Comandos de Limdep

Por Douglas Ramírez

GUÍA DE PARA EL USO DEL PROGRAMA LIMDEP

Economista: Douglas C. Ramírez Vera

A. Características del Limdep

El programa posee las siguientes procesos

- ✓ Estadísticas descriptivas (media, desviación, Skewnes, Curtosis, máx. y mín., etc.)
- ✓ Estratificación.
- ✓ Identificación de series de tiempo, correlaciones parciales y autocorrelaciones.
- ✓ Correlaciones, histogramas, plots.
- ✓ Análisis de regresión por MCO y MCG
- ✓ - En el análisis de regresión
    - o Estimadores robustos frente a problemas de Heterocedasticidad.
    - o Estimadores robustos frente a problemas de autocorrelación.
    - o Heterocedasticidad multiplicativa
    - o Heterocedasticidad grupal y autocorrelación en MODELOS DE CORTE TRANSVERSAL.
    - o Estimación Box-Cox
    - o Efectos aleatorios y fijos en modelos A PANEL.
    - o Estimación de modelos no lineales
    - o Modelos de rezagos distribuidos y ARIMA y ARMAX.
    - o Ecuaciones simultaneas.
    - o Estimación por máximo verosimilitud con información completa
    - o Modelos Logit, Probit y multinomial
    - o Modelos binomiales y de Poisson
    - o Modelos Tobit y de datos censurados (de frontera).
    - o Modelos de fronteras estocásticas
    - o Y otras facilidades (ver manual)

B. Mínimos Cuadrados Ordinarios en Limdep

El método de estimación de mínimos cuadrados (MCO o “Ordinary Least Squares Regresión”) consiste en minimizar la suma cuadrada de los desvíos entre los valores observados y el modelo muestra estimado para un conjunto de parámetros.

C. Sobre las Salidas

Las salidas dependerán del tipo de procedimiento utilizado para estimar el modelo, en principio los modelos que requieren procesos más complejos y convergen sus resultados por procesos iterativos parten de una estimación de mínimo cuadrática (MCO o LS abreviatura en inglés) y de ahí obtienen las soluciones correspondientes. Véase a continuación ejemplos de salidas.

```
--> REGRESS       ; Lhs = GRADE
             ; Rhs = GPA,TUCE,PSI $

+-----------------------------------------------------------------------+
| Ordinary    least squares regression    Weighting variable = none     |
| Dep. var. = GRADE    Mean=   .3437500000    , S.D.=   .4825587044     |
| Model size: Observations =      32, Parameters =   3, Deg.Fr.=     29 |
| Residuals:  Sum of squares= 5.447726127    , Std.Dev.=         .43342 |
| Fit:        R-squared=  .245337, Adjusted R-squared =          .19329 |
| Model test: F[  2,     29] =    4.71,    Prob value =          .01688 |
| Diagnostic: Log-L =    -17.0774, Restricted(b=0) Log-L =     -21.5812 |
|             LogAmemiyaPrCrt.=   -1.582, Akaike Info. Crt.=      1.255 |
| Autocorrel: Durbin-Watson Statistic =   2.48763,   Rho =      -.24381 |
+-----------------------------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |t-ratio |P[|T|>t] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 GPA       .1629089505      .13748772        1.185   .2457  3.1171875
 TUCE     -.1358674309E-01  .19622205E-01    -.692   .4942  21.937500
 PSI       .3649747953      .15535102        2.349   .0258  .43750000
```

En esta salida se puede describir por renglones su significado, partiendo de la primera fila o renglón.

1) La primera fila describe el método de estimación, en este caso: mínimos cuadrados ordinarios sin ponderación.

2) En el segundo renglón define a la variable dependiente (para el ejemplo GRADE) y su valor medio o promedio muestral de la variable dependiente que mide la tendencia central, este es igual a:

$$\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_i$$

En la misma línea, se arroja el valor de la desviación estándar muestral de la variable dependiente que es igual a

$$S.D = \sqrt{\frac{\sum_{t=1}^{T}(y_i - \bar{y})^2}{T-1}}$$

Este estadístico mide la dispersión o escala de la variable y.

3) En la tercera línea se describe el tamaño del modelo, en términos del numero de observaciones (T =100 observaciones), el número de parámetros (k =3) y sus grados de libertad[1] (Degree Freedom; T-k=97), los grados de libertad me mide él numero de parámetros fijos que establecen una restricción al estadístico utilizado sea una T de Student, un F o una Chi cuadrado[2].

4) La cuarta línea indica los valores residuales estimados de la muestra. El objetivo del método de los mínimos cuadrados ordinarios es minimizar los desvíos al cuadrado por lo tanto, es natural calcular dicho valor, este en si mismo no tiene mucho valor, pero sirve como dato de otros diagnósticos que se describen más adelante y es útil para comparar modelos y probar hipótesis. La formula es:

$$SCR = \sum_{t=1}^{T} e_t^2$$

A su vez, se muestra en esta línea la desviación estándar de la regresión (Std. Dev) que es la raiz cuadrada de la varianza de la estimación construida con los residuos que son calculados con los betas estimados (k parámetros)

[1] Cuando se utilizan estadísticas muéstrales, es necesario determinar él numero de variables que pueden variar. Como por ejemplo, si la suma de cuatros números es 20, pueden escribirse varias combinaciones de tres números, pero el cuarto número es obligado para que el total de la suma de 20. Si se seleccionan 7,4 y 1 como los tres números, el cuarto numero debe ser 8, de manera que la suma de todos sea 20. Debido a una restricción, se dice que "se pierde un grado de libertad".

[2] Por ejemplo, supóngase que se sabe que la media de cuatro números es 5. los cuatro números son 7,4,1 y 8. Las desviaciones de estos números respecto a la media deben ser en total cero (0). Las desviaciones de +2, -1, -4, y +3 dan en total 0. Si las desviaciones de +2, -1 y –4 se conocen, entonces el valor de +3 es fijo (obligatoriamente) para satisfacer la condición de que la suma de las desviaciones debe ser igual a cero (0). Así, 1 grado de libertad se pierde en el problema de muestreo que comprende la desviación estándar de la muestra ya que se conoce un número, la media aritmética.

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^{T} e_t^2}{T-k}$$

S^2 es una estimación de la dispersión de la perturbación en la regresión (σ^2) y, por consiguiente se usa para evaluar la bondad del ajuste del modelo y estimar el error de pronóstico. Cuanto mayor sea S^2 es peor el ajuste del modelo y será probable cometer errores de pronósticos mayores. El error estándar de la regresión es un estimador de σ y por tanto se usa S para representarlo y su ecuación es

$$\hat{\sigma} = \sqrt{\left(\sum_{t=1}^{T} e_t^2 \Big/ T-k \right)}$$

5) El renglón número cinco, indica los estadísticos asociados a la bondad de ajuste. Muestra el Coeficiente de Determinación (R^2 ó R-squared). Si la regresión incluye la ordenada en el origen, el R cuadrado debe estar entre cero y uno ($0 \leq R^2 \leq 1$). En ese caso representa el porcentaje de la varianza de la variable dependiente que es explicada por el modelo estimado y por tanto mide el éxito de la ecuación de regresión, dentro de la muestra para predecir a la misma (en el ejemplo es del 24,5%). La ecuación es:

$$R^2 = 1 - \frac{\sum_{t=1}^{T} e_t^2}{\sum (y_t - \bar{y})^2} = 1 - \frac{\frac{1}{T}\sum_{t=1}^{T} e_t^2}{\frac{1}{T}\sum_{t=1}^{T} (y_t - \bar{y})^2}$$

Nótese que el termino superior se aproxima a la varianza de la regresión y el termino inferior se aproxima a la varianza muestral de y. El R- cuadrado ajustado (Adjusted R-squared ó $\overline{R}^2$) se interpreta igual al coeficiente de determinación, pero la ecuación incorpora correcciones de acuerdo con los grados de libertad que se usaron para ajustar el modelo, a fin de compensar una visión optimista de buen ajuste y pronostico. En consecuencia el $\overline{R}^2$ es una medida más confiable que el R^2 simple. La ecuación es:

$$\overline{R}^2 = 1 - \frac{\frac{1}{T-k}\sum_{t=1}^{T} e_t^2}{\frac{1}{T-1}\sum_{t=1}^{T} (y_t - \bar{y})^2}$$

Nótese que el numerador es la varianza de la regresión y el denominador es la varianza muestral de la variable dependiente o la variación total.

6) Este renglón esta constituido dos elementos, la prueba asociadas a la validez del modelo general y la probabilidad del "**Error Tipo I"** de la prueba. La primera prueba muestra el F calculado con k-1 grados de libertad en el numerador y T-k grados de libertad en el denominador. Se emplea para comprobar la hipótesis de que los coeficientes de todas la variables (pendientes) excepto la ordenada en el origen, son cero.[3] Su formula es.

$$F = \frac{(SCR_{restringida} - SCR)/(k-1)}{SCR/(T-k)}$$

Donde SCR res es la suma de los residuos elevados al cuadrado, procedentes de una regresión *restringida* con sólo la ordenada en el origen y SCR son la suma a cuadrado de los residuos con el modelo estimado. La prueba consiste en cuanto aumenta la parte no explicada si el modelo elimina todas las variables excepto la constante. Si aumenta mucho implica que cuando menos alguna de ellas posee posibilidades predictivas.

7) El diagnostico del modelo se evalua a través de varios indicadores. El primero es el logaritmo de la *función de verosimilitud* que es la función de densidad conjunta de los datos, considerada como función de los parámetros estimados del modelo. Una estrategia de estimación puede ser maximizar la función, para ello se busca determinar el conjunto de los parámetros que logran dicho objetivo. El valor se compara con la misma función que sólo estima el valor medio (ordenada en el origen) versus la inclusión de las variables explicativas (pendientes). Su ecuación es la siguiente

$$\log L(\mu,\sigma^2) = -\tfrac{n}{2}\left[\ln(2\pi) + \ln\sigma^2\right] - \tfrac{1}{2\sigma^2}\sum_{t=1}^{T} e_t^2$$

Al igual que la suma al cuadrado de los residuos, este no tiene aplicación directa, pero es útil para comparar modelos y comprobar hipótesis.

La línea dos del séptimo renglón arroja dos estadísticos, uno es el criterio de información del logaritmo de Amemiya (AIC) y el otro el criterio de predicción de Akaike (PC), sus ecuaciones a saber son;

$$AIC = \ln\left(\frac{e'e}{T}\right) + (2k/T) = \ln\left(\sum_{t=1}^{T} e_t^2 / T\right) + \left(2k/T\right)$$

$$PC = \frac{e'e}{T-k}\left(1+\frac{k}{T}\right) = S^2\left(1+\frac{k}{T}\right)$$

Se utilizan para elegir entre modelos alternos de pronósticos y son criterios que penalizan por los grados de libertad. A medida que ambos valores disminuyen (se minimizan) más "adecuado" es el modelo.

8) El octavo renglón (novena línea) se refiere a la presencia o ausencia de autocorrelación serial de primer orden y muestra el estadístico Durbin-Watson y el valor del coeficiente de autocorrelación serial. La prueba de Durbin Watson funciona dentro del contexto del modelo.

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$$
$$\varepsilon_t = \varphi_0 + \varphi_1 \varepsilon_{t-1} + \nu_t$$
$$\nu_t \overset{iid}{\approx} N(0,\sigma^2)$$

La perturbación se correlaciona cuando $\varphi_1 \neq 0$. La hipótesis a rechazar es cuando $\varphi_1 = 0$. Cuando esto ocurre la estimación es adecuada. De lo contrario los errores siguen un proceso AR(1). La ecuación del estadístico Durbin-Watson.

$$DW = \frac{\sum_{t=1}^{T}\left(e_t - e_{t-1}\right)^2}{\sum_{t=1}^{T} e_t^{\,2}}$$

El DW se distribuye entre los valores de 0 a 4, estado su valor esperado alrededor de dos si es significativamente menor a dos existe evidencia de autocorrelación positiva, si por el contrario es significativamente superior a dos muestra evidencia de autocorrelación negativa. El DW y el coeficiente de autocorrelación serial se relacionan aproximadamente por la siguiente relación.

$$DW \approx 2(1-\hat{\rho})$$

9) En este renglón último se encuentran seis (6) columnas y filas igual al número de parámetros. La primera columna señala las variables incluidas en el modelo. La segunda columna indica el valor del parámetro estimado (pendientes y ordenada) o coeficiente, la tercera columna muestra el valor de la desviación estándar del parámetro, la cuarta columna muestra el valor de la prueba t student (calculado) de significación de los valores medios estimados de los parámetros, estas columnas se construye dividiendo la segunda columna

[3] No se restringe que todos los parámetros son cero ya que la ordenada, ante la ausencia de pendientes seria

con la tercera columna. La quinta columna muestra la respectiva probabilidad del error tipo uno para cada coeficiente o parámetro estimado. La última columna muestra el valor medio de cada variable incluida en el modelo.

D. Operación con Limdep

Al iniciar el programa se despliega automáticamente un archivo "*PROJECT*" el cual muestra los archivos de salida y entrada de las operaciones y donde se guardaran los resultados de las operaciones, el menú, en **File**, ejecuta **New** y muestra dos opciones una es "*PROJECT*" , toma la opción "*TEXT/COMMAND DOCUMENT*", donde se guardara el conjunto de instrucciones de las secciones estos archivos (TRACE FILE) pueden ser guardados con un nombre asociado a la actividad desarrollada y reutilizados en una sesión posterior. Para guardar diríjase a Archivo (FILE) y guarde como (SAVE AS) ponga el nombre y se guardara con el programa con extensión LIM (<Nombre.lim>).

Los comandos básicos de operación se resumen en estas líneas.

1. Inicio, creando el archivo donde se guardan los resultados.

Open; Output=<*nombre del archivo de salida.extensión***>$** [Se omite en ambiente Windows]

2. Lectura de datos, en este caso en formato libre

Read; File=<*nombre del archivo. Extensión***>; Names; Format=<***Extensión***>$**

3. Transformación de variables

CREATE; <Nueva Variable**>=<**operador**>(<**variable a transformar**>);...$**

4. Estadísticos descriptivos

DSTAT; **Rhs=*$** Realiza un análisis estadístico descriptivo de todas las variables.

Ejecución de la operación sintaxis general

[Comando de Operación];Lhs=<Variable(s) dependiente(s) >; Rhs=<Variables independientes separadas por comas>,<otras especificaciones>,...$

---

igual al promedio de y, el cual por lo general no es cero.

Al final de cada instrucción esta concluye con el símbolo de peso ($), y las instrucciones se separan con punto coma (;), él listado de variables se separa con las comas (,). Los [Comando de Operación] son nombres que señalan el procedimiento de estimación, TOBIT, REGRESS, PROBIT. La variable sobre la que se realiza la regresión le antecede el comando de asignación Lhs= (left hand side) y a las variables que explican la regresión le antecede el comando Rhs= (right hand side).

En ambiente Windows, en la barra de menú se encuentra **Run,** al desplegarlo se ven las diferentes opciones para ejecutar los comandos, que se pueden realizar por línea (instrucción) o por selección de un conjunto de instrucciones.

El Limdep al recibir cada instrucción la compila y muestra en el archivo de salida el error de procedimiento, esto ayuda a depurar el programa corrigiendo los errores de sintaxis o de lógica computacional.

<u>Ejemplo</u>

```
READ        ;NOBS=25; NVAR=4; NAMES=ESTB, Ki, Li, Qi $
CREATE      ; VA= Qi ; Y = Log(Qi) ;  X1 = Log(Ki)
            ; X2 = Log(Li) ;  X3 = (X1-X2)^2$
NAMELIST    ; X = One,X1,X2,X3,X4 $
MATRIX      ; XX = <X'X> $
MATRIX      ; XY = X'Y $
MATRIX      ; BETAS= XX* X'Y $
```

SINTAXIS

Los comando poseen la siguiente forma general

```
VERBO  ;  especificación  ;  especificación  ;  ...  ;  especificación
MODEL COMMAND          ; Lhs   = variable dependiente
                       ; Rhs   = lista de variables independientes ;
Otras especificaciones del modelo  a estimar; ... $
```

<u>Ejemplo</u>

```
REGRESS          ; Lhs = <Variable dependiente>
                 ; Rhs = <Variables explicativas>
                 ; Panel
                 ; Str = variable $.
```

Ejemplo:

```
? Lectura de datos dentro del programa
READ ; nvar = 4 ; nobs=25 ; names =  Y,  X1,  X2, X3 $
?Estadisticos Descriptivos
DSTAT ; Rhs = Y,  X1,  X2, X3 $
?Ploteo
PLOT  ;  Lhs = X1 ;  Rhs = Y  $
? Regresión
REGRESS         ; Lhs = Y
 ; Rhs =ONE, X1, X2, X3
; Res=Error; Kepp=Yhat  $.
/* Transformación de variables
Obtener el logaritmo, y renombrar las variables */
Create  ; LY=log(Y)      ; LX1=log(X1)  ; LX2=log(X2)  ; LX3=log(X3) $
? Regresión
REGRESS         ; Lhs = Y
 ; Rhs =ONE, X1, X2, X3
; cls: B(2)+B(3)+B(4)=0
; Res=Error; Kepp=Yhat  $.
```

E. Comandos Frecuentes

Comando REGRESS

Este comando estima los parámetros por el método de los mínimos cuadrados ordinarios, conocido también como modelo de regresión lineal

$$y = b'x + e$$

```
REGRESS        ;  Lhs = variable dependiente  ;  Rhs  =  regresores $
```

Ejemplo

```
REGRESS        ; Lhs=Y ; Rhs=ONE,Y[-1] $
```

Si quiere estimar la constante debe incluirse la palabra ONE que incluye un vector columna de unos.

Ejemplo

```
REGRESS     ; Lhs = Y
            ; Rhs  = One,X1,X2
            ; RES=ERROR; KEEP=YHAT$
```

Opciones de Regress.

; Rh2 = <Lista de variables>. Es una lista de variables extras. Cada variable asignada en Rh2 es una variable a ser añadida al modelo, una a la vez, para evaluar sus efectos en:
(1) Su efecto en el coeficiente de determinación R^2 y
(2) La significacia del coeficiente a través de la prueba t , si la Razon t es significativa o no.
; Plot elabora un ploteo o gráfico de dos dimensiones de los residuos de la estimación
; Plot (variable) Realiza un gráfico de los residuos contra una variable del modelo.
; List Imprime los valores de los valores predichos o estimados de la variable dependiente, de los residuos y de los intervalos de confianza de la predicción as un 95% de confianza.
;Keep = <Nombre>. Asigna a una variable los valores predichos o estimados de la variable dependiente de la regresión.
; Res = <Nombre>. Asigna a una variable los valores predichos o estimados de los residuos de la regresión.
; Res ; Standard Construye los residuos estándar (restando la media y dividendo por la desviación.
; Fill Construye la predicción con los valores no utilizados (omitidos) en la regresión..
; CLS: Estima la regresión por minimos cuadrados restringidosy prueba si la restricción es valida a través de la prueba F.
; CUSUM = <Nombre> Estima los valores del Cusum y del CUSUM-cuadrado y prueba la estabilidad estructural de la estimación (modelo).

Comando Namelist

NAMELIST es usado para definir un conjunto de variables y sirve como sinónimo de ellas permite evitar reescribir una larga lista de variables y simplifica el trabajo.

Sintaxis

NAMELIST; <Nombre de la variable>=< Lista de variables separadas por comas> $

También se puede construir instrucciones anidadas.

NAMELIST ; XALL = X1, X2, X3$

NAMELIST ; Set1 =X1,X2$

NAMELIST ; Set2 =X1,X2,X3$

NAMELIST ; Set3 = X*$

NAMELIST ; Bigset = XALL,Set1,Set1 $

Nótese que Bigset esta construida sobre la base de otras lista de nombres. ***NAMELIST*** puede contener hasta 100 nombres de variables y pueden definirse hasta 10 listado con 100 variables. Si necesita espacio y necesita limpiar la memoria puede usar la siguiente instrucción

NAMELIST ; Delete name(s) $

Ejemplo

NAMELIST ; Delete BIGSET $

Comando DSTAT

Las salidas típicas incluyen los valores medios, la desviación estándar, el skewness, la kurtosis, los valores mínimos, máximos, el numero de casos u observaciones, los valores perdidos para cada variable especifica

DSTAT Produce las estadísticas descriptivas de un conjunto de variables.

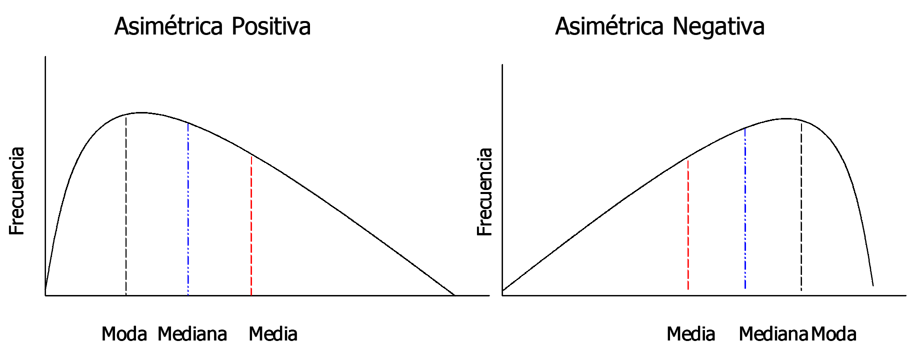
Las estadísticas descriptivas son la media, la desviación estándar, el coeficiente de asimetría o sesgo (Skewness), el coeficiente de concentración (Kurtosis), valores máximos, mínimos y él numero de casos validos.

El Skewness se describe por la siguiente formula.

$$S = \frac{\frac{1}{T}\sum_{t=1}^{T}(y_t - y)^3}{\left(\sqrt{\frac{\sum_{t=1}^{T}(y_t - y)^2}{T-1}}\right)^3} = \frac{\mu_3}{SD^3}$$

Donde S es el coeficiente de asimetría (Skewness) y SD es la desviación estándar muestral de la variable, μ_3 es el tercer momento de la distribución[4].

El sesgo o asimetría se mueve entre valores de –3 y 3 si existe asimetría en una distribución normal su valor sería igual a cero. Con frecuencia una distribución no es simétrica alrededor de ningún valor, en lugar de ello posee una de sus colas más larga (delgada) que la otra. Si la cola superior de la distribución (cola derecha) es más gruesa (gorda) que la cola inferior el Skewness (asimetría o sesgo) sería positivo, en caso contrario sería negativo.

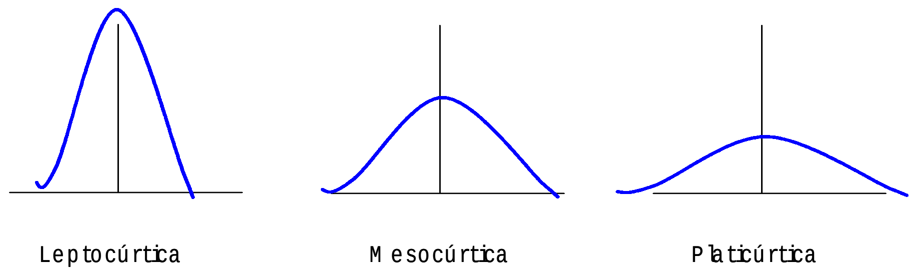


La Kurtosis mide el grado de agudeza (concentración) de una distribución (curtosis) y es obtenido por el siguiente estimador.

$$K = \frac{\frac{1}{T}\sum_{t=1}^{T}(y_t - y)^4}{\left(\sqrt{\frac{\sum_{t=1}^{T}(y_t - y)^2}{T-1}}\right)^4} = \frac{\mu_4}{SD^4}$$

Donde K es el coeficiente de concentración (curtosis) y SD es la desviación estándar muestral de la variable, μ_4 es el cuarto momento de la distribución. La curtosis de una distribución normal es igual a 3. Si la distribución posee colas más gruesas que la normal (menos aguda) su curtosis será mayor que tres

[4] El primer momento de la distribución es la media y el segundo momento es la varianza.



<u>Sintaxis</u>

DSTAT ; Rhs = Lista de variables $

Ejemplo

DSTAT; Rhs = LY,LX $

Opciones:

; All Elabora todos las estadísticas descriptivas incluyendo el skewness y la kurtosis.

; WTS = variable usada para ponderar o dar un peso.

; OUTPUT = 1$ Calcula y muestra la matriz de covarianza.

; OUTPUT = 2 $ Calcula y muestra la matriz de correlación.

; OUTPUT = 3 $ Calcula y muestra la matriz de covarianza y de correlación.

Las matrices son construidas utilizando todos los datos que se encuentra en los vectores o variables listados

; QUANTILES a partir de la muestra construye los deciles, percentiles y cuartiles.

; PLOT Grafica los cuartiles normalizados de los datos.

Quintiles

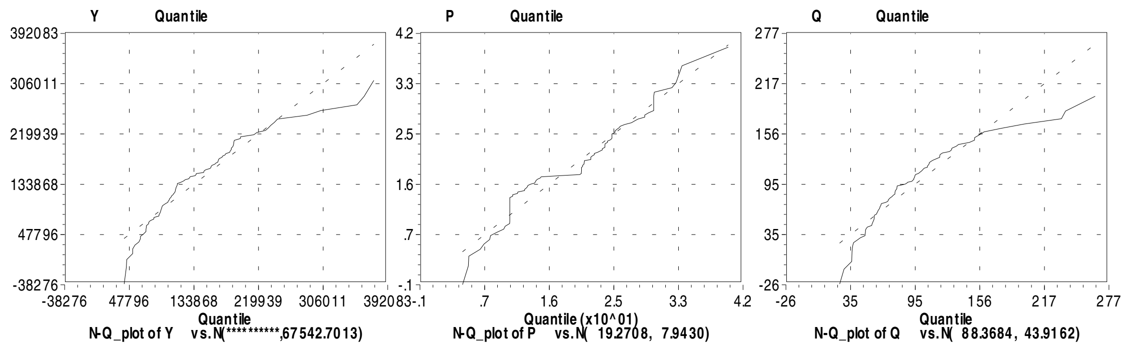
Sintaxis.

DSTAT ; RHS = <LISTA DE VARIABLES>; QUANTILES $

Ejemplo.

DSTAT ; Rhs = Q, P, Y; Plot $

Esta instrucción, además de proveer los estadísticos descriptivos, construye gráficos de probabilidad normal (“Normal Q-Q Plot” o Gráfico de Quintiles reales versus teórico de una distribución normal), que es la prueba más simple y visual para diagnosticar si los datos de la muestra al ser comparados con una distribución teórica se aproximan o no a una normal.



Aunque esta prueba es atractiva por su simplicidad presenta problemas para muestras pequeñas ya que el número de categorías o su anchura puede distorsionar la representación visual[5]. La distribución normal sigue una línea recta en diagonal (ver gráficos anexos), comparándola con el grafico de los valores de los datos observados de la muestra. Si una distribución es normal, la línea que representa la distribución real de los datos sigue de cerca de a la diagonal.

DSTAT ; Rhs = Q, P, Y; QUANTILES $

Esta instrucción provee los estadísticos descriptivos y los datos (lista) de los percentiles, los cuartiles y la mediana de la distribución de la muestra observada.

Por ejemplo la siguiente instrucción del comando DSTAT obtiene las siguientes salidas: Una relación de las estadísticas descriptivas; media, desviación, valores mínimos, máximos, número de casos, Skewness, Kurtosis y autocorrelación, las matrices de Varianzas y Covarianzas y la matriz de correlación entre las variables, los Percentiles y la mediana, los límites cuartilíticos y los gráficos de normalidad

DSTAT;Rhs=PNB,INV,RATE;Output=3;Quantiles;Plot;All;AR1$

Comando PLOT

PLOT es usado para producir gráficos simples de nubes de puntos

Sintaxis

[5] Para más detalles ver:
Visauta V., B. 2002. Análisis Estadístico con SPSS para Windows® Mc Graw Hill Madrid Vol. I .
Hair et al. 1999. Análisis Multivariante. Prentice Hall Madrid, 5ª Edición.

PLOT ; Lhs = variable del eje horizontal ; Rhs = variable(s) del eje vertical $

<u>Opciones</u>

; Fill conecta los puntos muéstrales con una línea en orden correlativo a como se encuentran listados

; Grid Coloca una rejilla o cuadricula de líneas en la figura.

; **Regression** muestra el ajuste lineal de la regresión

; **Limits** = LY,UY Establece los limites inferiores (LY, low Y) y superiores (UY, upper Y) para el eje vertical de las abscisas (para los valores dados).

; **Endpoints** = LX,UX Establece los limites inferiores (LX, low X) y superiores (UX, upper Y) para el eje horizontal de las abscisas (para los valores dados).

F. Lectura de archivo desde una hoja de calculo

A veces el número de datos es muy grande y por tanto es preferible que se encuentren en un soporte de base de datos que sea legible por el programa y no "pegados" dentro del "*TEXT DOCUMENT"*. Los datos deben encontrarse en una hoja de calculo y no en formato "libro" para evitar riesgos se sugiere grabar en formatos de hoja de configuración menor como Excel 4, wks, wk1 o wq1 que son sólo hojas de calculo. Además la hoja debe ser "limpia" en el sentido de sólo poseer datos y no formulas, así como ningún formato extraño este debe ser el más simple (sin negrillas ni subrayados, ni bordes, ni sombras) ya que puede ser interpretado como ausencia de dato. Por último la hoja a ser leída no de estar compartida por ningún otro programa es decir debe estar cerrada.

<u>Ejemplo</u> : Lectura de un archivo lotus

```
Reset;$
Read     ; File=A:\limdep\Coquimbo.wks; Names; Format=wks$
Namelist       ; X=ONE, CD, PVI, SEXO, EDAD, EDUC, ING$
Regress; LHS=Y       ; RHS= X ; Keep =BX ;  res=E$
Plot     ;Lhs=bx ;   Rhs=E$
Create   ; E2=E*E$
Calc     ; SCR=Sum(E2)$
```

Ejemplo:

Lectura desde un archivo Excel.

```
Reset;$
? LECTURA DE LOS DATOS
Read; File=C:\DATOS\Producto.xls; Names; Format=xls$
? TRANSFORMACION DE LAS VARIABLES
Create; K2=K*K; L2=L*L; LK=LOG(K); LL=LOG(L); LY=LOG(Y)$
? ASIGNACION DE LAS VARIABLES A UNA MATRIZ
NAMELIST; X= K,L , K2, L2; LX= LK, LL$
```

G. Uso del Comando de Calculo

El comando se puede ejecutar con las primera cuatro letras (de hecho todos los comandos de Limdep son interpretado con sólo escribir las primera cuatro letras. CALC o calculate). Permite obtener un numero o escalar a partir de un vector u obtener un valor dado de acuerdo a un conjunto de parámetros como el valor en tablas de la distribución Normal, F, T o Chi cuadrado

Sintaxis

Calculate ; <nombre valor resultante > = <formula varible> $

Ejemplo

```
CREATE   ; LQ = LOG(Q);LP = LOG(P);  LY = LOG(Y)$
REGRESS  ; LHS=LQ; RHS= ONE, LP
         ; RES=ERRL0; KEEP=LQE0$
Create   ; Er02=ERRL0*ERRL0$
Calc     ; SCR=Sum(Er02) $
```

Opciones.

Sum(variable) = Valor de la suma total de la variable

Xbr(variable) = Valor de la media de la muestra

Sdv(variable) = Desviación Estándar

Var(variable) = Varianza

Min(variable) = Valor mínimo

Max(variable) = Valor máximo

Ntb(P) = Obtiene el valor de la Distribución Normal Estándar, para una probabilidad P.

Ttb(P,d) = Obtiene el valor de la distribución t (de Student) con d grados de libertad, para una probabilidad P

Ctb(P,d) = Obtiene el valor de la distribución chi-cuadrado con d grados de libertad

Ftb(P,n,d) = Obtiene el valor de la distribución F con n grados de libertad en el numerador y d grados de libertad en el denominador

Ntb(P,m,s) = Obtiene el valor de la distribución Normal con media y desviación definida por el analista.

Para los momentos con el comando calc

Sum(variable**)** **=** Obtiene el valor de la suma de la columna.

Xbar(variable**)** **=** Obtiene el valor medio de la variable.

Sdv(variable**)** **=** Devuelve el valor de la desviación estándar.

Var(variable**)** = Obtiene la varianza muestral.

Min(variable**)** **=** Da el valor mínimo

Max(variable**)** **=** Da el valor mínimo.

Cov(variable, variable**)** = Covarianza muestral

Cor(variable, variable**)** = Correlación muestral.

Dot(variable, variable**)** = La suma del producto de las variables.

Si **X** es una matriz de vectores columnas de las variables independientes, incluyendo el vector de unos, (ver comando **Namelist**) e **Y** es un vector columna de la variable dependiente, el comando calc se puede utilizar para computar los siguientes resultados.

Rsq(X,Y) = Calcula el coeficiente de determinación

Xss(X,Y) = Calcula la suma cuadrada explicada de la regresión (SCE)

Ess(X,Y) = Calcula la suma cuadrada del error de la regresión (SCR)

Tss(X,Y) = Calcula la suma cuadrada total de la regresión (SCT)

Ser(X,Y) = Calcula el error estándar de la regresión

Lik(X,Y) = Calcula el valor de la función de verosimilitud de la regresión (*Log-Likelihood Function*)

H. Uso de los Comandos de Transformación “Create”.

Use el comando CREATE para realizar transformaciones en las variables que son vectores columnas.

Sintaxis

CREATE ; <nombre nueva variable > = <formula variable > $

Puede realizar varias transformaciones simultaneas

CREATE ; variable <nombre1> = <formula1>

; variable <nombre2> = <formula2> ;... $

Ejemplos:

CREATE ; Y = X + 1 ; Z = X* Y $

CREATE ; LAGY = Y[-1] $

Puede incluir cualquiera de las siguientes opciones de transformaciones.

<*Nombre*>=**Log** (*X*) Logaritmo natural del vector x,

<*Nombre*>=**Exp** (*X*) Eleva al exponente al vector x,

<*Nombre*>=**Abs** (*X*) Obtiene el valor absoluto de los valores,

<*Nombre*>=**Sqr** (*X*) Calcula la raíz cuadrada del vector.

<*Nombre*>=**Sin** (*X*) Obtiene el valor del seno dado los valores del vector.

<*Nombre*>=**Cos** (*X*) Obtiene el valor del coseno, dado los valores.

<*Nombre*>=**Tan** (*X*) Obtiene el valor de la tangente, dado los valores.

<*Nombre*>=**Fix** (*X*) Redondea alrededor del entero más cercano.

<*Nombre*>=**Trn** (*X1, X2*) Obtiene la tendencia = X1+(i-1)*x2 ; donde i = al numero de observaciones,

<*Nombre*>=**Int** (x) Obtiene la parte entera de los elementos del vector.

I. Uso de los Comandos de Matriz

Todas las entradas numéricas de LIMDEP usan una forma matricial en su lectura computacional. Así que cuando usamos. NAMELISTS ; se entiende como una matriz de NxK dimensiones, donde N es el tamaño de la muestra y K el numero de variables o vectores columnas.

- VARIABLES ; son matrices de Nx1 o vectores columnas.
- SCALARS ; son matrices de 1x1 matrices o, cuando apropiadamente, es un escalar.
- MATRICES ; Es el resultado de las salidas computacionales y tienen la misma jerarquía que las anteriores.

La forma usual de trabajar con matrices por comandos es

MATRIX ; <Nombre> = los comandos apropiados y las expresiones $

<u>Ejemplo</u>

MATRIX ; DELETE <Nombre> $

MATRIX ; F = Q'R $? Donde Q y R son un mix de elementos como matrices, escalares, variables o números

MATRIX ; XX = Dtrm(X'X) $ obtiene el valor del determinante de la matriz

CALC ; LOGDET = Log(XX) $ Calcula elemento por elemento el logaritmo de la matriz

Algunas operaciones.

A'B = Transpuesta de la Matriz A por la Matriz B

<A> = inv(A) ; Inversa de la matriz A

X'X = Matriz de primeros momentos

X'Y = Matriz de productos cruzados

<u>Ejemplo.</u>

MATRIX ; XX = <X'X> $

MATRIX ; Xy = X'Y $

MATRIX ; Betas = <X'X>*XY $

Introducir directamente datos de una matriz?

MATRIX ; A = [1,2,3,4 / 4,3,2,1 / 0,2,0,1/2,5,6,0] $ $\begin{bmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \\ 0 & 2 & 0 & 1 \\ 2 & 5 & 6 & 0 \end{bmatrix}$

J. Otros Comandos.

SORT Ordena una variable de acuerdo a un orden de mayor a menor.

SORT ; Lhs = Ing$ Se ordena en función de la variable Ing.

SORT ; Lhs = name ; Rhs = * $ En este caso se aplica a toda la data

SORT ; All ; DESCENDING$ En este caso se ordena toda la data en sentido descendente.

Histogramas.

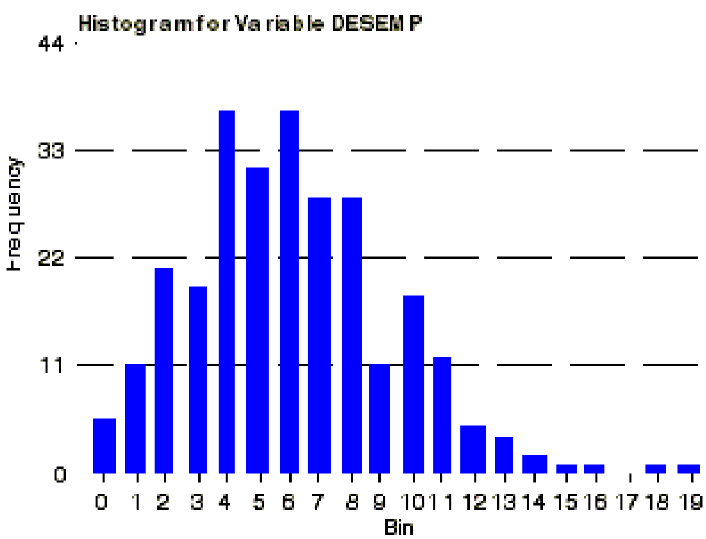
El histograma de los datos es la representación grafica de una única variable. Al formar el recuento de frecuencias en categorías, puede verse el perfil de la distribución de la variable. Es utilizado para realizar una comparación visual con la distribución Normal.

EL comando HISTOGRAM construye el grafico de un histograma por variable.

Sintaxis

HIST ; RHS =< variable>$

HISTOGRAM;Rhs=DESEMP$



El programa asume y construye automáticamente los intervalos, pero puede indicarse el numero de columnas del hsitorgrama, este caso 5 columnas

HISTOGRAM;Rhs=Q;Int=5$

También puede establecer la proporción de datos de cada columna por ejemplo decirle que cada una tenga el 25% de los datos lo cual hará que tenga 4 columnas.

HISTOGRAM;Rhs=Q;Bin=0.25$

Puede a su vez establecer los limites de graficación en datos continuos por ejemplo para cuatro intervalos se establece el primer tramo (L0-L1) con limite inferior L0 y superior L1 y así sucesivamente.

```
HISTOGRAM ; Rhs=<variables>
          ; limits =L0, L1,
                    L2, L3,
                    L4, L5,
                    L6, L7$
HISTOGRAM ; Rhs=Q
          ; limits =24.525, 57.116,
                    57.116,76.764,
                    76.764, 106.866,
                    106.866, 264.194$
```

Por ultimo se puede establecer el número de tramos y el límite inferior y superior.

HISTOGRAM ; Rhs=Q ; Int = 5 ; limits =24.525, 264.194$