



PRESENTACION IX

ANALISIS REGIONAL DE FRECUENCIAS CON PLATAFORMA OPEN SOURCE R Y R ESTUDIO

Curso Internacional

Análisis Regional de Frecuencia de Sequías basado en L-momentos

Jorge Nuñez C.

CAZALAC

Santiago-Chile

4-5 de Abril de 2011

¿Alternativas para la aplicación del procedimiento Análisis Regional de Frecuencias basado en L-momentos?

Una de las dificultades que ha tenido la difusión de la metodología Análisis Regional de Frecuencias basado en L-momentos es la carencia de herramientas informática para su fácil implementación.

La gran cantidad de cálculos requeridos para la aplicación de la metodología hace necesario el uso de programas de cálculo computacional. Por otro lado, este procedimiento no está disponible en los software de cálculo estadístico convencionales.

Las herramientas disponibles para la aplicación completa han ido evolucionando desde los antiguos códigos FORTRAN desarrollados por Hosking, hasta software en ambiente windows, como L-RAP.

L-RAP

- L-RAP es el nombre del paquete de cómputo Linear Regional Analysis of Precipitation, desarrollado por M.G.S. Engineering Consultants, del estado de Washington, EEUU.
- Este software, en ambiente windows, desarrolla las primeras 4 etapas del procedimiento visto en el curso. No incluye la elaboración de mapas, la cual es ejecutada por la aplicación L-MAP.
- Si bien es un software que es altamente escalable en su utilización y realiza todos los procedimientos necesarios para la aplicación del ARF-LM, la única limitante es que es un producto comercial y tiene un costo asociado.

R y R ESTUDIO

- Todas las rutinas requeridas por el Análisis Regional de Frecuencias basado en L-momentos, desarrolladas originalmente por Hosking en FORTRAN, han sido trasladadas por él y otros investigadores a la plataforma R.
- Esto tiene la ventaja de poner todas las rutinas a disposición gratuita del público en general, tener un respaldo científico a todos los procedimientos, automatizar procedimiento que demandan una gran cantidad de cálculos numéricos y estadístico y requerir una gran capacidad del usuario final para entender cada una de las operaciones involucradas en el cálculo.
- La principal desventaja es que R es un lenguaje complicado de aprender y con una alta curva de aprendizaje. Sin embargo, en este tutorial se pretende resolver las dificultades asociadas mediante:
 - El uso de R STUDIO, una plataforma un poco más amigable para la ejecución de R
 - La presentación del listado de comandos mínimo y necesario para aplicar toda la metodología

1. Ejecute RStudio y active, en la ventana derecha, en el menú packages, los paquetes que serán necesarios: lmom, lmomRFA y nsRFA

The screenshot shows the RStudio interface. The console window on the left displays the following text:

```
Platform: i386-pc-mingw32/i386 (32-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace restored from ~/.RData]

Package lmomco (1.2.2) loaded.
Warning message:
package 'lmomco' was built under R version 2.12.1
Loading required package: lmom

Attaching package: 'lmom'

The following object(s) are masked from 'package:lmomco':

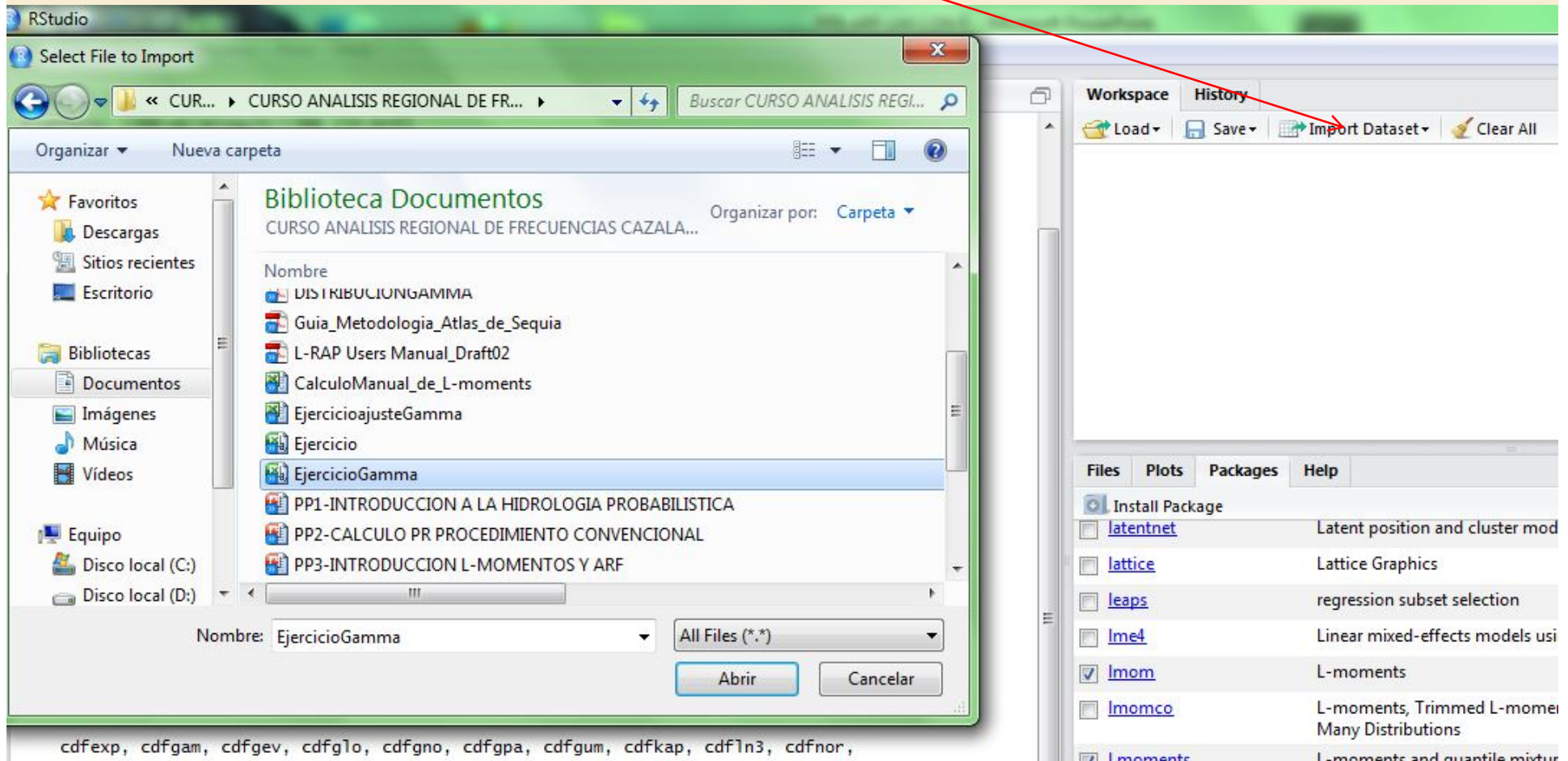
cdfexp, cdfgam, cdfgev, cdfglo, cdfgno, cdfgpa, cdfgum, cdfkap, cdfln3, cdfnor,
cdfpe3, cdfwak, cdfwei, quaexp, quagam, quagev, quaglo, quagno, quagpa, quagum,
quakap, qualn3, quanor, quape3, quawak, quawei

Warning messages:
1: package 'lmomRFA' was built under R version 2.12.1
2: package 'lmom' was built under R version 2.12.1
Warning message:
package 'lmom' was built under R version 2.12.1
Warning message:
package 'Lmoments' was built under R version 2.12.1
>
```

The Packages pane on the right shows a list of installed and available packages. A red box highlights the following packages:

Package Name	Description
<input type="checkbox"/> latentnet	Latent position and cluster models for statistical networks
<input type="checkbox"/> lattice	Lattice Graphics
<input type="checkbox"/> leaps	regression subset selection
<input type="checkbox"/> lmed	Linear mixed-effects models using S4 classes
<input checked="" type="checkbox"/> lmom	L-moments
<input type="checkbox"/> lmomco	L-moments, Trimmed L-moments, L-comoments, Censored L-moments, and Many Distributions
<input checked="" type="checkbox"/> Lmoments	L-moments and quantile mixtures
<input type="checkbox"/> lmomRFA	Regional frequency analysis using L-moments
<input type="checkbox"/> lntest	Testing Linear Regression Models
<input type="checkbox"/> locfit	Local Regression, Likelihood and Density Estimation.
<input type="checkbox"/> logspline	Log spline density estimation routines
<input type="checkbox"/> mapproj	Map Projections
<input type="checkbox"/> maps	Draw Geographical Maps
<input type="checkbox"/> maptools	Tools for reading and handling spatial objects

2. Importe un archivo csv (generado por Excel) con los datos a analizar (ImportDataSet->Text. Este archivo debe tener un formato específico.



3. Indique si las variables tienen un nombre en la cabeza de la columna y cuál es el identificador de separación de columnas. Una buena lectura de datos aparece abajo, en Data Frame. Luego presione Import.

Import Dataset

Name: EjercicioGamma

Input File: OBS v1;OBSv2;OBSv3;OBSv4;OBSv5;OBSv6
17.8;9.5;5.0;0.8;6.7;27.1
23.5;37.1;7.4;1.5;9.3;35.0
37.1;38.5;15.1;3.9;15.4;40.5
46.8;55.4;15.8;16.2;24.1;41.6
48.1;56.0;17.4;19.2;28.3;48.2
50.3;60.5;24.0;27.2;29.7;62.7
54.6;68.1;31.9;29.9;40.9;68.3
55.7;69.4;34.8;38.4;41.1;68.9
62.5;79.9;42.0;52.8;41.9;75.2
67.5;83.3;43.3;54.5;45.9;99.6
68.8;84.5;48.9;56.9;47.0;135.9
70.4;85.2;52.2;61.3;48.9;141.8
72.4;90.5;59.8;71.9;50.1;152.4
91.7;93.7;70.3;72.2;61.6;160.8
91.9;94.0;84.2;72.8;66.8;163.6
97.5;95.7;87.2;83.2;69.6;165.1
123.4;99.3;88.4;85.8;73.7;165.5
129.5;109.8;86.6;80.7;75.6;165.8

Heading: Yes No

Separator: Semicolon

Quote: Double quote (")

Data Frame

OBS.V1	OBSV2	OBSV3	OBSV4	OBSV5	OBSV6
17.8	9.5	5.0	0.8	6.7	27.1
23.5	37.1	7.4	1.5	9.3	35.0
37.1	38.5	15.1	3.9	15.4	40.5
46.8	55.4	15.8	16.2	24.1	41.6
48.1	56.0	17.4	19.2	28.3	48.2
50.3	60.5	24.0	27.2	29.7	62.7
54.6	68.1	31.9	29.9	40.9	68.3
55.7	69.4	34.8	38.4	41.1	68.9
62.5	79.9	42.0	52.8	41.9	75.2
67.5	83.3	43.3	54.5	45.9	99.6
68.8	84.5	48.9	56.9	47.0	135.9
70.4	85.2	52.2	61.3	48.9	141.8
72.4	90.5	59.8	71.9	50.1	152.4
91.7	93.7	70.3	72.2	61.6	160.8
91.9	94.0	84.2	72.8	66.8	163.6
97.5	95.7	87.2	83.2	69.6	165.1
123.4	99.3	88.4	85.8	73.7	165.5
129.5	109.8	86.6	80.7	75.6	165.8

Import Cancel

4. A la izquierda aparecerá la matriz de datos a utilizar. A la derecha, se indica el nombre de este set de datos, número de observaciones y cantidad de variables

The screenshot shows the RStudio interface with a data matrix loaded into the workspace. The matrix has 15 rows and 6 columns, labeled OBS.v1 through OBS.v6. The right-hand pane shows the 'Data' section with the dataset name 'EjercicioGamma' and its dimensions '30 obs. of 6 variables'. The bottom pane shows the console with a list of installed packages.

	OBS.v1	OBS.v2	OBS.v3	OBS.v4	OBS.v5	OBS.v6
1	17.8	9.5	5.0	0.8	6.7	27.1
2	23.5	37.1	7.4	1.5	9.3	35.0
3	37.1	38.5	15.1	3.9	15.4	40.5
4	46.8	55.4	15.8	16.2	24.1	41.6
5	48.1	56.0	17.4	19.2	28.3	48.2
6	50.3	60.5	24.0	27.2	29.7	62.7
7	54.6	68.1	31.9	29.9	40.9	68.3
8	55.7	69.4	34.8	38.4	41.1	68.9
9	62.5	79.9	42.0	52.8	41.9	75.2
10	67.5	83.3	43.3	54.5	45.9	99.6
11	68.8	84.5	48.9	56.9	47.0	135.9
12	70.4	85.2	52.2	61.3	48.9	141.8
13	72.4	90.5	59.8	71.9	50.1	152.4
14	91.7	93.7	70.3	72.2	61.6	160.8
15	91.9	94.0	84.2	72.8	66.8	163.6

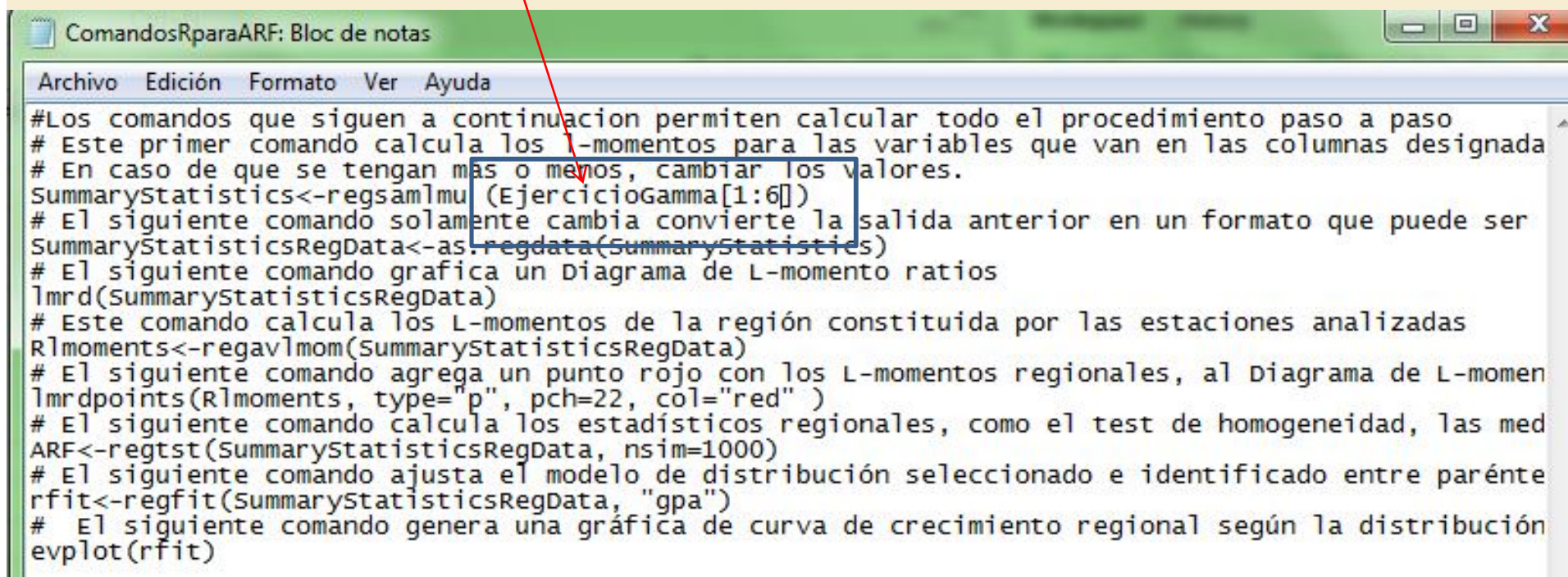
Workspace History
Load Save Import Dataset Clear All
Data
EjercicioGamma 30 obs. of 6 variables

Files Plots Packages Help
Install Package
 latentnet Latent position and cluster mc
 lattice Lattice Graphics
 leaps regression subset selection
 lme4 Linear mixed-effects models u
 lmom L-moments
 lmomco L-moments, Trimmed L-mom
Many Distributions

Console ~/
cdfexp, cdfgam, cdfgev, cdfglo, cdfgno, cdfgpa, cdfgum, cdfkap, cdfln3, cdfnor,
cdfpe3, cdfwak, cdfwei, quaexp, quagam, quagev, quaglo, quagno, quagpa, quagum,
quakap, qualn3, quanor, quape3, quawak, quawei

5. Ahora puede ingresar los comandos que ejecutan de manera completa el procedimiento ARF-LM. Estos comandos los puede copiar pegar, fila por fila, o todas las filas de una sola vez, en la consola de comandos, al lado izquierdo, después del símbolo prompt ">". Sólo debe cambiar dos cosas para analizar sus propios datos.

A) Donde dice (EjercicioGamma[1:8]) debe poner el nombre de su archivo cargado en el sistema (csv) y debe indicar el número correcto de variables que tiene cargadas. Si su archivo se llama Ecuador y tiene 30 variables, entonces cambiar por (Ecuador[1:30])



```
ComandosRparaARF: Bloc de notas
Archivo Edición Formato Ver Ayuda
#Los comandos que siguen a continuación permiten calcular todo el procedimiento paso a paso
# Este primer comando calcula los l-momentos para las variables que van en las columnas designada
# En caso de que se tengan más o menos, cambiar los valores.
SummaryStatistics<-regsamlmu (EjercicioGamma[1:6])
# El siguiente comando solamente cambia convierte la salida anterior en un formato que puede ser
SummaryStatisticsRegData<-as.regdata(SummaryStatistics)
# El siguiente comando grafica un Diagrama de L-momento ratios
lmerd(SummaryStatisticsRegData)
# Este comando calcula los L-momentos de la región constituida por las estaciones analizadas
Rlmoments<-regavlmmom(SummaryStatisticsRegData)
# El siguiente comando agrega un punto rojo con los L-momentos regionales, al Diagrama de L-momen
lmerdpoints(Rlmoments, type="p", pch=22, col="red" )
# El siguiente comando calcula los estadísticos regionales, como el test de homogeneidad, las med
ARF<-regtst(SummaryStatisticsRegData, nsim=1000)
# El siguiente comando ajusta el modelo de distribución seleccionado e identificado entre parénte
rfit<-regfit(SummaryStatisticsRegData, "gpa")
# El siguiente comando genera una gráfica de curva de crecimiento regional según la distribución
evplot(rfit)
```

B) Donde (SummaryStatisticsRegData, "gpa"), las siglas entre comillas deben ser reemplazadas por las que corresponden a la distribución seleccionadas. En este caso, "gpa" corresponde a la Pareto generalizada, pero otras tienen otro código de letras.

C) Finalmente, al final donde dice `cdfgpa(0.22, para=rfit$para)`, la gpa debe cambiarse por el código de letras correspondiente a la distribución seleccionada y el valor 0.22 por otro de interés. El 0.22 corresponde al cuantil para el cual se quiere calcular la probabilidad de no excedencia. En todos los casos representa una fracción de la media. Aquí, sería equivalente a un 22% de la media, o sea, un déficit de 78% respecto a la media.

```
lmrpoints(RImoments, type="p", pch=22, col="red")
# El siguiente comando calcula los estadísticos regionales, como el test de homogeneidad, las med
ARF<-regtst(SummaryStatisticsRegData, nsim=1000)
# El siguiente comando ajusta el modelo de distribución seleccionado e identificado entre paréntes
rfit<-regfit(SummaryStatisticsRegData, "gpa")
# El siguiente comando genera una gráfica de curva de crecimiento regional según la distribución
evplot(rfit)

# El siguiente comando, calcula los cuantiles regionales para distintas probabilidades acumuladas
RegionalQuantiles<-regquant(seq(0.05, 0.95, by=0.05), rfit)
# El siguiente comando une todas las curvas de cuantiles en una sola matriz de salida
TableQuantiles<-cbind(seq(0.05, 0.95, by=0.05), RegionalQuantiles)
# Los siguientes comandos generan una gráfica de cuantiles
rgc <- reggfunc(rfit)
rgc(seq(0.05, 0.95, by=0.05))
curve(rgc, 0.01, 0.99, xlab="Non-exceedence Probability, F", ylab="Growth curve")

# Los siguientes comandos calculan la probabilidad acumulada para determinados cuantiles, específic
# Si se quiere llevar la probabilidad a periodo de retorno en la cola inferior, basta dividir 1 p
# Por ejemplo, 0.22 equivale al cuantil 0.22 con respecto a la media. Es decir, un 22% de la medi
cdfgpa(0.22, para = rfit$para)
cdfgpa(0.38, para = rfit$para)
```

Puede ver los gráficos de salida en la ventana izquierda, abajo.

A la derecha, arriba, en **Values** se muestran todas las variables de resultado. Si quiere ver los valores de cada una, escriba su nombre en la consola y presione ENTER.

The screenshot displays a software interface with three main sections:

- Table:** A table with 15 rows and 3 columns. The columns are labeled 'row.names', 'V1', and 'RegionalQuantiles'. The data shows a linear relationship between 'row.names' and 'RegionalQuantiles'.
- Console:** A terminal window showing the command `> SummaryStatistics` and its output. The output is a table with columns: name, n, l_1, t, t_3, t_4, t_5. It lists 6 observations (OBS.v1 to OBS.v6) with their respective parameters.
- Plot:** A graph with 'L-skewness' on the x-axis (ranging from 0.0 to 0.6) and 'L-kurtosis' on the y-axis (ranging from 0.0 to 0.4). The plot shows several curves representing different distributions: GLO (blue), GEV (green), GPA (red), GNO (black), and PE3 (cyan). Data points are marked with '+' and 'U' symbols.

row.names	V1	RegionalQuantiles
1	0.05	0.1601795
2	0.1	0.2243529
3	0.15	0.2908766
4	0.2	0.3599830
5	0.25	0.4319439
6	0.3	0.5070798
7	0.35	0.5857739
8	0.4	0.6684891
9	0.45	0.7557939
10	0.5	0.8483980
11	0.55	0.9472061
12	0.6	1.0534001
13	0.65	1.1685722
14	0.7	1.2949504
15	0.75	1.4358132

```
> SummaryStatistics
  name n    l_1      t      t_3      t_4      t_5
1 OBS.v1 30 115.09333 0.3548764 0.22261331 0.11945259 0.07311876
2 OBS.v2 30 117.47333 0.3232761 0.23214059 0.16615117 -0.01379691
3 OBS.v3 30 84.78333 0.3982796 0.12955259 0.04152221 0.02673311
4 OBS.v4 30 86.10667 0.4089581 0.17398092 0.16829536 0.10490611
5 OBS.v5 30 110.28667 0.4894767 0.35704005 0.15366678 0.06214752
6 OBS.v6 30 161.32000 0.3248017 0.04249387 0.04859176 0.10192165
```

RStudio

File Edit View Workspace Plots Help

EjercicioGamma x TableQuantiles x

	row.names	V1	RegionalQuantiles
1	0.05	0.05	0.1601795
2	0.1	0.10	0.2243529
3	0.15	0.15	0.2908766
4	0.2	0.20	0.3599830
5	0.25	0.25	0.4319439

19 ob

Console ~/

```

> SummaryStatistics
  name n   l_1      t      t_3      t_4      t_5
1 OBS.v1 30 115.09333 0.3548764 0.22261331 0.11945259 0.07311876
2 OBSv2 30 117.47333 0.3232761 0.23214059 0.16615117 -0.01379691
3 OBSv3 30 84.78333 0.3982796 0.12955259 0.04152221 0.02673311
4 OBSv4 30 86.10667 0.4089581 0.17398092 0.16829536 0.10490611
5 OBSv5 30 110.28667 0.4894767 0.35704005 0.15366678 0.06214752
6 OBSv6 30 161.32000 0.3248017 0.04249387 0.04859176 0.10192165
> ARF
Discordancy measures (critical value 1.65)
0.41 1.06 0.82 1.54 1.33 0.83

Heterogeneity measures (based on 1000 simulations)
1.55 1.19 0.63

Goodness-of-fit measures (based on 1000 simulations)
  glo gev gno pe3 gpa
  2.86 1.55 1.28 0.69 -1.45
> Rlmoments
  l_1 l_2 t_3 t_4 t_5
1.0000000 0.3832781 0.1929702 0.1162800 0.0591717
> |

```

Así, primero se muestran los L-momentos y L-momento-ratios de cada una de las estaciones (variables)

Luego, los valores de discordancia para cada estación. En este caso, no hay estaciones discordantes.

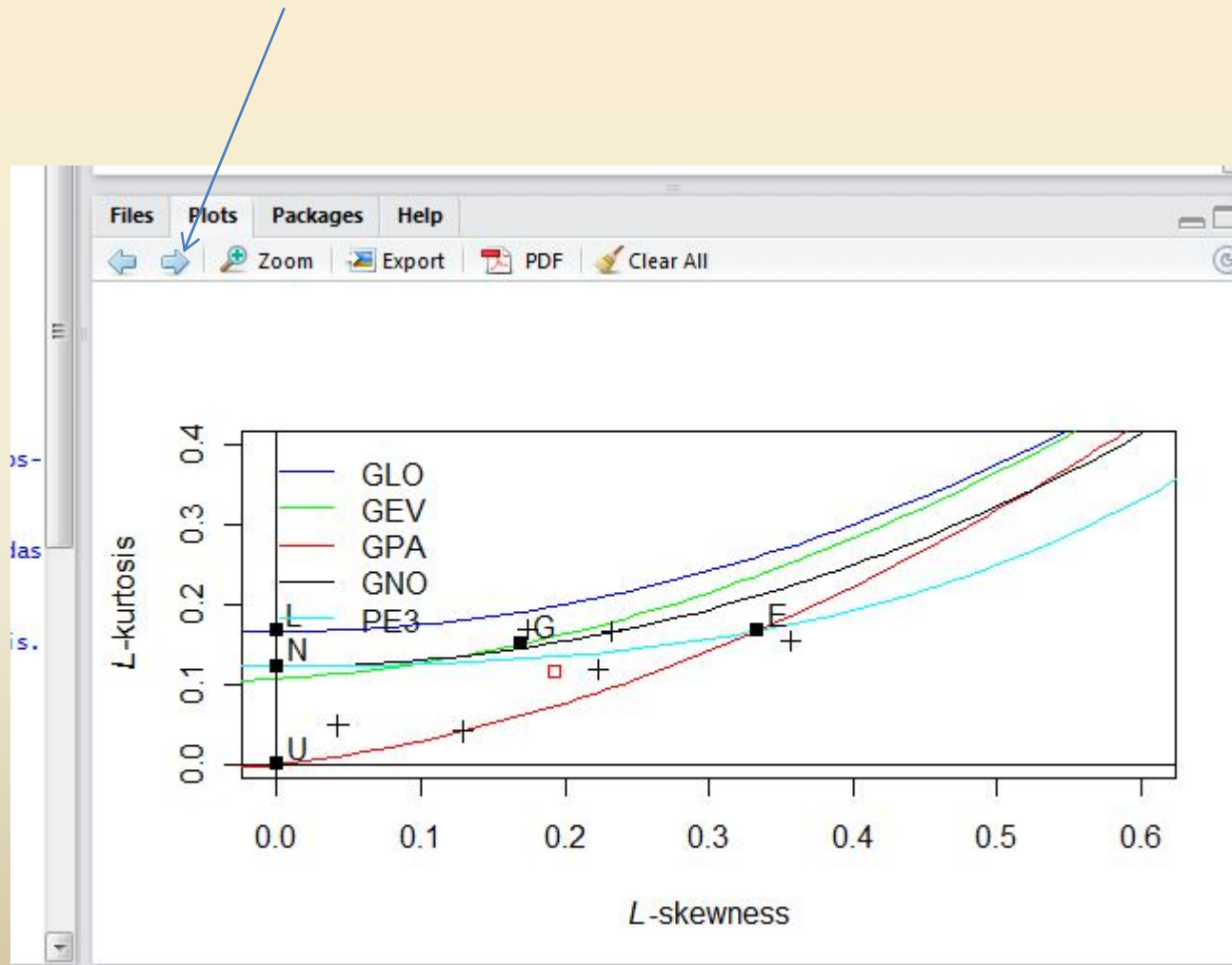
Luego, la medida de heterogeneidad H1 (primer valor) es menor a 2, lo que indica que la región es homogénea.

Luego, se ajustaron bien las distribuciones GEV, Gno, PE3 y GPAR. Puede seleccionar PE3 y corre nuevamente el programa.

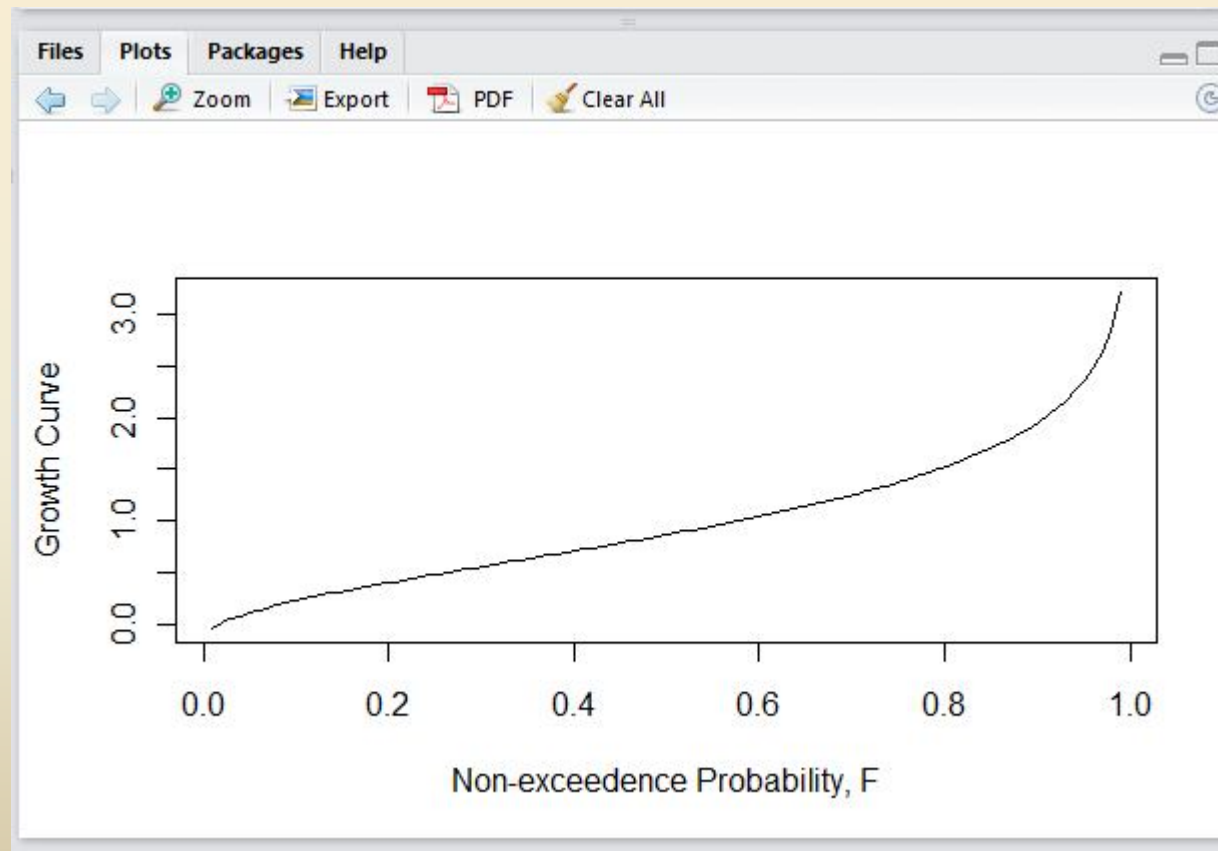
Luego, se muestran los L-momentos regionales.

En la ventana derecha, abajo, puede visualizar los distintos gráficos. En este diagrama de L-momento-ratios se muestran con "+" los L-momentos de las estaciones, con un cuadrado rojo, se muestran los L-momentos regionales y las líneas representan las distintas distribuciones.

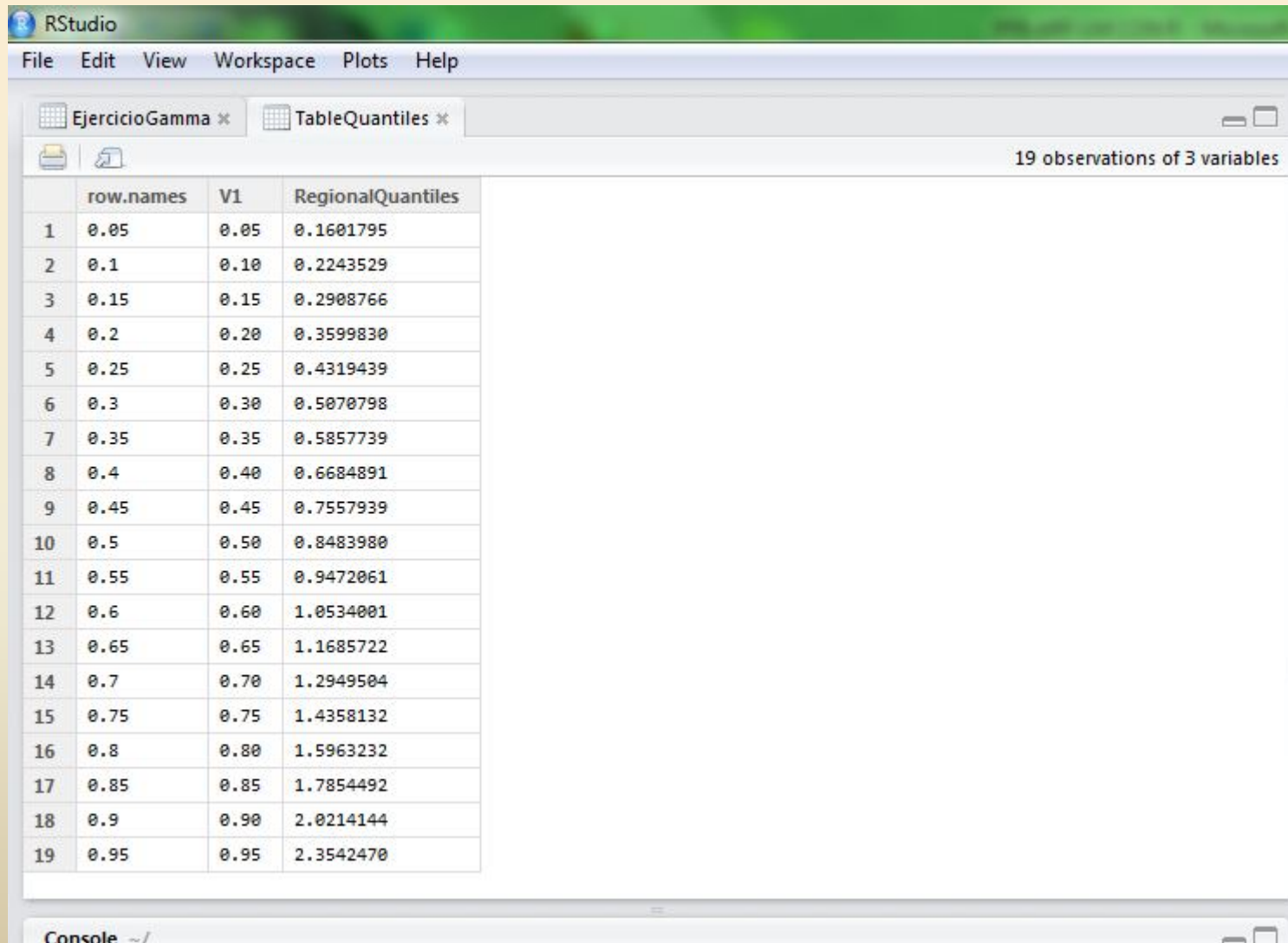
Presione la flecha para visualizar otro gráfico.



Este otro gráfico presenta la curva de crecimiento regional, adimensional. Esta curva puede usarse para calcular la curva específica de cada estación multiplicándola por la media de la estación.



Finalmente, esta ventana muestra la curva regional de manera numérica.



The screenshot displays the RStudio interface with a data table open. The table has 19 rows and 3 columns. The columns are labeled 'row.names', 'V1', and 'RegionalQuantiles'. The data shows a linear relationship between 'row.names' and 'RegionalQuantiles', with 'V1' values matching the 'row.names' values. The 'RegionalQuantiles' values increase as 'row.names' increases, starting from 0.1601795 for row 1 and ending at 2.3542470 for row 19.

	row.names	V1	RegionalQuantiles
1	0.05	0.05	0.1601795
2	0.1	0.10	0.2243529
3	0.15	0.15	0.2908766
4	0.2	0.20	0.3599830
5	0.25	0.25	0.4319439
6	0.3	0.30	0.5070798
7	0.35	0.35	0.5857739
8	0.4	0.40	0.6684891
9	0.45	0.45	0.7557939
10	0.5	0.50	0.8483980
11	0.55	0.55	0.9472061
12	0.6	0.60	1.0534001
13	0.65	0.65	1.1685722
14	0.7	0.70	1.2949504
15	0.75	0.75	1.4358132
16	0.8	0.80	1.5963232
17	0.85	0.85	1.7854492
18	0.9	0.90	2.0214144
19	0.95	0.95	2.3542470