

ESTADISTICA CON EViews: ESTADISTICA DESCRIPTIVA

F. MARMOL
UNIVERSIDAD CARLOS III DE MADRID

Curso 2001/2002

1 PRELIMINARES

CONJUNTO DE DATOS

Como ilustración, estudiaremos el fichero de datos “**estres.txt**” (véase la página web de la asignatura).

El fichero contiene 15 observaciones para analizar la relación existente entre el grado de estrés de los trabajadores (Y), medido a partir de su tensión nerviosa, y el tamaño de la empresa en la que trabajan ($X1$), el número de años que llevan en el puesto de trabajo actual ($X2$), el salario anual percibido ($X3$) y la edad del trabajador ($X4$).

INICIO DE LA SESION

Seleccionar **File New Workfile** y elegir un nombre para este Workfile.
INTRO.

Una vez en la correspondiente pantalla, debemos elegir la frecuencia de nuestros datos y especificar el tamaño muestral.

Al trabajar con datos de sección cruzada, debemos elegir la opción **undated frequency**. Véase el HELP del programa para especificar datos de series temporales.

En el ejemplo, debemos escribir las siguientes instrucciones:

```
File New Workfile
nombre: estres
undated frequency
start observation: 1
end observation: 15
```

con lo que crearemos un Workfile con el nombre “estres.wfl” diseñado para una muestra de 15 datos.

Este Workfile podrá ser utilizado en otras sesiones con la orden:

```
FILE
OPEN WORKFILE
“estres.wk1”
```

INTRODUCCION DE LOS DATOS

El siguiente paso es introducir los datos.

Primera posibilidad: Crear uno/a mismo/a la base de datos.

Para ello, vamos a

Object New

y seleccionamos

Serie

dándole un nombre. Repetiremos este paso para cada serie que deseemos introducir.

Los nombres aparecerán en el Workfile.

A continuación se procederá a la introducción de los datos. Para ello, hacemos click sobre el icono de la serie en cuestión y en el botón **Edit +/-** de la nueva ventana. Introducimos los datos de la serie. Antes de cerrar la ventana, hacer click otra vez sobre el botón **Edit +/-**.

Segunda posibilidad: la base de datos ya está creada.

En este caso, debemos seguir estos pasos:

File

Import

En este punto aparecerán diferentes posibilidades, según el formato que tenga la base de datos. En el caso particular del fichero “estres.txt”, es un fichero ASCII. La opción que tenemos que elegir es

Read Text-Lotus-Excel

Evidentemente, ya del nombre de esta orden podemos deducir que también sirve para importar ficheros de datos escritos en Excel o en Lotus.

Con ambas posibilidades, verificar que están todos los datos correctamente introducidos con **View**

GENERAR NUEVAS SERIES

A partir de las series introducidas, podemos generar nuevas series a partir del comando

GENR

Por ejemplo, si tenemos en nuestro Workfile una serie de nombre S1, podemos obtener el logaritmo de esta serie con la siguiente orden:

GENR LS1=LOG(S1)

Con esta orden hemos creado una nueva serie denominada LS1 que es el logaritmo de S1.

Véase el Help del programa.

DIBUJAR UN GRAFICO

En la ventana del Workfile, seleccionar la serie apretando dos veces el botón izquierdo del ratón sobre el nombre de la serie. Ir a **View** y seleccionar **Line Graph**. EViews incluye muchas opciones gráficas. Para verlas, apretando dos veces el botón izquierdo del ratón sobre el dibujo de la serie.

Los gráficos pueden ser grabados para su posterior uso.

TAMAÑO MUESTRAL

A este respecto, uno de los procedimientos del Workfile es la posibilidad de cambiar la muestra con la que vamos a trabajar. Para efectuarlo seleccionamos **Procs/Sample** en su Barra de Herramientas o hacemos doble click en la línea del Workfile que nos indica la muestra actual, dos líneas por debajo de dicha barra. La opción **Sample** también aparece como otro botón en la barra de herramientas.

Cualquiera que sea el camino que tomemos, aparece un cuadro de diálogo en el que se presenta por defecto, en la parte superior, la muestra existente en ese momento, que, de no haber sido modificada, coincidirá con el rango especificado al inicio de la sesión de ordenador. En dicha ventana podemos cambiar la muestra escribiendo la observación inicial y final del intervalo de trabajo que queramos (siempre, por supuesto, que estemos dentro del rango).

La muestra seleccionada no tiene por qué ser un intervalo continuo, sino que también podemos escoger varios intervalos o bien aquellos elementos de la muestra que cumplan una determinada condición.

Para escoger varios intervalos no solapados dentro de la muestra disponible, indicaremos en esa misma ventana las observaciones iniciales y finales de cada uno de los intervalos elegidos, con la única condición de que guarden entre sí un orden de menor a mayor.

Por ejemplo, supongamos que los datos son anuales y que el rango especificado al inicio de la sesión fue de

1964 1986

Ahora, si escribimos

1965 1968 1970 1975 1980 1986

estaríamos seleccionando una muestra que comprendería tres intervalos de 1965 a 1968, de 1970 a 1975 y de 1980 a 1986. Todas las observaciones que no pertenezcan a tales intervalos no se incluirían dentro de la muestra. Si alguno de estos intervalos hubiera estado formado solamente por una observación, la forma de seleccionarlo sería repitiendo la misma fecha como observación inicial y final. Por ejemplo, si en el caso anterior el primer intervalo fuera únicamente el año 1965, las dos primeras fechas deberían ser

1965 1965.

Eviews presenta ciertas abreviaturas que pueden utilizarse a la hora de describir el intervalo. Así, si queremos volver al total del rango desde el ejemplo anterior, en lugar de escribir de nuevo 1964 1986 podemos escribir **@all**. Otras abreviaturas posibles son **@first** y **@last** para la primera y última observación, respectivamente, del rango total. Por ejemplo

@first 1967 1980 @last

seleccionaría los intervalos que van desde 1964 a 1967 y de 1980 a 1986.

Igualmente, si queremos reservar un número concreto de observaciones a partir de una determinada, se puede indicar escribiendo la primera fecha como fecha de inicio y la primera fecha + el número de observaciones que

completan el intervalo deseado como fecha final. Así, escribiendo 1970 1970+12 escogeríamos los 13 datos que hay entre 1970 y 1982.

Si adicionalmente queremos que los elementos de la muestra seleccionada cumplan una determinada condición, podemos utilizar la ventana inferior del cuadro de diálogo de **Sample**. En este caso, la muestra resultante serán la intersección del conjunto de observaciones indicadas en la ventana superior y del conjunto de observaciones definidas por la condición incluida en la ventana inferior. Por ejemplo, si una de las variables se denomina $X1$, por ejemplo, entonces si escribimos respectivamente en cada ventana:

```
1964 1986
X1 > 0
```

estaríamos seleccionando como muestra las observaciones de todas las variables comprendidas entre 1964 y 1986 para las que la variable $X1$ es positiva.

De igual forma, si queremos seleccionar aquellas observaciones para las que $X1$ ha decrecido con respecto al año anterior las órdenes de cada ventana serían:

```
1964 1986
X1 < X1(-1)
```

También pueden construirse condiciones más complejas utilizando las órdenes **AND** y **OR**. Por ejemplo,

```
1964 1986
X1 > 0 and X2 >= 1
```

selecciona las observaciones para las que $X1$ es positiva y $X2$ es mayor o igual que 1.

Como otro ejemplo,

```
1965 1975 1980 @last
(X1 >= -10 and X1 <= 100) or X2 >= 1
```

selecciona las observaciones comprendidas entre los años 1965 a 1975 y 1980 a 1986 para las que $X1$ esté comprendida entre -10 y 100 o $X2$ sea mayor o igual a 1.

Si dispusiéramos de más observaciones y, por lo tanto, quisiéramos aumentar el rango fijado, el procedimiento a seguir sería seleccionar en la barra de herramientas del Workfile **Procs/Change workfile Range** o de forma más rápida, hacer doble clic sobre la palabra **Range** que aparece debajo de la barra de herramientas del Workfile.

2 ESTADISTICA DESCRIPTIVA

A la hora de especificar el modelo, es siempre aconsejable en **primer lugar realizar una primera inspección gráfica y numérica de las series**, pues esto nos va a permitir analizar el grado y la forma de la relación existente entre las variables.

2.1 ANALISIS INDIVIDUAL

Estando en el Workfile de interés, hacer doble click con el botón izquierdo del ratón sobre la serie de interés.

Ir a **View, Descriptive Statistics, Histogram and Stats**.

Muestra el histograma de la serie, X , así como los momentos muestrales más relevantes:

número de observaciones: n

media: \bar{X}

mediana

valor máximo: $X_{(n)}$

valor mínimo: $X_{(1)}$

desviación estándar de la serie (raíz cuadrada de la cuasivarianza muestral):

$$s_X = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$$

coeficiente de asimetría: $AS = \frac{1}{n} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\hat{\sigma}_X} \right)^3$, donde $\hat{\sigma}_X = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$,

coeficiente de curtosis: $CUR = \frac{1}{n} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\hat{\sigma}_X} \right)^4$,

contraste de Jarque-Bera:

$$JB = \frac{n-k}{6} \left(AS^2 + \frac{1}{4} (CUR - 3)^2 \right)$$

donde k es el número de parámetros estimados. Su hipótesis nula es que la serie tiene una distribución normal. Se distribuye cuando esta hipótesis nula es cierta como una chi-cuadrado con dos grados de libertad.

Probability: $P(|JB| > JB^*)$, donde JB^* es el valor que se obtiene de JB al sustituir en su fórmula los datos que tenemos.

Si el valor de **Probability** es menor que 0.05, consideramos que la distribución no es normal.

Cuanto menor sea el valor de **Probability**, menos evidencia de que la serie tiene una distribución normal.

Por otro lado, con las órdenes:

View, Distribution Graphs,

podemos ver la función de distribución empírica de la serie,

CDF-Survivor-Quantile

y podemos ver el QQ plot de la serie,

QUANTILE-QUANTILE

2.2 Análisis conjunto

Lo primero que deberíamos realizar es abrir las series como un grupo:

Seleccionamos las variables (en orden) mediante la tecla CTRL y el botón izquierdo del ratón.

Hacemos doble clic en cualquier parte del área sombreada en la selección y pulsamos **Open Group** en el cuadro de diálogo resultante. Esta operación nos abre la ventana del objeto grupo deseado. Los gráficos y el análisis descriptivo individual y conjunto de las series se obtienen a partir de las representaciones (**Views**) que aparecen en la ventana del objeto grupo.

Así, por ejemplo, mediante **View/Multiple Graphs/Scatter/First variable against all** obtendríamos la representación de los pares de valores $j = 1, 2, 3, 4$, para el fichero “estres.txt” de la Figura 1.

En nuestro caso, los gráficos de la Figura 1 muestran una clara relación positiva lineal entre el grado de estrés (Y) y cada una de las variables consideradas, si bien parece que la relación es más fuerte con el tamaño de la empresa ($X1$) y el salario anual ($X3$).

Una visión más general de la relación entre el conjunto de variables consideradas se obtendría mediante **View/Multiple Graphs/Scatter/Matrix of all pairs (SCATMAT)**. En este caso la salida adopta la forma de una matriz en la que cada elemento recoge la representación gráfica de cada par de variables, incluidas también las variables explicativas (Figura 2).

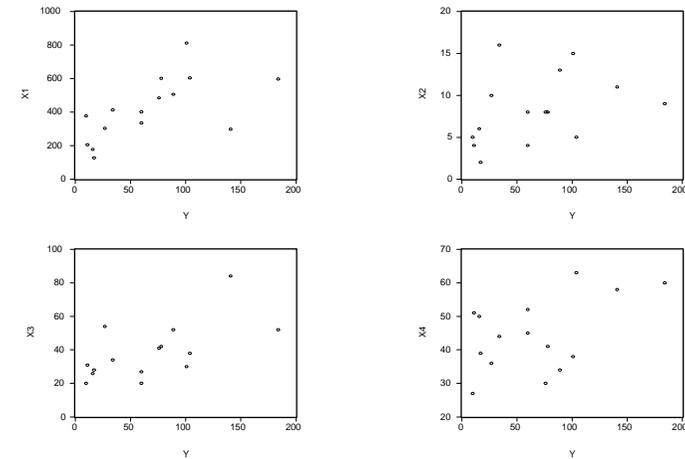


Figura 1:

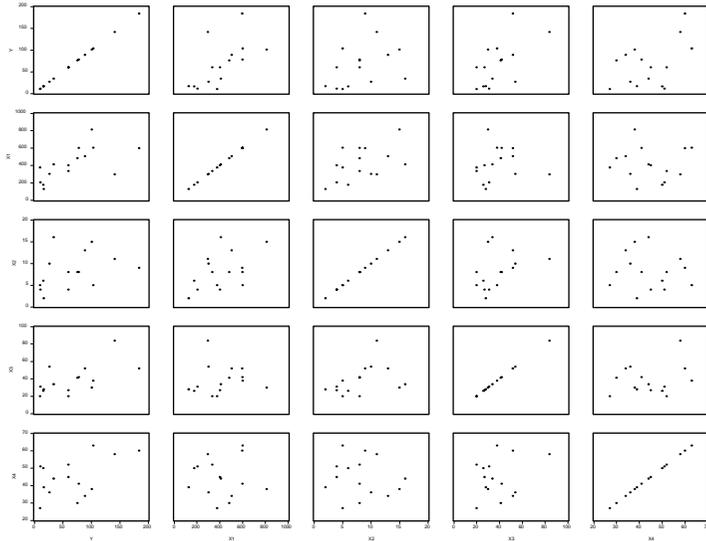


Figura 2:

El resumen numérico de la información que suministran estos gráficos se recoge en la matriz de correlaciones entre las variables, que se obtiene usando **View/Correlations**, y cuyo resultado para el ejemplo que estamos siguiendo es:

| | Y | X1 | X2 | X3 | X4 |
|----|----------|-----------|-----------|----------|-----------|
| Y | 1.000000 | 0.621569 | 0.355443 | 0.614634 | 0.485174 |
| X1 | 0.621569 | 1.000000 | 0.501089 | 0.088364 | -0.018855 |
| X2 | 0.355443 | 0.501089 | 1.000000 | 0.384154 | -0.113456 |
| X3 | 0.614634 | 0.088364 | 0.384154 | 1.000000 | 0.260760 |
| X4 | 0.485174 | -0.018855 | -0.113456 | 0.260760 | 1.000000 |

De esta matriz de correlaciones vemos que tanto $X1$ como $X3$ son las variables que presentan mayor correlación lineal con el nivel de estrés y que existe una cierta correlación lineal entre $X1$ y $X2$, así como entre $X2$ y $X3$.

Las características descriptivas básicas de cada serie se obtienen con **View/Descriptive Stats/Common Sample**. Los resultados para nuestro ejemplo son:

| | Y | X_1 | X_2 | X_3 | X_4 |
|----------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| Mean | 67.20000 | 415.7333 | 8.266667 | 38.60000 | 44.53333 |
| Median | 60.00000 | 401.0000 | 8.000000 | 34.00000 | 44.00000 |
| Maximum | 184.0000 | 812.0000 | 16.00000 | 84.00000 | 63.00000 |
| Minimum | 10.00000 | 127.0000 | 2.000000 | 20.00000 | 27.00000 |
| Std. Dev. | 51.16388 | 187.5132 | 4.148436 | 16.74515 | 10.94706 |
| Skewness | 0.755415 | 0.353083 | 0.431821 | 1.311686 | 0.143565 |
| Kurtosis | 2.869032 | 2.527598 | 2.237346 | 4.549391 | 1.981337 |
| Jarque-Bera Probability | 1.437350 0.487398 | 0.451147 0.798058 | 0.829700 0.660439 | 5.801683 0.054977 | 0.700074 0.704662 |
| Observations | 15 | 15 | 15 | 15 | 15 |

3 NOTAS SOBRE @-FUNCIONES

Los estadísticos que aparecen en los resultados de la estimación se almacenan temporalmente en @-funciones y pueden ser utilizados a la hora de generar series, escalares, vectores o matrices. Hay dos tipos de @-funciones, la que devuelven un escalar y las que contienen un vector o matriz. Véase la lista completa en el Help del programa).

3.1 DESCRIPTIVE STATISTICS FUNCTIONS

These functions compute descriptive statistics for the specified sample, excluding missing values if necessary. The default sample is the current work-file sample. You can specify a sample as the last argument of any descriptive statistic function, either as a string (in double quotes) or by the name of a sample object. For example, `@mean(x,"1945:01 1979:12")` or `@var(y,s2)` where `s2` is the name of a sample object.

To determine the number of observations available for a given series, use the `@obs` function. Note that where appropriate, EViews will perform casewise exclusion of data with missing values. For example, `@cov(x,y)` and `@cor(x,y)` will use only those observations for which data on both X and Y are available.

In the following table, arguments in square brackets [] are optional arguments:

- [s]: sample expression in double quotes or name of a sample object.

For `@quantile`, you must provide the method option argument in order to include a sample argument.

| Function | Description |
|----------------------|--|
| @cor(x,y[,s]) | returns the correlation between X and Y |
| @cov(x,y[,s]) | returns the covariance between X and Y |
| @inner(x,y[,s]) | returns the inner product of X and Y |
| @obs(x[,s]) | returns the number of non-missing observations for X in the current sample |
| @mean(x[,s]) | average of the values in X |
| @median(x[,s]) | computes the median of the X |
| @min(x[,s]) | minimum of the values in X |
| @max(x[,s]) | maximum of the values in X |
| @quantile(x,q[,m,s]) | returns the q-th quantile of the series X (*) |
| @stdev(x[,s]) | square root of the unbiased sample variance (division by $n - 1$) |
| @sum(x[,s]) | returns the sum of X |
| @sumsq(x[,s]) | sum of the squares of X |
| @var(x[,s]) | variance of the values in X (division by n) |

(*): m is an optional argument that specifies the following methods:

- 1: rankit (default)
- 2: ordinary
- 3: van der Waerden
- 4: Blom
- 5: Tukey

For example,

@quantile(x,0.5)

returns the median of X (using the default rankit method).

3.2 STATISTICAL DISTRIBUTION FUNCTIONS

The following functions provide access to the density or probability functions, cumulative distribution, quantile functions, and random number generators

for a number of standard statistical distributions.

There are four functions associated with each distribution. The first character of each function name identifies the type of function:

| Function Type | Beginning of Name |
|-------------------------------|-------------------|
| Cumulative distribution (CDF) | @c |
| Density or probability | @d |
| Quantile (inverse CDF) | @q |
| Random number generator | @r |

The remainder of the function name identifies the distribution. For example, the functions for the beta distribution are **@cbeta**, **@dbeta**, **@qbeta** and **@rbeta**.

When used with series arguments, EViews will evaluate the function for each observation in the current sample. As with other functions, **NA** or invalid inputs will yield **Not Available** values. For values outside of the support, the functions will return zero.

Note that the CDFs are assumed to be right-continuous: $F_X(k) = P(X \leq k)$. The quantile functions will return the smallest value where the CDF evaluated at the value equals or exceeds the probability of interest: $q_X(p) = q^* : F_X(q^*) \geq p$. The inequalities are only relevant for discrete distributions.

The information provided below should be sufficient to identify the meaning of the parameters for each distribution.

NAME OF THE DISTRIBUTION:

- **Beta:** **@cbeta(x,a,b)**, **@dbeta(x,a,b)**, **@qbeta(p,a,b)**, **@rbeta(a,b)**.
- **Binomial:** **@cbinom(x,n,p)**, **@dbinom(x,n,p)**, **@qbinom(s,n,p)**, **@rbinom(n,p)**.

- Chi-square: @cchisq(x,v), @dchisq(x,v), @qchisq(p,v), @rchisq(v).
- Exponential: @cexp(x,m), @dexp(x,m), @qexp(p,m), @rexp(m).
- Extreme Value (Type I): @cextreme(x), @dextreme(x), @qextreme(p), @cloglog(p), @rextreme.
- F-distribution: @cfdist(x,v1,v2), @dfdist(x,v1,v2), @qfdist(p,v1,v2), @rfdist(v1,v2).
- Gamma: @cgamma(x,r,s), @dgamma(x,r,s), @qgamma(p,r,s), @rgamma(r,s).
- Laplace: @claplace(x), @dlaplace(x), @qlaplace(x), @rlaplace.
- Logistic: @clogistic(x), @dlogistic(x), @qlogistic(p), @rlogistic.
- Lognormal: @clognorm(x,m,s), @dlognorm(x,m,s), @qlognorm(p,m,s), @rlognorm(m,s).
- Negative Binomial: @cnegbin(x,n,p), @dnegbin(x,n,p), @qnegbin(s,n,p), @rnegbin(n,p).
- Normal Distribution: @cnorm(x), @dnorm(x), @qnorm(p), @rnorm, nrnd.
- Pareto: @cpareto(x,a,k), @dpareto(x,a,k), @qpareto(p,a,k), @rpareto(a,k).
- Poisson: @cpoisson(x,m), @dpoisson(x,m), @qpoisson(p,m), @rpoisson(m).

- t-Distribution: @ctdist(x,v), @dtdist(x,v), @qtdist(p,v), @rt-dist(v).
- Uniform: @cunif(x,a,b), @dunif(x,a,b), @qunif(p,a,b), @runif(a,b), rnd.
- Weibull: @cweib(x,m,a), @dweib(x,m,a), @qweib(p,m,a), @rweib(m,a).

3.3 ADDITIONAL DISTRIBUTION FUNCTIONS

The following utility functions were designed to facilitate the computation of p -values for common statistical tests. While these results may be derived using the distribution functions above, they are retained for convenience and backward compatibility.

Function: @chisq(x,v)

Distribution: Chi-square.

Description: Returns the probability that a Chi-squared statistic with v degrees of freedom exceeds X :

$$1 - \text{@cchisq}(x,v).$$

Function: @fdist(x,v1,v2)

Distribution: F-distribution

Description: Probability that an F-statistic with $v1$ numerator degrees of freedom and $v2$ denominator degrees of freedom exceeds X :

$$1 - \text{@cfdist}(x,v1,v2).$$

Function: @tdist(x,v)

Distribution: t-distribution

Description: Probability that a t-statistic with v degrees of freedom exceeds X in absolute value (two-sided p -value):

$$2*(1-@ctdist(@abs(x),v)).$$

3.4 otras 2-funciones de interés

Algunas @-funciones escalares de interés:

@abs(x): valor absoluto de x

@ceiling(x): menor número entero no menor que x . Ejemplo:@ceiling(2.34)=3;

@ceiling(4)=4

@exp(x): exponencial de x

@fact(x): factorial, $x!$. Ejemplo: @fact(3)=6; @fact(0)=1

@floor(x): mayor número entero no mayor que x . Ejemplo:@floor(1.23)=1;

@floor(3)=3

@inv(x): inverso de x .

@log(x): logaritmo natural de x .

@round(x): redondeo al número entero más cercano a x . Ejemplo:@round(-97.5)=98; @round(3.5)=4

@sqrt(x): raíz cuadrada de x

Por ejemplo, si quisiéramos generar un objeto escalar, llamado χ , que recoja el número de observaciones multiplicado por el coeficiente de determinación de esta ecuación, escribiríamos en la ventana de comandos:

```
scalar chi=@regobs*@r2
```

y pulsamos INTRO. El resultado sería un nuevo icono escalar, en la ventana del Workfile seguido del nombre “ χ ”. Para ver su contenido hacemos doble clic sobre el mismo y **su valor se muestra en la parte izquierda de la línea de estado.**

Eviews, como siempre, solamente almacena el valor más reciente de estas @-funciones. No obstante, podemos darle nombre a la ecuación y guardarla. Por ejemplo, se le llamamos $eq01$, tendríamos que escribir

```
scalar chi=eq01.@regobs*eq01.@r2
```

y pulsar INTRO. Evidentemente, si tenemos dos o más ecuaciones guardadas, podemos combinar las órdenes.