribuyen sin forma definida por toda la pantalla. Si el coeficiente de correlación es alto positivo, la recta de regresión partirá de cerca del origen de coordenadas (ángulo inferior izquierdo), y los puntos aparecerán cerca de ella; y si la correlación es alta y negativa, la recta de regresión partirá de extremo superior izquierdo. Los pares de datos coincidentes aparecen señalados como un solo par.

BOALOX INFORMATICA agradece su confianza al haber elegido este paquete de programas, y le brinda su apoyo y colaboración para solucionar cualquier problema que pueda presentársele.

A pesar de que estos programas han sido probados a fondo, los informes acerca de posibles errores son de gran valor. Agradecemos cualquier dato, comentario o sugerencia que pueda remitirnos.

BOALOX INFORMATICA, propietaria del "copyright" de este paquete de programas, se reserva todos los derechos de copia y difusión del material lógico y de los textos explicativos. La difusión, copia o utilización no autorizada de estos programas será perseguida con arreglo a la Ley.