

Análisis de datos climáticos y de salud utilizando estadísticas descriptivas y herramientas de visualización

Rémi COUSIN, DL
IRI



Objetivos

- Aprender como datos epidemiológicos están organizados en la Data Library
- Aprender como visualizar series temporales de datos epidemiológicos y climáticos en la Data Library
- Aprender como construir histogramas de datos epidemiológicos en la Data Library
- Aprender como calcular una climatología mensual
- Aprender como calcular varias medidas estadísticas de datos epidemiológicos y climáticos en la Data Library, incluyendo media, mediana, desviación típica, error cuadrático medio

Estructura de los Datos Epidemiológicos en la Data Library

Estructura de los Datos Epidemiológicos en la Data Library

En la Data Library, datos epidemiológicos están generalmente clasificados en bases de datos por país y enfermedad.

Todas las bases de datos epidemiológicas utilizadas en este curso están localizadas en la base de datos Climate Information and Public Health:


**expert
home .ciph**

En esta página pueden ver bases de datos para varios países como Eritrea, Colombia, Madagascar.


Estructura de los Datos Epidemiológicos en la Data Library

Seleccionar “Eritrea” muestra “malaria” como sub-base de datos en la jerarquía de la Data Library:

expert
home .ciph .Eritrea



Data Library
Finding Data
Tutorial
Questions & Answers
Function Documentation



help

home ciph Eritrea options [Help](#) [Expert Mode](#)

[home](#) [Climate Information for Public Health](#) [Eritrea](#)

served from [IRI/LDEO Climate Data Library](#)

home ciph Eritrea

home ciph Eritrea.

Documents

[overview](#) an outline showing sub-datasets of this dataset

Datasets and variables

[CAD](#) home ciph Eritrea CAD[**Monthly**]

[malaria](#) home ciph Eritrea malaria[**monthly96-03 climatology96-03**]

Estructura de los Datos Epidemiológicos en la Data Library



Data Library

Finding Data
Tutorial
Questions & Answers
Function Documentation



help

home ciph Eritrea malaria monthly96-03 options [Help](#) [Expert Mode](#)

   [Data Selection](#) [Data Downloads & Files](#) [Data Tables](#)

served from [IRI/LDEO Climate Data Library](#)

[home](#) [Climate Information for Public Health](#) [Eritrea](#) [malaria](#) [monthly96-03](#)

home ciph Eritrea malaria monthly96-03

home ciph Eritrea malaria monthly96-03.

Documents

[outline](#) an outline showing all sub-datasets and variables contained in this dataset

Datasets and variables

- [Cases per month](#) home ciph Eritrea malaria monthly96-03 cases[T district]
- [Incidence](#) home ciph Eritrea malaria monthly96-03 incidence[T district]
- [MOH SubZobas](#) home ciph Eritrea malaria monthly96-03 MOH_SubZobas[**the_geom mohcode zonecode_2 mohname_2 zonename_2**]
- [Population](#) home ciph Eritrea malaria monthly96-03 population[T district]

Independent Variables (Grids)

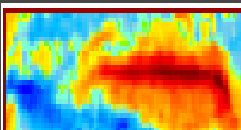

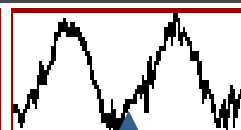
district grid: /district (ids) unordered [(ARETA) (MAKELAY KEYHI BAHRI) (DEBUB-DENKALIA) (ASSAB) (GHELAELO) (FORO) (DAHLAK) (MASSAWA) (GHINDAE) (SHIEB) (AFABET) (NAKFA) (KARORA) (ADOBHA) (ADI-TEKELEZAN) (ELABERED) (GELEB) (KEREN) (HAGAZ) (HALHAL) (HABERO) (ASMAT) (KERKEBET) (SELA) (HAMELMALO) (AGORDAT) (BARENTU) (DIGHE) (FORTO) (GOCHE) (ALAGCETA) (GOCO) (NGERA) (GEMCHI) (GOCOL) (GULID) (GIL) (DIBLO) (GELID) (TEGEMER) (GALAY) (GASH)

Visualizar Series Temporales de Datos Epidemiológicos

home ciph Eritrea malaria monthly96-03[cases incidence MOH SubZobas population]

expert
home .ciph .Eritrea .malaria .monthly96-03

OK
reset

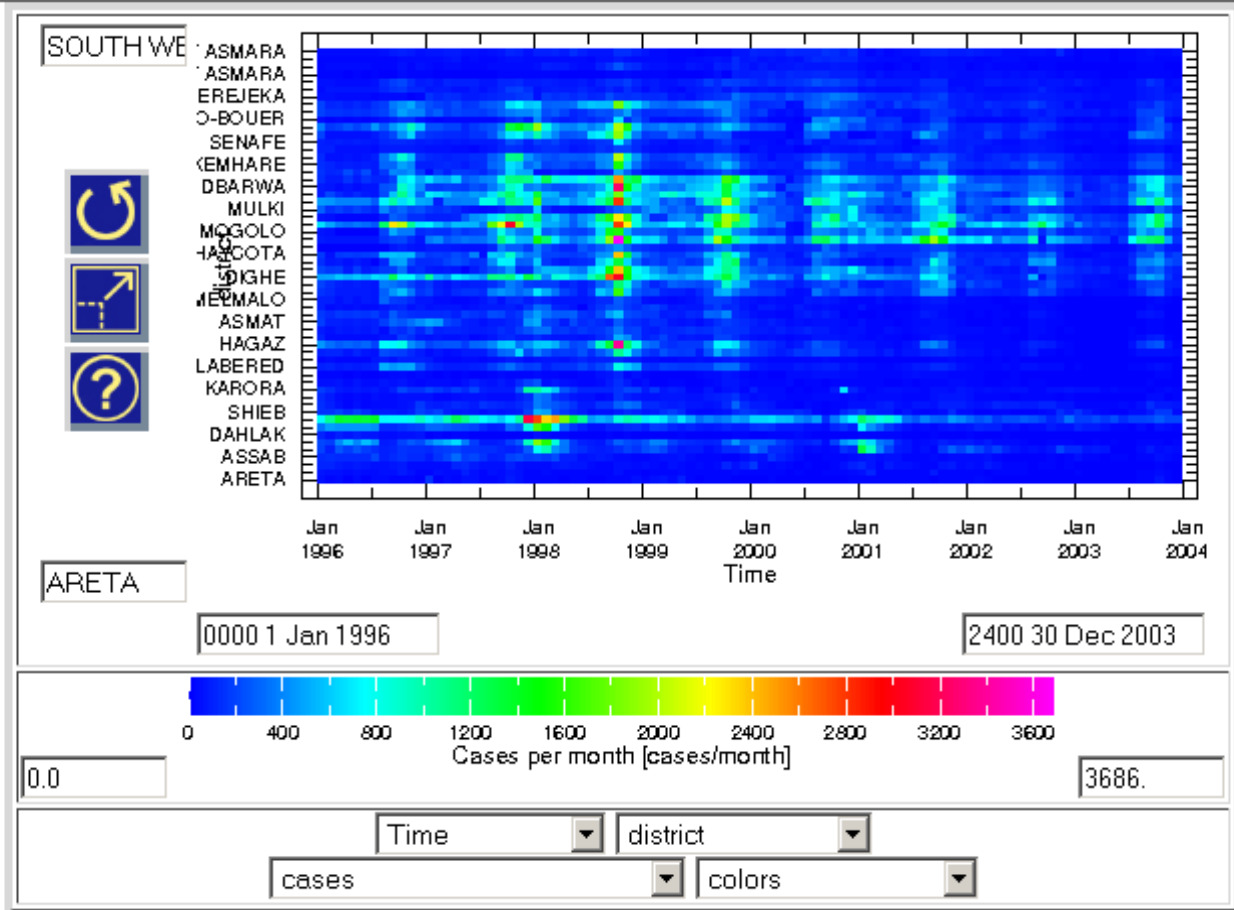
   [Data Selection](#) [Data Downloads & Files](#) [Data Tables](#)

Visualizar datos en gráfico 2D
(tiempo, distrito)

Visualizar datos en gráfico 1D
(ie, series temporales)

Visualizar Series Temporales de Datos Epidemiológicos

IRI Data Library home ciph Eritrea malaria monthly96-03 home ciph Eritrea malaria monthly96-03 Cases per month 1996-2003 ARETA - SOUTH WEST ASMARA



Visualizar Series Temporales de Datos Epidemiológicos

IRI

Data Library

home ciph Eritrea malaria monthly96-03

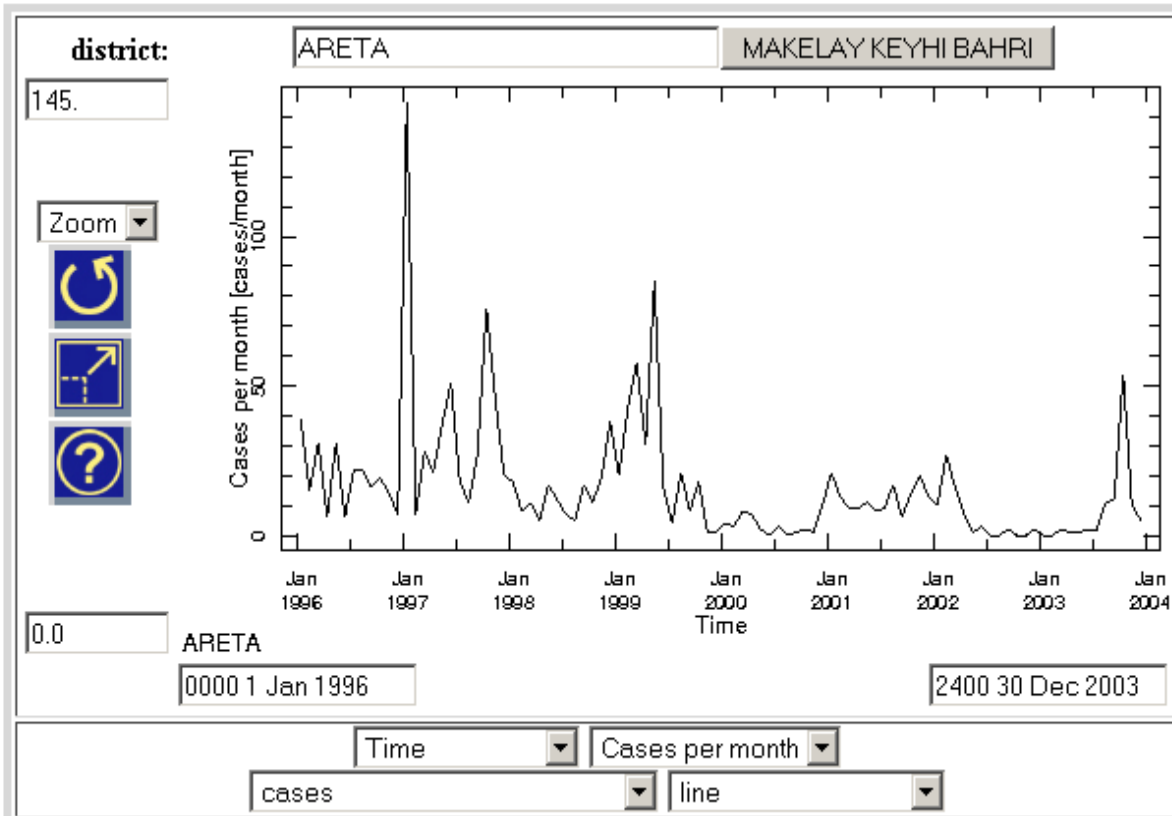
home ciph Eritrea malaria monthly96-03
Cases per month ARETA

1996-2003

ARETA - SOUTH WEST
ASMARA



Ministry of Health Sub-Zoba code	Ministry of Health Sub-Zoba name	Zone code	Zone name
	ids	ids	ids
101	ARETA	DK	SOUTHERN RED SEA



Construir Histogramas de Distribución de Datos

Construir Histogramas de Distribución de Datos

Histograma de Frecuencia

1. Un gráfico de una sola base de datos, agrupados en clases
2. Frecuencia está definida aquí como la cuenta de valores que pertenecen a cada clase
3. El histograma consiste en una serie de rectángulos de cuales la anchura está definida por las límites de las clases, y de cuales la altura está definida por la frecuencia de estas valores en este intervalo

Construir Histogramas de Distribución de Datos

Histograma de Frecuencia

4.El histograma ilustra varias características de los datos, incluyendo la posición, la dispersión, y la simetría

a. No hay reglamento en cuanto al número de clases o al intervalo de clase que utilizar

b. De 5 hasta 20 clases funciona para la mayoría de bases de datos

c. Anchuras de tamaño igual vienen de la división del rango por el número de clases

Construir Histogramas de Distribución de Datos

Ejercicio 1: Construir un histograma de frecuencia de datos de casos de malaria mensuales para un solo subzoba (distrito) en Eritrea en el periodo 1996-2003.

Etapas del Expert Mode:

1. Seleccionar una variable de una base de datos
2. Restrinja el rango de los dominios temporales, especiales, etc. de los datos al dominio que quiere
3. Utilice las funciones **DATA** y **RANGESTEP** para definir el rango y el tamaño de los intervalos de datos para el histograma
4. Aplique la función **distrib1D** para clasificar los datos en cuanto a los intervalos

Construir Histogramas de Distribución de Datos

Plantilla en Expert Mode:

Expert

dataset/variable selection

time/spatial domain restriction

DATA *lower upper step* RANGESTEP

distrib1D

Detalle de las Etapas:

En el ejemplo precedente, visualizamos una serie temporal de datos de casos de malaria mensuales para el subzoba de Areta en Eritrea. Vamos a visualizar estos datos como histograma de frecuencia.

Construir Histogramas de Distribución de Datos

Etapa 1: Seleccionar la base de datos de malaria de Eritrea y seleccionar la variable “Cases per month”:

```
home .ciph .Eritrea .malaria .monthly96-03 .cases
```

Etapa 2: Como queremos los datos para Areta, añade la función siguiente para seleccionar este distrito:

```
district (ARETA) VALUE
```

Etapas 3 & 4: Definir las clases que utilizar para construir el histograma.

Si miramos la serie temporal de Areta, la cuenta de casos por mes tiene valores aproximadamente entre 0 y 150.

Construir Histogramas de Distribución de Datos

Etapas 3 & 4: Definir las clases que utilizar para construir el histograma.

Si decidimos un intervalo de 10 casos por mes, necesitamos $150/10 = 15$ intervalos en el histograma, que es razonable. Cuando los intervalos están definidos, utilice los valores centrales de los intervalos “lower” y “upper” (5 and 145) para definirlos con la función RANGESTEP:

DATA 5 145 10 RANGESTEP

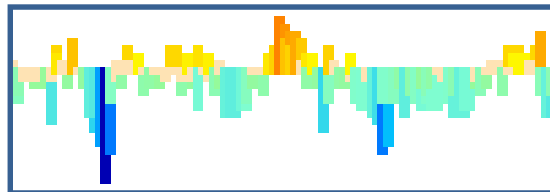
Finalmente, **distrib1D** utiliza los valores en RANGESTEP para calcular la distribución de la frecuencia de los datos, asignando frecuencias a intervalos definidos con una nueva retícula.

Construir Histogramas de Distribución de Datos

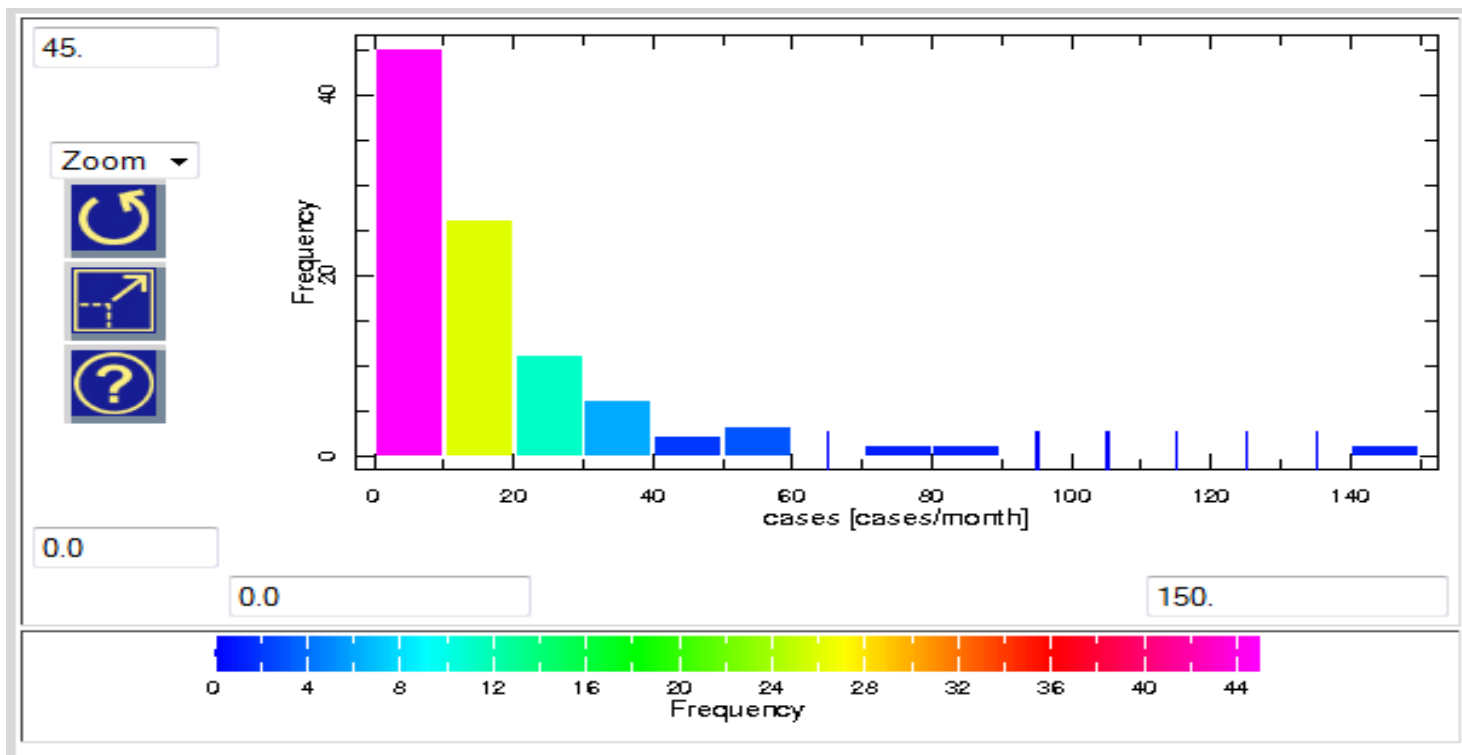
Al final, utilice la secuencia de funciones en Expert Mode para construir el histograma de frecuencia:

```
expert  
home .ciph .Eritrea .malaria .monthly96-03 .cases  
district (ARETA) VALUE  
DATA 5 145 10 RANGESTEP  
distrib1D
```

Note que una nueva retícula estaba creada que define los intervalos y tiene el mismo nombre que la variable original “cases”. Para visualizar el histograma, haga clic el icono que representa un histograma:



Construir Histogramas de Distribución de Datos



Ambos altura y color de las barras del histograma indican la frecuencia por la cual una cuenta de casos de malaria por mes ocurre en el subzoba de Areta durante el periodo 1996-2003.

Construir Histogramas de Distribución de Datos

Formas del Histograma

Unimodal: la distribución tiene un máximo único

Bimodal: tiene dos máximos distintos

Multimodal: tiene mas que dos máximos

Se puede describir el sesgo también en los histogramas

Sesgo Si una distribución no está simétrica en cuanto a su centro, está sesgada.

Construir Histogramas de Distribución de Datos

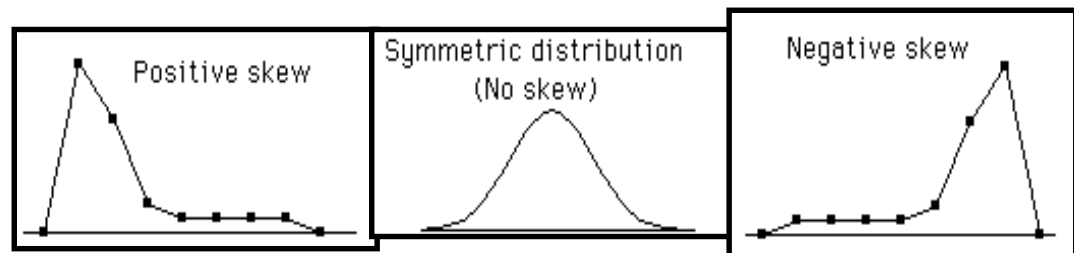
Sesgo

Características de distribuciones con sesgo negativo:

1. Cola a la izquierda
2. La mediana está siempre más grande que la media

Características de distribuciones con sesgo positivo:

1. Cola a la derecha
2. La media está siempre más grande que la mediana



Calcular la Tendencia Central

Una cantidad muy utilizada para describir una base de datos es su centro. El centro es un valor sólo, que da una aproximación razonable de la normalidad de los datos.

Hay varias maneras para aproximar el centro de una base de datos. Lo mas familiar es la media; sin embargo, utilizar solamente la media para aproximar la normalidad puede ser erróneo.

Para mejorar el entendimiento de lo que es normal, otras medidas de tendencia central, como la mediana, pueden ser utilizadas.

Calcular la Media

Lo siguiente son características de la media:

- Definida como la media aritmética de la base de datos
- Calculada adicionando todos los valores, pues dividiendo por el número de valores
- Una de las más simples medidas del centro que calcular
- Puede indicar una descripción incompleta de la tendencia central si no está acompañada por otras medidas
- Muy afectada por valores extremos (no está resistente a valores atípicas)

Calcular la Media

Ejercicio 2: Calcular la media temporal de los casos de malaria mensuales en Eritrea para el periodo de tiempo entero.

Etapas en Expert Mode:

1. Seleccionar una variable de una base de datos
2. Restrinja los dominios espaciales y temporales como quiere
3. Calcular la media sobre la retícula de tiempo

Calcular la Media

Plantilla en Expert Mode:

```
expert  
dataset/variable selection  
time/spatial domain restriction  
[grid]average
```

Detalle de las Etapas:

Etapas 1:

Vuelva a la base de datos de Eritrea y seleccione los datos de casos de malaria mensuales en Expert Mode:

```
home .ciph .Eritrea .malaria .monthly96-03  
.cases
```


Calcular la Media

Etapa 2: La retícula de tiempo para esta base de datos se extiende desde Enero 1996 hasta Diciembre 2003.

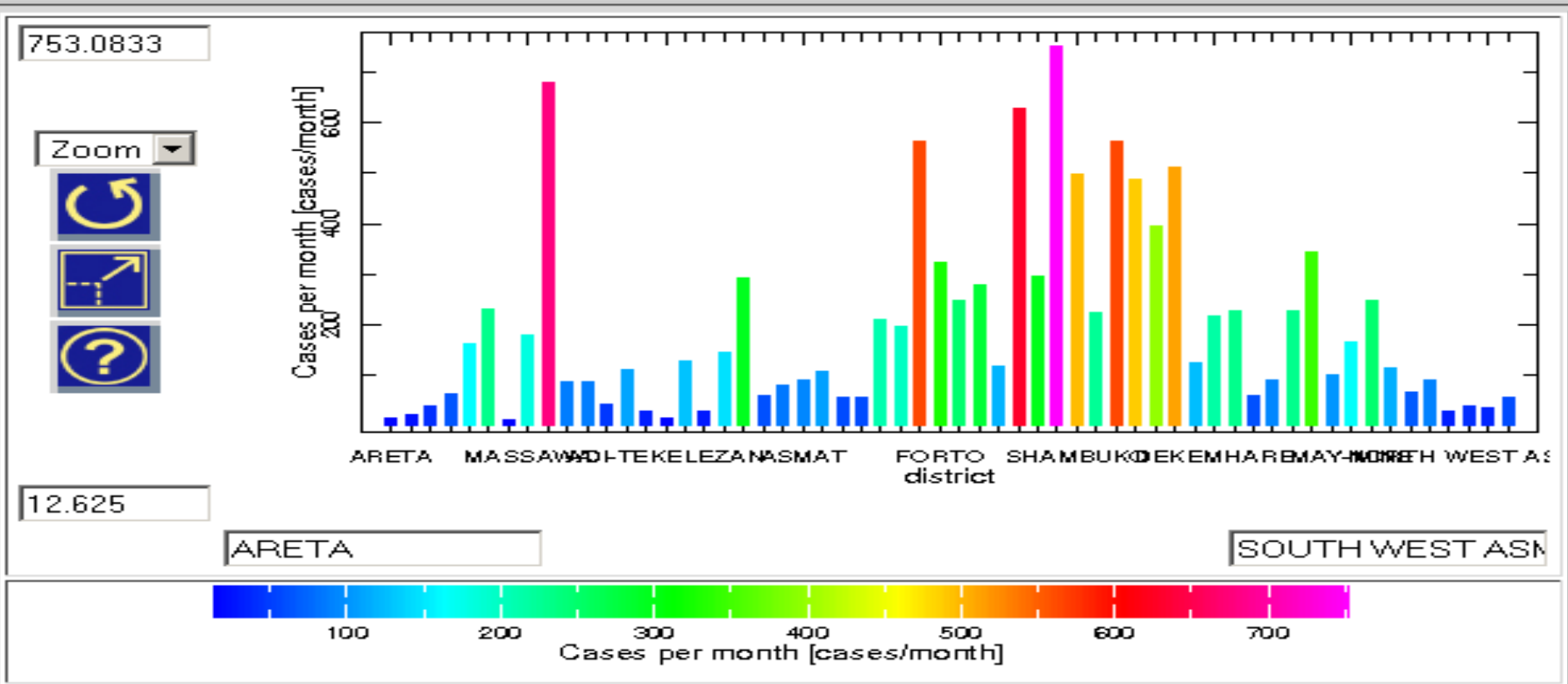
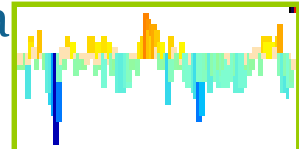
Etapa 3: Para calcular la media de los casos mensuales sobre este rango de tiempo (sin discriminar meses o estaciones), haga clic el lazo “Filters” y seleccione “Average over T”.

Haciéndolo, se vuelve a la página de base de datos de casos de malaria, pero con el resultado de la media sobre el tiempo. La media según el tiempo de los casos mensuales fue calculada para cada distrito. En Expert Mode, se traduce:

```
expert  
home .ciph .Eritrea .malaria .monthly96-03 .cases  
[T]average
```

Calcular la Media

Un histograma de la media de los casos por distritos puede ser construido haciendo clic el icono que representa un histograma



Calcular la Media

Table of

1. district
2. mean [home ciph Eritrea malaria monthly96-03 cases]

Additional Information

district ids	Cases per month cases/month
ARETA	16.11458
MAKELAY KEYHI BAHRI	20.82292
DEBUB-DENKALIA	38.45833
ASSAB	63.65625
GHELAELO	161.7396
FORO	229.8021
DAHLAK	12.625
MASSAWA	178.7604
GHINDAE	679.
SHIEB	86.10416
AFABET	86.54166
NAKFA	44.10417
KARORA	110.6042
ADOBHA	28.84375
ADI-TEKELEZAN	15.14583
ELABERED	128.6667

Calcular la Media

Ejercicio 3: Calcular una climatología mensual (media sobre múltiples años de valores mensuales) de la incidencia de malaria mensual en Eritrea.

En el ejercicio precedente, calculamos la media de largo plazo utilizando valores para el dominio temporal entero. En este ejercicio, la climatología mensual es la media de largo plazo para cada mes del año.

Etapas en Expert Mode:

1. Seleccione una variable de una base de datos
2. Restringa a los dominios especial, temporal, etc. que quiere
3. Aplique la función **yearly-climatology** a los datos mensuales

Calcular la Media

Plantilla en Expert Mode:

expert
dataset/variable selection
time/spatial domain restriction
yearly-climatology

Detalle de las Etapas:

Etapa 1:

Vuelva a la base de datos de malaria de Eritrea y seleccione los datos de incidencia de malaria mensual en Expert Mode:

home .ciph .Eritrea .malaria .monthly96-03 .incidence

Calcular la Media

Etapa 2: La retícula de tiempo para esta base de datos se extiende desde Enero 1996 hasta Diciembre 2003

Etapa 3: Para calcular la climatología mensual de la incidencia de malaria mensual sobre este dominio temporal, haga clic el lazo “Filters” y seleccione “Monthly Climatology”.

Haciéndolo, se vuelve a la página de base de datos de incidencia de malaria, pero con el resultado del calculo. La media según los años de las incidencias mensuales fue calculada para cada mes del año y para cada distrito.

En la sección “Grids”, note que hay una nueva retícula temporal con 12 puntos – uno para cada mes del año.

Calcular la Media

En Expert Mode se traduce:

expert

home .ciph .Eritrea .malaria .monthly96-03

.incidence

yearly-climatology

yearly-climatology es la función de la Data Library que calcula climatologías mensuales. Como entrada, requiere una variable que depende de una retícula temporal mensual que debe llamarse "T". Este cálculo puede ser efectuado directamente en la ventana del Expert Mode.

Calcular la Media

Este gráfico de la incidencia de la malaria muestra que muchos distritos tienen un pico de casos de incidencia de malaria por Octubre; otros tienen un pico por el principio del año; y algunos tienen ambos. Hay claramente una estacionalidad en la aparición de la malaria en Eritrea.

Porque la población puede cambiar de manera importante entre los subzobas de Eritrea, la incidencia puede mostrar la estacionalidad de la malaria mejor que los casos mismos.

Si miramos los casos, la estacionalidad temporal de la malaria habría podido ser escondida por la variabilidad espacial de la población entre subzobas.

Calcular la Mediana

La mediana es otra medida de la tendencia central de una serie de valores. Características de la mediana siguen:

- Una medida robusta y resistente de la tendencia central
- Definida como el valor a la mitad (o al centro) cuando las observaciones están ordenadas desde el valor menor hasta el valor mayor
- Divide la base de datos en dos partes de tamaño igual, con 50% de los valores menores que la mediana y 50% mayores que ella
- También llamada el 50 percentil
- Insensible a valores extremos

Calcular la Mediana

Ejercicio 4: Calcular la mediana de los casos de malaria en Eritrea según el periodo temporal entero

Etapas en Expert Mode:

1. Seleccione una variable de una base de datos
2. Restrinja los dominios temporal, espacial, etc. que quiere
3. Aplique la función **medianover** a los datos según la retícula de tiempo

Calcular la Mediana

Plantilla en Expert Mode:

```
expert  
dataset/variable selection  
time/spatial domain restriction  
[grid] medianover
```

Detalle de las Etapas:

Etapa 1:

Vuelva a la base de dato de malaria de Eritrea y seleccione los datos de casos de malaria mensuales en Expert Mode:

```
expert  
home .ciph .Eritrea .malaria .monthly96-03 .cases
```

Calcular la Mediana

Etapa 2:

Utilice el dominio temporal entero de la base de datos

Etapa 3:

Para calcular la mediana de los casos de malaria mensuales para el periodo desde Enero 1996 hasta Diciembre 2003 para cada distrito, aplique la función **medianover** según la retícula de tiempo. En Expert Mode, se traduce:

```
expert  
home .ciph .Eritrea .malaria .monthly96-03 .cases  
[T] medianover
```

Calcular la Mediana

Table of

1. district
2. home ciph Eritrea malaria monthly96-03 Cases per month 0.5

[Additional Information](#)

district	Cases per month
ids	cases/month
ARETA	10.375
MAKELAY KEYHI BAHRI	15.5
DEBUB-DENKALIA	32.5
ASSAB	37.
GHELAELO	120.
FORO	116.5
DAHLAK	4.933333
MASSAWA	90.8
GHINDAE	560.
SHIEB	70.
AFABET	73.33334
NAKFA	29.
KARORA	52.99999
ADOBHA	17.
ADI-TEKELEZAN	10.66667
ELABERED	80.66666

Table of

1. district
2. mean [home ciph Eritrea malaria monthly96-03 cases]

[Additional Information](#)

district	Cases per month
ids	cases/month
ARETA	16.11458
MAKELAY KEYHI BAHRI	20.82292
DEBUB-DENKALIA	38.45833
ASSAB	63.65625
GHELAELO	161.7396
FORO	229.8021
DAHLAK	12.625
MASSAWA	178.7604
GHINDAE	679.
SHIEB	86.10416
AFABET	86.54166
NAKFA	44.10417
KARORA	110.6042
ADOBHA	28.84375
ADI-TEKELEZAN	15.14583
ELABERED	128.6667

Medidas de Dispersión

Además del cálculo del centro (es decir la media o la mediana) de una serie de observaciones, es interesante también de caracterizar la dispersión de los valores por el centro.

Varias medidas de dispersión pueden ser calculadas para una base de datos. Vemos aquí como calcular el **rango**, las **anomalías cuadráticas medias**, y la **desviación típica** en la Data Library.

Calcular el Rango en la Data Library

El rango de una serie de valores es la diferencia entre el valor mayor y el menor.

Características del rango:

- Definido como la diferencia entre el valor mayor y el menor
- Una medida de la variabilidad muy simple que calcular
- Sin embargo, depende solamente de valores extremos y da ninguna información en cuanto a la distribución de lo demás de los datos.

Calcular el Rango en la Data Library

Ejercicio 5: Calcular el rango de la incidencia de malaria en Octubre en Eritrea para los años 1996-2003 para todos los subzobas

Etapas en Expert Mode:

1. Seleccione una variable de una base de datos
2. Restrinja el dominio temporal, espacial, etc. que quiere
3. Calcule el valor máximo según una retícula o más
4. Seleccione otra copia de los datos originales
5. Restrinja el dominio temporal, espacial, etc. que quiere (como en Etapa 2)
6. Calcule el valor mínimo según una retícula o más
7. Sustraiga el valor mínimo del valor máximo

Calcular el Rango en la Data Library

Plantilla en Expert Mode:

```
expert  
dataset/variable selection  
time/spatial domain restriction  
[grid] maxover  
dataset/variable selection  
time/spatial domain restriction  
[grid]minover  
sub
```

Calcular el Rango en la Data Library

Detalle de las Etapas:

Etapa 1:

Vuelva a la base de datos de malaria de Eritrea y seleccione los datos de incidencia de malaria mensual en Expert Mode:

expert

home .ciph .Eritrea .malaria .monthly96-03 .incidence

Etapa 2:

En este ejercicio, queremos solamente los valores de Octubre. Para seleccionar en la retícula de tiempo, haga clic el lazo “Data Selection” en la barra de función por la parte superior de la página.

Calcular el Rango en la Data Library

En la página siguiente, hay una ventana de texto en la sección “Setting Ranges” donde se puede leer “Jan 1996 to Dec 2003”.

Para seleccionar los meses de Octubre solamente, sustituya al texto existente las tres primeras letras de Octubre en ingles, con la primera letra en mayúscula: “Oct”.

Si quiere también indicar el rango de los años, puede utilizar “Oct 1996-2003”, por ejemplo.

Calcular el Rango en la Data Library

Cuando la selección temporal está correcta, haga clic el botón “Restrict Ranges”. Haciéndolo, la información sobre la retícula de tiempo en la ventana griega por la parte superior de la página cambia según la selección que ha hecho. Asegúrese que el texto para la retícula de tiempo se lee como lo siguiente:

The current settings for the grids are

- grid: /T (months since 1960-01-01) ordered (Oct 1996) to (Oct 2003) by 12. N= 8 pts :grid
- grid: /district (ids) unordered [(ARETA) (MAKELAY KEYHI BAHRI) (DEBUB-DENKALLIA) ... (SOUTH WEST ASMARA)] N= 58 pts :grid

If this is what you want, choose

Stop Selecting

Setting Ranges

If you want to restrict the range along a grid, choose here.

name	range
T Time	Oct
<input type="button" value="Restrict Ranges"/>	

Calcular el Rango en la Data Library

Si la información está correcta, haga clic el botón “Stop Selecting” para volver a la página principal de la base de datos con su selección temporal.

Además, note que la selección del tiempo para Octubre aparece en la ventana del Expert Mode:

```
expert  
home .ciph .Eritrea .malaria .monthly96-03 .incidence  
T (Oct) VALUES
```

T (Oct) VALUES es el código en Ingrid para seleccionar todos los valores de Octubre en la base de datos. La opción que incluye el rango de los años es:

```
T (Oct 1996-2003) VALUES
```

Calcular el Rango en la Data Library

Etapa 3:

Ahora que la selección temporal esta aplicada a la base de datos, calcule los valores máximo y mínimo (en este orden), y pues sustráigalos para calcular el rango.

Para calcular el valor máximo, haga clic el lazo “Filters” en la barra de función por la parte superior de la página, y seleccione “Maximum over T”.

Haciéndolo, se vuelve a la página principal de la base de datos, y en la ventana del Expert Mode, se puede leer:

[T] maxover

Calcular el Rango en la Data Library

Eso es la función de Ingrid para calcular el valor máximo según la retícula de tiempo. Note que la retícula de tiempo desapareció – la variable no depende tampoco del tiempo.

Etapa 4 & 5:

Para calcular el mínimo, hay que referenciar de nuevo la incidencia de malaria para los meses de Octubre:

```
home .ciph .Eritrea .malaria .monthly96-03 .incidence  
T (Oct) VALUES
```

copiando este texto en la ventana de Expert Mode debajo del texto **[T] maxover**, como sigue:

Calcular el Rango en la Data Library

expert

home .ciph .Eritrea .malaria .monthly96-03 .incidence

T (Oct) VALUES

[T]maxover

home .ciph .Eritrea .malaria .monthly96-03 .incidence

T (Oct) VALUES

Etapa 6:

Tenemos de nuevo la incidencia de malaria para los meses de Octubre que manipular. El proceso para calcular el mínimo es similar que el que calcula el máximo. Finalmente:

Calcular el Rango en la Data Library

expert

home .ciph .Eritrea .malaria .monthly96-03 .incidence

T (Oct) VALUES

[T]maxover

home .ciph .Eritrea .malaria .monthly96-03 .incidence

T (Oct) VALUES

[T]minover

Etapa 7:

Ahora tenemos dos variables que manipular – el máximo y el mínimo de la incidencia de malaria en los meses de Octubre, en este orden.

Para sustraer el mínimo del máximo para calcular el rango, la función que utilizar es **sub**:

Calcular el Rango en la Data Library

```
expert
```

```
home .ciph .Eritrea .malaria .monthly96-03 .incidence
```

```
T (Oct) VALUES
```

```
[T]maxover
```

```
home .ciph .Eritrea .malaria .monthly96-03 .incidence
```

```
T (Oct) VALUES
```

```
[T]minover
```

```
sub
```

Otra opción para copiar la selección de la incidencia y de su dominio temporal es utilizar la construcción **a: :a: :a**. En este caso, el código en Expert Mode es:

```
expert
```

```
home .ciph .Eritrea .malaria .monthly96-03 .incidence
```

```
T (Oct) VALUES
```

```
a: [T]maxover :a: [T]minover :a
```

```
sub
```

Calcular el Rango en la Data Library

Table of

1. district
2. [max - min] [home ciph Eritrea malaria monthly96-03 incidence]

Additional Information

district	Incidence
ids	cases/month per thousand people
ARETA	3.86
MAKELAY KEYHI BAHRI	4.73
DEBUB-DENKALIA	7.56
ASSAB	8.71
GHELAELO	5.29
FORO	5.55
DAHLAK	12.45
MASSAWA	7.55
GHINDAE	11.71
SHIEB	2.56
AFABET	2.54
NAKFA	2.36
KARORA	4.3
ADOBHA	5.5
ADI-TEKELEZAN	2.84
ELABERED	16.09
GELEB	3.07
KEREN	13.97

Calcular la RMSA en la Data Library

La anomalía cuadrática media (RMSA para Root Mean Square Anomaly en inglés) es otra medida de la dispersión de los datos por una valor central.

Características de la RMSA incluyen:

- La RMSA no está robusta o resistente a los valores atípicos y puede ser errónea cuando aplicada a distribuciones sesgadas.
- Muy parecida a la desviación típica, sino que está utilizada para bases de datos de grande tamaño (es decir, el divisor es N en lugar de $N-1$)

- La RMSA está calculada utilizando la formula siguiente:

$$S_N = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}$$

Donde \bar{x} es la media, x_i es cada valor de los datos, and N es el número de observaciones

El termino $x_i - \bar{x}$ calcula las anomalías

Se utiliza a menudo como medida de error

Calcular la RMSA en la Data Library

Ejercicio 6: Calcular la anomalía cuadrática media (RMSA) de la incidencia de malaria por distritos para el periodo desde Enero 1996 hasta Diciembre 2003

Etapas en Expert Mode:

1. Seleccione una variable de una base de datos
2. Restrinja los dominios temporales, espaciales, etc. que quiere
3. Calcule la anomalía cuadrática media (RMSA) según la retícula de tiempo utilizando **rmsaover**

Calcular la RMSA en la Data Library

Plantilla en Expert Mode:

```
expert  
dataset/variable selection  
time/spatial domain restriction  
[grid] rmsaover
```

Detalle de las Etapas:

Etapas 1:

Vuelva de nuevo a la base de datos de malaria de Eritrea y seleccione los datos de incidencia de malaria mensual en Expert Mode:

```
expert  
home .ciph .Eritrea .malaria .monthly96-03 .incidence
```


Calcular la RMSA en la Data Library

Etapa 2:

Utilice el dominio temporal entero de la base de datos

Etapa 3:

La anomalía cuadrática media hace parte de las opciones de la página “Filters”. Haga clic el lazo “Filters”, y pues seleccione la opción “RMSA (root mean square with mean removed) over T” (asegúrese de que seleccionó “T” sólo). Esta selección vuelve a la página principal de la base de datos, con el calculo de la RMSA para el periodo desde Enero 1996 hasta Diciembre 2003 para cada subzoba.

Calcular la RMSA en la Data Library

Table of

1. district
2. root mean sq anom [home ciph Eritrea malaria monthly96-03 incidence]

Additional Information

district	Incidence
ids	cases/month per thousand people
ARETA	1.053626
MAKELAY KEYHI BAHRI	1.347368
DEBUB-DENKALIA	2.673937
ASSAB	3.110789
GHELAELO	7.973109
FORO	6.3288
DAHLAK	6.380796
MASSAWA	7.411683
GHINDAE	8.675712
SHIEB	1.534154
AFABET	0.6972591
NAKFA	0.9831892
KARORA	4.158183
ADOBHA	1.59148
ADI-TEKELEZAN	0.4860355
ELABERED	2.986312
GELEB	0.8042555
KEREN	2.169164

Calcular la Desviación Típica en la Data Library

Características de la desviación típica incluyen:

- La desviación típica es la raíz cuadrada de la varianza
- La desviación típica no está robusta o resistente a valores atípicos, y puede ser errónea cuando calculada para distribuciones de datos sesgadas
- Permite describir la varianza de los datos

Calcular la Desviación Típica en la Data Library

- la desviación típica está calculada según la formula siguiente:

$$S_{N-1} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}$$

Note que la formula es cuasi la misma que la de la RMSA, sino que el denominador en la división es “N – 1” en lugar de “N”.

Calcular la Desviación Típica en la Data Library

Suponiendo que la distribución de los datos es una distribución normal, la desviación típica da la información siguiente:

1. La media ± 1 desviación típica contiene aproximadamente 68% de las observaciones en la serie de datos
2. La media ± 2 desviación típica contiene aproximadamente 95% de las observaciones en la serie de datos
3. La media ± 3 desviación típica contiene aproximadamente 99.7% de las observaciones en la serie de datos

Calcular la Desviación Típica en la Data Library

Ejercicio 7: Calcular la desviación típica de la incidencia de malaria para cada distrito para el periodo desde Enero 1996 hasta Diciembre 2003. Aplicando la desviación típica en este caso a una distribución sesgada, tiene que acordárselo al interpretar el resultado.

Etapas en Expert Mode:

1. Seleccione una variable de una base de datos
2. Restrinja los dominios temporales, espaciales, etc. que quiere
3. Utilice **dataflag** y **sum** para sumar el número de observaciones que no están faltantes

Calcular la Desviación Típica en la Data Library

4. Copia la cuenta utilizando **dup** y sustraiga 1. utilizando **sub** para calcular el grado de libertad (N-1)
5. Divida la cuenta de observaciones (N) por el grado de libertad utilizando **div**
6. Calcule la raíz cuadrada de este cociente utilizando **sqrt**
7. Seleccione la variable original de nuevo
8. Restrínjala de misma manera que en la etapa 2
9. Calcule la RMSA según la retícula de tiempo aplicando **rmsaover**
10. Multiplique utilizando **mul**

Calcular la Desviación Típica en la Data Library

Plantilla Expert Mode:

```
expert  
dataset/variable selection  
time/spatial domain restriction  
dataflag  
[grid]sum  
dup  
1 sub  
div  
sqrt  
dataset/variable selection  
time/spatial domain restriction  
[grid]rmsaover  
mul
```


Calcular la Desviación Típica en la Data Library

Detalle de las Etapas:

Etapa 1:

Vuelva de nuevo a la base de datos de malaria de Eritrea y seleccione los datos de incidencia de malaria mensual en Expert Mode:

```
expert  
home .ciph .Eritrea .malaria .monthly96-03  
.incidence
```

Etapa 2:

Utilice el periodo de tiempo entero de la base de datos

Calcular la Desviación Típica en la Data Library

Etapas 3-10:

La sola diferencia entre el cálculo de la RMSA y el de la desviación típica es el denominador $N-1$ (en lugar de N). A pesar de que esta diferencia cambia poco el resultado para bases de datos de grande tamaño, es importante para bases de datos más pequeñas.

El cálculo de la desviación típica utiliza la función **rmsaover** y corrige el denominador en otra parte del código. Aquí está el código para calcular la desviación típica en este ejercicio:

Calcular la Desviación Típica en la Data Library

```
expert
home .ciph .Eritrea .malaria .monthly96-
    03 .incidence
dataflag
[T]sum
dup
1. sub
div
sqrt
home .ciph .Eritrea .malaria .monthly96-
    03 .incidence
[T]rmsaover
mul
```

Calcular la Desviación Típica en la Data Library

```
expert  
home .ciph .Eritrea .malaria .monthly96-03 .incidence  
dataflag  
[T]sum  
dup  
1. sub  
div  
sqrt
```


La primera parte del código calcula este factor  que debe ser multiplicado al resultado de **[T]rmsaover** (que incluye un denominador N) para finalizar el cálculo. Así, **dataflag** y **[T]sum** calculan N, **dup** duplica N, **1. sub** sustrae 1, **div** calcula el cociente, y **sqrt** calcula la raíz cuadrada del cociente.

Table of

1. district
2. $\sqrt{\text{total dataflag (home ciph Eritrea malaria monthly96-03 incidence) / (* root mean sq anom [home ciph Eritrea malaria monthly96-03 incidence])}}$

Additional Information

district	sqrt[dataflag[Incidence]]
ids	unitless cases/month per thousand people
ARETA	1.059157
MAKELAY KEYHI BAHRI	1.354441
DEBUB-DENKALIA	2.687974
ASSAB	3.127119
GHELAELO	8.014962
FORO	6.362022
DAHLAK	6.414291
MASSAWA	7.450589
GHINDAE	8.721253
SHIEB	1.542207
AFABET	0.7009193
NAKFA	0.9883503
KARORA	4.18001
ADOBHA	1.599835
ADI-TEKELEZAN	0.4885868
ELABERED	3.001988
GELEB	0.8084774

Calcular Medias de Datos Climáticos sobre Distritos

La Data Library tiene capacidad similar a la de Sistemas de Información Geográfica (SIG), como utilizar juntamente datos raster (en retícula) y datos vectoriales. Una manera es de reticular regiones definidas por fronteras políticas (o otras) y calcular medias ponderadas de los datos climáticos en retícula que pertenecen a estas regiones.

Como los datos sanitarios y de enfermedades están generalmente colectados por regiones administrativas, o distritos sanitarios, y como los datos climáticos están generalmente disponibles en retícula o por estación meteorológica, es útil agregar o hacer la media de datos climáticos para que correspondan a las fronteras utilizadas por el sector de la salud.

Calcular Medias de Datos Climáticos sobre Distritos

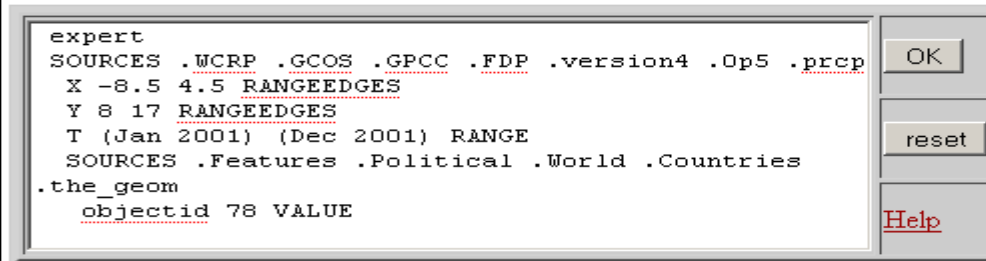
Calcular la media espacial de una base de datos en retícula según una geometría, requiere las tres etapas siguientes en Expert Mode:

1. Seleccione la base de datos en retícula
2. Seleccione la variable geometría en la base de datos de geometrías que quiere
3. Aplique la función **[X Y] weighted-average**.

Por su puesto, asegúrese que las geometrías coinciden con el dominio espacial de la base de datos seleccionado.

Calcular Medias de Datos Climaticos sobre Distritos

WCRP GCOS GPCC FDP version4 0p5 prcp[X Y | T]
Features Political World Countries the_geom 78[| objectid]



Features Political World Countries the_geom is



En este ejemplo se calcula la media espacial de la precipitación en 2001 sobre Burkina Faso. Note que el dominio espacial de la base de datos de precipitación en retícula fue restringido a una área que contiene Burkina Faso para acelerar el calculo. El “objectid” para Burkina Faso es “78”, y está seleccionado aquí. Con la variable de geometría al final de la pila de objetos y la selección de una de las geometrías, una mapa del mundo automáticamente muestra dicha geometría en rojo.

Calcular Medias de Datos Climáticos sobre Distritos

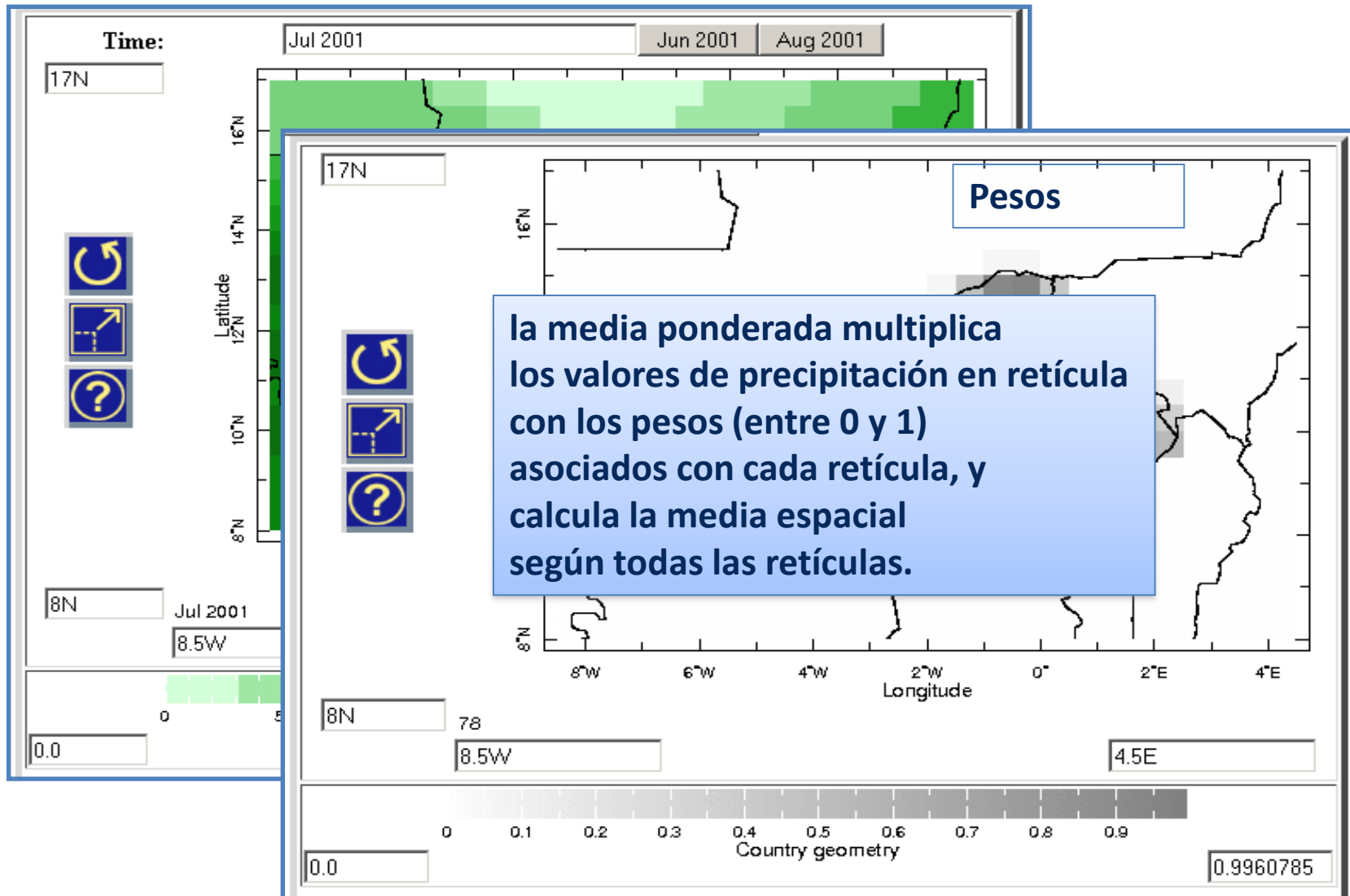
La etapa final es de aplicar la función **[X Y] weighted-average**.

Aplica una media ponderada de manera que las geometrías definidas por las fronteras están primero reticuladas para que correspondan a la base de datos en retícula de que se hace la media; en este caso, una base de datos de precipitación de resolución de 0.5° de longitud y latitud.

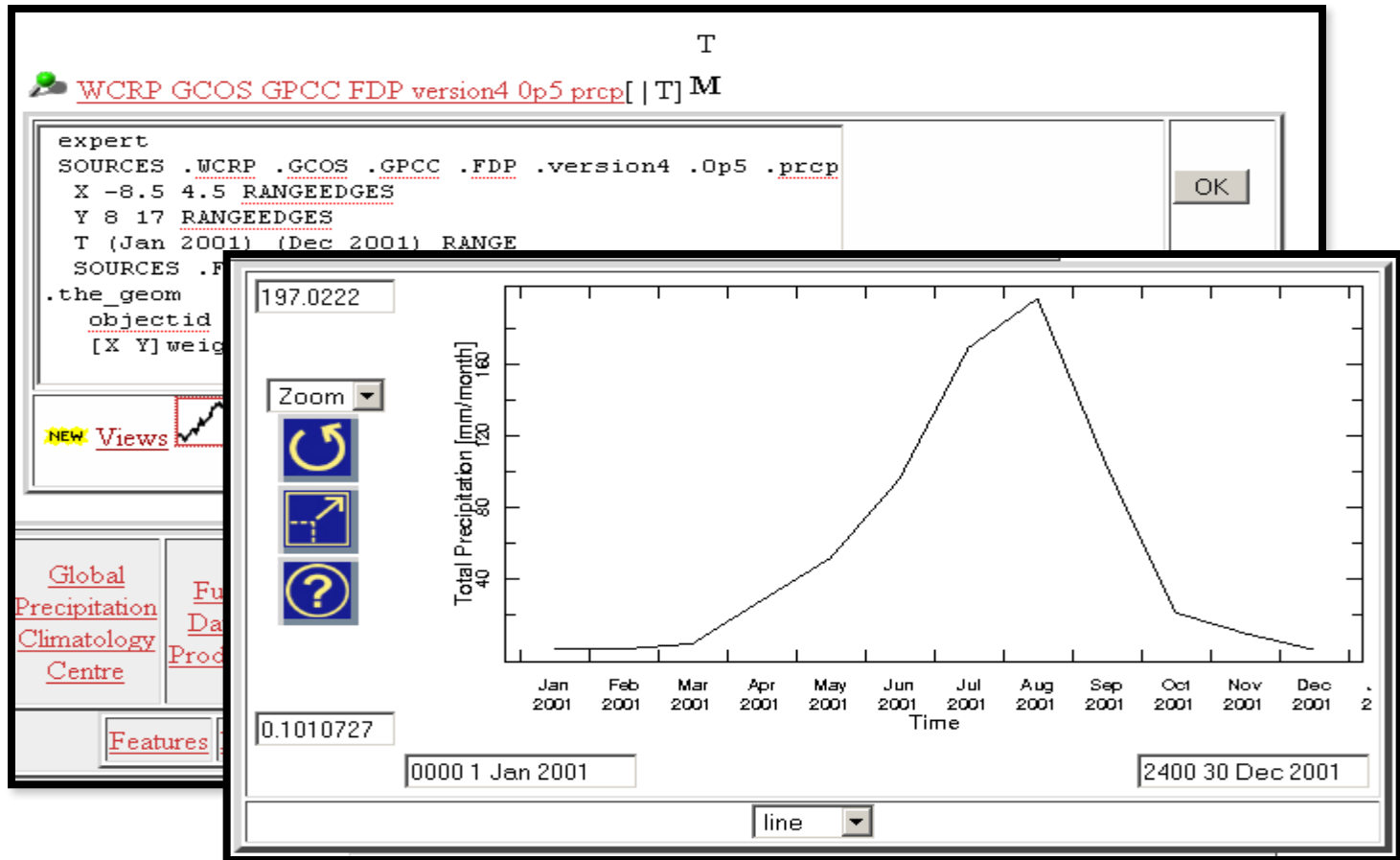
Pues, las retículas de las geometrías en retícula están asignadas pesos entre 0 y 1 que indican la fracción de cada retícula que pertenece a la geometría.

Cada peso está aplicado al valor de precipitación que corresponde a cada retícula antes de aplicar la media.

Calcular Medias de Datos Climáticos sobre Distritos



Calcular Medias de Datos Climáticos sobre Distritos



El resultado es una serie temporal sola para Burkina Faso.