

# Lección 1 — Producción de cemento

Marcos Bujosa

## Objetivo

Realizaremos una práctica similar a la anterior con los datos de *producción de cemento* descargados desde la base de datos del [Banco de España](#).

Guión: [P-L01-B-produccionCemento.inp](#)

### Objetivo

1. Repetir el ejercicio visto en clase pero con otros datos.
2. Tener en cuenta que al usar datos obtenidos de distintas fuentes, quizá sea necesario realizar algunos ajustes para que Gretl los lea.
3. Mostrar datos.
4. Transformarlos
5. Generar gráficos.

## Datos

### Producción de cemento en España

Producción de cemento (en miles de Toneladas métricas). Fuente Ministerio de Industria y Turismo (descargado desde la base de datos del [Banco de España](#)).

### Descarga de datos

Vaya a <https://www.bde.es/webbe/es/estadisticas/temas/estadisticas-economicas-generales.html>. En dicha página realice los siguientes pasos: Pulse en *Datos --> Producción y demanda --> Industria, energía y construcción --> ``Descargar``* (bajo Producción y consumo aparente de acero y cemento)

- Elija el formato *csv*.
- (Si lo necesita, copie el fichero que se ha descargado, *be2311.csv*, en la carpeta de datos de su proyecto)

### Carga de los datos en Gretl

Y ahora carguemos los datos:

**Archivo --> Abrir Archivo de datos --> Archivo de usuario.**

Se abrirá una ventana; abajo a la derecha verá un menú desplegable. Seleccione **Ficheros CSV (.csv)** y busque el fichero que ha descargado (*be3211.csv*) en el subdirectorio donde lo guardó.

Conteste **Si** o **No** a la primera pregunta.

INDEPENDIENTEMENTE DE LO QUE HAYA CONTESTADO OBTENDRÁ UN ERROR.

## Actividad 1 - Análisis de error de carga de los datos

Gretl espera un fichero formado por columnas y donde la primera fila contiene los nombres (entre comillas) de cada una de las columnas. El resto de filas solo contienen datos numéricos. Si hay datos ausentes, Gretl espera que esté indicado con NA.

Sin embargo el fichero `be3211.csv` que hemos descargado:

- tiene 6 filas con información no numérica.
- Los datos ausentes están marcados con "\_":

|                             |   |                             |
|-----------------------------|---|-----------------------------|
| CÓDIGO DE LA SERIE          | D_1ID10000  | D_1KC02000                  |
| NÚMERO SECUENCIAL           | 247145  | 247556                      |
| ALIAS DE LA SERIE           | BE_23_11.1  | BE_23_11.2                  |
| DESCRIPCIÓN DE LA SERIE     | Estadísticas generales. Producción interna de acero | Estadísticas generales. Exp |
| DESCRIPCIÓN DE LAS UNIDADES | Miles de toneladas métricas                         | Miles de toneladas métricas |
| FRECUENCIA                  | MENSUAL   | MENSUAL                     |
| ENE 1955                    | —   | —                           |
| FEB 1955                    | —   | —                           |
| MAR 1955                    | —   | —                           |

- Y las dos últimas filas también contienen información no numérica sobre los datos:

|          |   |  |          |
|----------|---|--|----------|
| ABR 2025 | —   | —  | —        |
| MAY 2025 | —   | —  | —        |
| JUN 2025 | —   | —  | —        |
| FUENTE   | UNESID  | UNESID                                       | UNESID   |
| NOTAS    | Estadísticas generales. Producción interna de acero | Estadísticas generales. Exportación de acero | Estadíst |

Otros problemas que pueden surgir se deben a las distintas codificaciones de los ficheros en las bases de datos (`iso-latin-1-dos`, `utf-8-unix`, etc.), que afectan a como se interpretan los caracteres no ASCII (acentos, eñes, etc.) en cada sistema operativo.

## Actividad 2 - Depuración del fichero .csv y lectura de los datos

- Borre las tres primeras líneas de texto (la que empieza con `CÓDIGO`, la que empieza con `NÚMERO` y la que empieza con `ALIAS`).
- En la cuarta línea cambie `"DESCRIPCIÓN DE LA SERIE"` por `"Fechas"` (como cabecera de la primera columna; que contiene la indicación del mes de cada dato).
- Acorte los nombres de la series borrando las cadenas de texto `Estadísticas generales._` en todos los nombres de las series:
  - Así por ejemplo `" Estadísticas generales. Producción interna de acero "` debe quedar como `" Producción interna de acero "`.
  - Repita esto con los nombres de todas las series (tal como aparecen son demasiado largos para que se visualicen correctamente en Gretl).
- Borre las dos líneas siguientes: la que describe las `UNIDADES` y la que describe la `"FRECUENCIA"`
- Borre también las dos últimas líneas del fichero: La que empieza con `"FUENTE"` y la que empieza con `"NOTAS"` al final del fichero.

- Sustituya todos los "\_" por NA (son muchos... hágalo usando la herramienta de sustitución de cadenas de texto que ofrezca su editor de texto).
- Guarde el fichero **be2311.csv** con todas esas modificaciones.

Ahora el comienzo de su fichero es:

| Fechas   | Producción interna de acero | Exportación de acero | Importación de acero | Disponibilidades de acero | Pr |
|----------|-----------------------------|----------------------|----------------------|---------------------------|----|
| ENE 1955 | NA                          | NA                   | NA                   | NA                        |    |
| FEB 1955 | NA                          | NA                   | NA                   | NA                        |    |

Y las filas finales solo contienen datos numéricos (exceptuando la primera columna):

|          |    |    |    |    |         |         |        |         |
|----------|----|----|----|----|---------|---------|--------|---------|
| ABR 2025 | NA | NA | NA | NA | 1434.52 | 246.75  | 52.812 | 1240.89 |
| MAY 2025 | NA | NA | NA | NA | 1617.17 | 284.805 | 53.084 | 1408.22 |
| JUN 2025 | NA | NA | NA | NA | 1712.92 | 299.159 | 53.873 | 1410.96 |

Además, los valores ausentes están indicados con NA.

## Carga de los datos en Gretl desde el fichero modificado

Y ahora ya podemos leer los datos con Gretl siguiendo los pasos: **Archivo -->Abrir Archivo de datos -->Archivo de usuario.**

- Se abrirá una ventana; en la esquina inferior derecha de la ventana seleccione **Ficheros CSV (.csv)**
- Busque su fichero modificado (**be3211.csv**) en el subdirectorio donde lo haya guardado.
- Conteste **Si** o **No** a la primera pregunta. Gretl no es capaz de interpretar lo que aparece en la primera columna como fechas. Conteste lo que conteste, cargará los datos inicialmente como si fueran de sección cruzada...pero, a continuación, nos preguntará si deseamos que se interpreten como series temporales o datos de panel.
- Conteste **Si**.
- Marque **Serie temporal** y pulse en **Adelante**.
- Indique que la frecuencia es **Mensual** y pulse en **Adelante**.
- Ajuste la fecha de la primera observación a enero de 1955; es decir: 1955:01 y pulse en **Adelante**.
- Finalmente pulse en **Aplicar**

*o bien teclee en línea de comandos:*

```
open ../datos/be2311.csv
setobs 12 1955:01 --time-series
```

Sustituya "**../datos/**" por la ruta al directorio donde ha guardado sus datos.

- Ayuda Para consultar la documentación sobre cualquier comando, puede emplear el menú desplegable **Ayuda** que aparece arriba, a la derecha de la ventana principal de **Gretl**.
  - **Ayuda ->Guía de Instrucciones** y "pinche" sobre **open**

*o bien teclee en línea de comandos:*

```
help open
```

## Guardar los datos en formato .gdt

Para no tener que repetir este proceso en el futuro con estos mismo datos, podemos guardarlos en el formato de Gretl. De esa manera podremos cargar los datos de manera inmediata en el futuro.

*o bien teclee en línea de comandos:*

```
store "../datos/be2311.gdt" --gzipped=1
```

Sustituya la cadena “ ../datos/ ” por la ruta al directorio donde quiere guardar sus datos.

## Actividad 3 - Gráfico de series temporales

### ■ Scatter plot

- Marque la variable **Producciondecemento** (pulsando ctrl y pinchando con el botón derecho del ratón sobre ella). Elija **Gráfico de series temporales**

*o bien teclee en línea de comandos:* `gnuplot Producciondecemento --time-series --with-lines`

### ■ Guardar gráfico como *icono* para editarlo más tarde

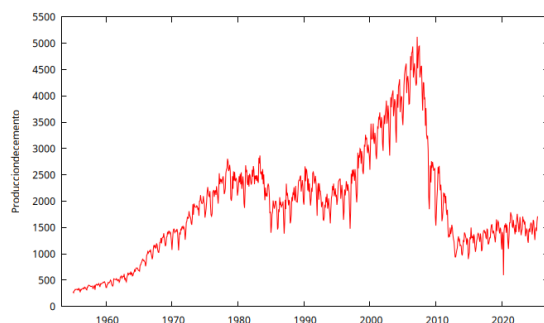
- “Pinche” con el botón derecho sobre la ventana del gráfico.
- Seleccione **Guardar a sesión como icono**

*o bien teclee en línea de comandos:*

```
Cemento <- gnuplot Producciondecemento --time-series --with-lines --output="Cemento.png"
```

*Cemento es el nombre con el que se guardará el icono*

(el comando `--output=` seguido de un nombre entre comillas es para que Gretl genere un fichero .png con el nombre indicado y que contenga la figura. Yo lo he añadido para poder insertar el gráfico en este documento; pero no es necesario para generar el gráfico ni el icono).



En la zona inferior izquierda de la ventana principal puede ver una serie de iconos. Uno de ellos es la **vista de iconos de sesión**.

## Actividad 4 - Gráfico rango media

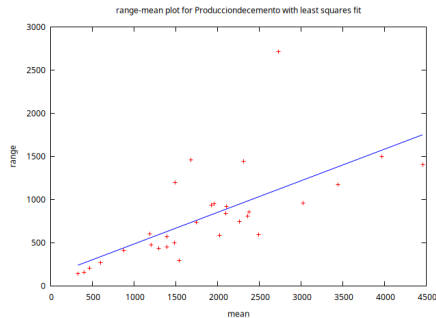
Para explorar si conviene tomar logaritmos porque la variabilidad de la serie crece con el nivel, realizamos un gráfico *rango-media*.

- En la ventana principal de **Gretl**, marque con el ratón la variable: **Producciondecemento**.
- Con la variable: **Producciondecemento** marcada, .en la parte superior ventana principal de **Gretl**, “Pinche” en el menú desplegable **Variable ->Gráfico rango-media**.
- Pulse **Aceptar** en la ventana emergente que se ha abierto.

- Aparecen dos ventanas, una con los detalles del rango y la media de las distintas sub-muestras. Otra con el gráfico *rango-media*. Pulse con el botón derecho del ratón sobre la figura si quiere guardar el gráfico como icono

*o bien teclee en línea de comandos:*

```
rmplot Producciondecemento --output="RangoMedia-Cemento.png"
```



Hay una significativa correlación positiva entre la volatilidad y el nivel de la serie y Gretl lo hace notar incluyendo la recta de regresión sobre la nube de puntos.

Por tanto, tomaremos logaritmos.

## Actividad 5 - Transformar logarítmicamente los datos

Seleccione con el ratón la variable **Producciondecemento** y luego pulse en el menú desplegable **Añadir** que aparece arriba, en el centro de la ventana principal de **Gretl**.

- **Añadir -> Logaritmos de las variables seleccionadas**

*o bien teclee en línea de comandos:*

```
logs Producciondecemento
```

Entre las variables aparecerá una nueva con el prefijo **l\_**, es decir, en este caso aparecerá la variable **l\_Producciondecemento**.

Genere el gráfico de series temporales de esta nueva serie y guárdelo como un nuevo icono (use un nombre suficientemente descriptivo para el icono, por ejemplo **LogsCemento**)

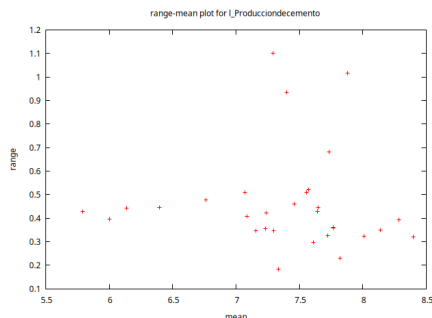
```
LogsCemento <- gnuplot l_Producciondecemento --time-series --with-lines
```

### Gráfico rango-media

Compruebe que la correlación entre el nivel y la variabilidad de los datos ha desaparecido con un gráfico *rango-media* de los datos transformados logarítmicamente.

*en línea de comandos:*

```
rmplot l_Producciondecemento --output="RangoMedia-LogCemento.png"
```



Ya no hay una significativa correlación entre la volatilidad y el nivel de la serie en logaritmos. En consecuencia Gretl ya no incluye la recta de regresión sobre la nube de puntos.

## Gráfico de los datos en logaritmos

### ■ Scatter plot

- Marque la variable `l_Producciondecemento` (pulsando `ctrl` y pinchando con el botón derecho del ratón sobre ella). Elija **Gráfico de series temporales**

*o bien teclee en línea de comandos:* `gnuplot l_Producciondecemento --time-series --with-lines`

### ■ Guardar gráfico como *icono* para editarlo más tarde

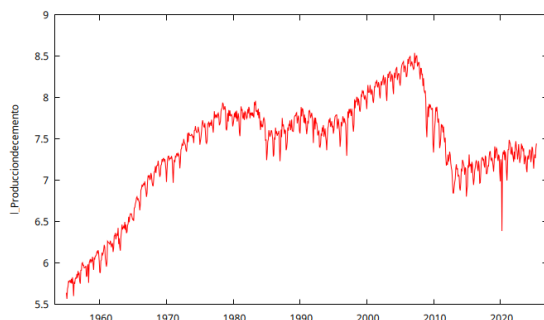
- “Pinche” con el botón derecho sobre la ventana del gráfico.
- Seleccione **Guardar a sesión como icono**

*o bien teclee en línea de comandos:*

```
LogCemento <- gnuplot l_Producciondecemento --time-series --with-lines --output="LogCemento.png"
```

*Cemento es el nombre con el que se guardará el icono*

(el comando `--output=` seguido de un nombre entre comillas es para que Gretl genere un fichero `.png` con el nombre indicado y que contenga la figura. Yo lo he añadido para poder insertar el gráfico en este documento; pero no es necesario para generar el gráfico ni el icono).



Claramente el nivel local de la serie evoluciona en el tiempo. Por tanto, es necesario tomar una primera diferencia de los datos. También se aprecia una pauta estacional, por lo que también será necesario tomar una diferencia estacional. Comencemos por la diferencia regular...

## Actividad 6 - Primera diferencia de los datos en logaritmos

Seleccione con el ratón la variable `l_Producciondecemento` y luego pulse en el menú desplegable **Añadir** que aparece arriba, en el centro de la ventana principal de **Gretl**.

### ■ **Añadir** -> **Primeras diferencias de las variables seleccionadas**

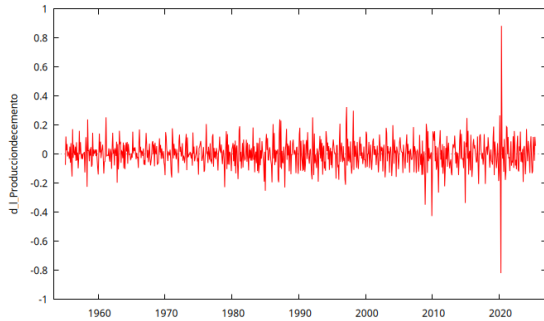
*o bien teclee en línea de comandos:*

```
diff l_Producciondecemento
```

Entre las variables aparecerá una nueva con el prefijo `d_`, es decir, en este caso aparecerá la variable `d_l_Producciondecemento`.

Genere el gráfico de series temporales de esta nueva serie y guárdelo como un nuevo icono (Use un nombre suficientemente descriptivo, por ejemplo `D_LogsCemento`)

```
D_LogCemento <- gnuplot d_l_Producciondecemento --time-series --with-lines --output="D_LogCemento.png"
```



Observe que en la serie obtenida ya muestra tendencia; pero si se aprecia el acusadísimo efecto del confinamiento del año 2020 (que aparece repetido y con el efecto contrario un periodo más tarde debido a la diferenciación regular)

## Actividad 7 - Diferencia de orden 12 (o estacional) de la primera diferencia de los datos en logaritmos

Seleccione con el ratón la variable `d_l_Producciondecemento` y luego pulse en el menú desplegable **Añadir** que aparece arriba, en el centro de la ventana principal de [Gretl](#).

- **Añadir ->Diferencias estacionales de las variables seleccionadas**

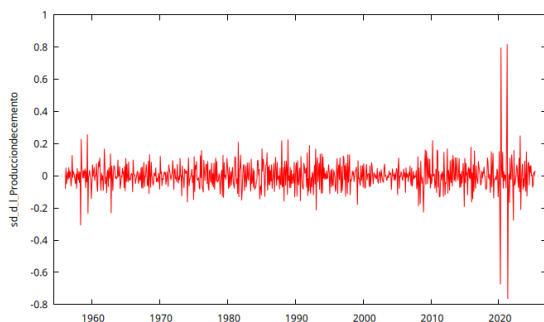
*o bien teclee en línea de comandos:*

```
sdiff d_l_Producciondecemento
```

Entre las variables aparecerá una nueva con el prefijo `sd_`, es decir, en este caso aparecerá la variable `sd_d_l_Producciondecemento`.

Genere el gráfico de series temporales de esta nueva serie y guárdelo como un nuevo icono (Use un nombre suficientemente descriptivo, por ejemplo `D12_D_LogsCemento`)

```
D12_D_LogsCemento <- gnuplot sd_d_l_Producciondecemento --time-series --with-lines --output="D12_D_LogCemento.png"
```



Es evidente que la serie resultante, tras aplicar una diferencia regular y otra estacional, ya no muestra ni tendencia ni componentes cíclicos estacionales. Se puede notar claramente el impacto significativo del confinamiento de 2020, que se manifiesta de manera repetida con un efecto opuesto un período más tarde (debido a la diferencia regular), y, posteriormente, ambos picos se repiten un año después como resultado de la diferenciación estacional.

Guión completo: [P-L01-B-produccionCemento.inp](#)