BIBLIOTECA DE PROGRAMAS

EL PROGRAMA BMDP3S DE ESTADISTICA NO PARAMETRICA

A.- Descripción general

El programa BMDP3S, ejecuta sobre un conjunto de datos dado, uno o más de los siguientes test estadísticos no paramétricos:

a) Test de los signos.
b) Test de rangos con signo de Wilcoxon.
c) Coeficiente de correlación por rangos de Kendall.
d) Coeficiente de correlación por rangos de Spearman
e) El análisis de varianza en dos formas de Friedman
f) Coeficiente de concordancia de Kendall.
g) U-Test de Mann-Whitney.
h) El análisis de la varianza en una forma de Kruskal-Wallis.

El número de palabras (M) de memoria de computador son:

a) Para el test de los signos M =A+3B+VT+1.200
b) Para el test de Wilcoxon M=A+3B+VT+3CU+1.200
c) Para Kendall y Spearman M=A+B+2CU+1.200
d) Para Friedman y Kendall M=A+CU(VT+3)+1.200
e) Para Kruskal-Wallis y Mann-Whitney M=A+5GN+2CU+1200

El significado de las variables usadas es el siguiente:

CU = número de casos usados
VT = número total de variables
GN = número de grupos
A = VT(CU+14) + 6CU
B = VT(VT+1)/2

B.- Forma de los datos de entrada

La forma de los datos de entrada depende del tipo de test deseado;

1) Cuando se desea obtener el test de los signos, el de Wilcoxon o los coeficientes de correlación por rangos de Kendall o Spearman para n casos de al menos dos observaciones por caso, los datos se deben leer en n fichas, teniendo cada ficha los datos correspondientes a un caso.

2) Las mismas reglas se siguen para el test de Friedman y el coeficiente de concordancia de Kendall.

3) Cuando se desea el test de Kruskal-Wallis para n observaciones en k muestras independientes, los datos deben ordenarse en n fichas, con dos datos por ficha, el primero de los cuales contiene la observación y el segundo contiene el número de orden de la muestra a la cual el dato anterior pertenece.

C.- Procedimiento computacional

Paso 1.- Los datos se leen caso a caso. Se computan las medias y desviaciones típicas de todas las variables usadas y se escriben, así como sus valores máximo y mínimo.

Paso 2.- Los siguientes test se aplican a datos que consisten de n pares de muestras

$$(X_1,Y_1),(X_2,Y_2), \ldots ,(X_n,Y_n)$$

Test de los signos

Para cada par de variables, se computa la diferencia $D_i = X_i-Y_i$, y se cuentan el número total de diferencias positivas NPLUS y negativas NMINUS. Después se computa el número total de diferencias distintas de cero T=NPLUS+NMINUS, y el test estadístico L=mín(NPLUS,NMINUS), y en la hipótesis de que las dos muestras procedan de una misma población, se calcula la probabilidad P de que la variable aleatoria ξ=mín($_1$, $_2$) sea menor o igual que L, siendo $\xi_1(X_1,X_2,\ldots,X_n,Y_1,Y_2,\ldots,Y_n)$ el número de signos positivos de las diferencias X_i-Y_i y $_2$ el número de signos negativos.

Si T 25 se computa P usando la distribución binomial:

$$P = 2^{-T}\sum_{i=0}^{L} \frac{T!}{i!(n-i)!}$$

Si M 25, se calcula P usando la distribución normal aproximada de la binomial.

Para cada par de variables la salida consiste de:

1) El número de diferencias distintas de cero T.

2) El número L = min(NPLUS,NMINUS).

3) La probabilidad P.

Test de los rangos con signo de Wilcoxon.

Para cada par de variables (X_i,Y_i), i=1,2,...,n , se computa la diferencia $D_i=X_i-Y_i$, y se ordenan de menor a mayor los valores absolutos de las diferencias distintas de cero, asignando a cada diferencia su número de orden en la ordenación anterior, teniendo en cuenta que si hay varias diferencias iguales en valor absoluto, a cada una de ellas se le asigna la media de los números de orden que les correspondiesen si fuesen distintas, y estuviesen entre los inmediatos menor y mayor A continuación se asocia a cada rango el signo de la diferencia con él asociada, y se computan el número total de diferencias distintas de cero T; así como la suma, SUMN, de los rangos negativos el estadístico L = min(SUMP,SUMN), y la probabilidad de que este estadístico sea menor o igual al valor L_o de la muestra particular, suponiendo que L sigue una ley normal.

Para cada par de variables la salida consiste en:

1) El número T, de diferencias distintas de cero entre pares.

2) El número L_o

3) La probabilidad P.

Coeficiente de correlación por rangos de Kendall.

Para el par de muestras (X_i,Y_i), i=1,2,...,n ; se ordenan, separadamente, de menor a mayor los valores X_i e Y_i, asignando a cada valor $X_i(Y_j)$, el número $r(X_i)(r(X_j))$ que ocupa en la ordenación anterior, teniendo en cuenta que si hay varios iguales, a todos ellos se les asigna la media de los lugares que ocuparían si fueran distintos y estuvieran entre el inmediato inferior y el inmediato superior. Cada par (X_i,Y_i) se ordena según el rango de X_i y se calcula S mediante los rangos de la variable Y.

$$S = \sum_{i=1}^{n-1}\sum_{j=1}^{n} Sgn(r(Y_i)-r(Y_j)) \quad \text{donde} \quad Sgn\ (u)=\begin{cases} 1 & \text{si } u>0 \\ -1 & \text{si } u<0 \\ 0 & \text{si } u=0 \end{cases}$$

Sean NTX, NTY el número de grupos de valores iguales para X e Y respectivamente, t_{x_i}, t_{y_j} el número de valores del - i-ésimo o j-ésimo grupo respectivamente. Se computan entonces los siguientes factores de corrección:

$$TX = \frac{1}{2}\sum_{i=1}^{NTX} t_{x_i}(t_{x_i}-1) \quad y \quad TY = \frac{1}{2}\sum_{j=1}^{NTY} t_{y_i}(t_{y_i}-1)$$

Finalmente se computa el coeficiente de correlación de Kendall.

$$T=\begin{cases} \dfrac{2S}{n(n-1)} & \text{si } TX=TY=0 \\[2ex] \dfrac{S}{\sqrt{\frac{1}{2}n(n-1)-TX}\sqrt{\frac{1}{2}n(n-1)-TY}} & \text{si } TX \neq 0 \text{ ó } TY \neq 0 \end{cases}$$

Se imprime T para cada par de variables.

<u>Coeficiente de correlación por rangos de Spearman</u>.

Para el par de muestras (X_i,Y_i), i=1,2,...,n ; se ordenan, separadamente, de menor a mayor los valores X_i, Y_i, asignando a cada valor $X_i(Y_j)$, el número $r(X_i)(r(Y_j))$ que ocupa - en la ordenación anterior, teniendo en cuenta que si hay varios iguales, a todos ellos se les asigna la media de los lugares que ocuparían si fueran distintos y estuvieran entre el inmediato inferior y el inmediato superior.

Se computa a continuación, la suma de cuadrados de diferencias de rangos $(r(X_i)-r(Y_j))$ y se obtiene:

$$D = \sum_{i=1}^{n} (r(X_i)-r(Y_j))^2$$

Los valores NTX, NTY, t_{x_i}, t_{y_j} ; tienen el mismo significado que en el apartado anterior. Para estos valores se computan los siguientes factores de corrección:

$$TX= \frac{1}{12}\sum_{i=1}^{NTX} t_{x_i}^3 - t_{x_i} \quad y \quad TY= \frac{1}{12}\sum_{i=1}^{NTY} t_{y_i}^3 - t_{y_i}$$

Finalmente se computa y se imprime, para cada par de variables, el coeficiente de correlación por rangos de Spearman RS.

$$RS=\begin{cases} 1 - \dfrac{6D}{n^3-n} & \text{si } TX=TY=0 \\ \dfrac{A+B-D}{2AB} & \text{si } TX\neq 0 \text{ ó } TY\neq 0 \end{cases} \quad \text{donde} \quad A= \frac{n^3-n}{12} - TX \qquad B= \frac{n^3-n}{12} - TY$$

Paso 3.- Los siguientes test son aplicables a una muestra aleatoria simple de extensión n y k variables emparejadas

$$(X_{11},X_{12},\dots,X_{1k}),(X_{21},X_{22},\dots,X_{2k}),\dots,(X_{n1},X_{n2},\dots,X_{nk})$$

Análisis de la varianza de Friedman.

Cada conjunto de variables de una muestra $(X_{i1}, X_{i2}, \dots$ $\dots, X_{ik})$ se ordena de menor a mayor, asignando a cada X_{ij} el número de orden $r(X_{ij})$ en la ordenación anterior, teniendo en cuenta, que cuando hay varios valores iguales, a todos ellos se les asigna el valor medio de los valores que les correspondiesen si todos fueran distintos y estuvieran comprendidos entre los inmediatos superior e inferior. Al número $r(X_{ij})$ se le llama rango de X_{ij}. Se suman los rangos sobre cada una de las k variables. Para la j-ésima variable obtenemos:

$$R_j=\sum_{i=1}^{n} r(X_{ij})$$

donde R_j es la suma de los rangos para la j-ésima variable y $r(X_{ij})$ es el rango de X_{ij}, la j-ésima variable en la i-ésima muestra.

A continuación se computa el estadístico XR del test de Friedman:

$$XR= \frac{12}{nk(k-1)}\sum_{j=1}^{K} R_j^2 - 3n(k+1)$$

Se computa finalmente el nivel de significación del estadístico, suponiendo que la distribución de XR es una χ^2_{k-1} con k-1 grados de libertad.

La salida, para todas las variables especificadas, consiste en:

1) La suma de rangos de cada variable, R_j.
2) El estadístico XR del test de Friedman.
3) El nivel de significación del estadístico XR.

Coeficiente de concordancia de Kendall.

Se computa este coeficiente, W, transformando el estadístico XR de Friedman:

$$W = \frac{XR}{n(k-1)}$$

Paso 4.- Los siguientes test se aplican a k muestras aleatorias simples de una variable aleatoria, teniendo n_i observaciones la muestra i-ésima y siendo n= n_i Los datos muestrales estarían representados por:

$$(X_{11}, X_{21}, \ldots, X_{n_1 1}), (X_{12}, X_{22}, \ldots, X_{n_2 2}), \ldots, (X_{1k}, X_{2k}, \ldots, X_{n_k k})$$

Análisis de la varianza de Kruskal-Wallis.

Cada variable es observada n veces en k muestras o grupos siendo n_i la extensión de la i-ésima muestra. Se ordenan de 1 a N y de menor a mayor todas las observaciones, asignando a cada X_{ij} el número de orden $r(X_{ij})$ de la anterior ordenación, teniendo en cuenta, que cuando hay varios valores iguales, a todos ellos se les asigna el valor medio de los valores correspondientes, caso de que fueran todos distintos y estuvieran comprendidos entre los inmediatos superior e inferior. A continuación se suman los rangos de cada una de las muestras, obteniendo:

$$R_j = \sum_{i=1}^{n} r(X_{ij})$$

donde R_j es la suma de los rangos para el j-ésimo grupo y $r(X_{ij})$ es el rango de X_{ij}, y la i-ésima observación en el j-ésimo grupo.

Seguidamente se computa la corrección para los valores muestrales iguales

$$T = \sum_{i=1}^{NT} t_i^3 - t_i$$

donde NT es el número de conjuntos con valores iguales, y t_i es el número de elementos en el i-ésimo grupo.

BIBLIOGRAFIA

1) HAJEK JAROSLAU (1.969)
"Nonparametric Statistics"
Holden-Day

2) CONOVER, W.J. (1.971)
Practical Nonparametric Statistics.
New York

3) BMDP (1.975)
"Biomedical Computer Programs"
University of California Press

Miguel Sánchez García
Biblioteca de Programas.

MANUALES NECESARIOS PARA UTILIZAR LOS PAQUETES DE PROGRAMAS EXISTENTES

1.- BMD - Biomedical Computer Programs.
University of California Press, 1973.

2.- OSIRIS III. System and Program Description.
Institute for social research.
The University of Michigan.

3.- GH20-0849-3 Introduction to Mathematical Programming System.
Extended MPSX, Mixed Integer Programming (MIP) and Generalized Upper Bounding (GUB).

4.- GH20-0205-4 System/360 Scientific Subroutine Package.
Version III. Programmer's Manual.

5.- Institut des Sciences Economiques Université Catholique de Louvain. Parkstraet, 121, B-3000 Louvain.
Time Series Processor. User's Manual.

6.- GH10-8067 Sistema de Simulación de uso General para el Sistema/360 (GPSS/360).
Manual de usuario.
360-CS-17X(OS) y 360A-CS-19X(DOS).

7.- SYAL.

8.- CSMPX.