



Universidad Nacional Agraria La Molina

1er. Coloquio de



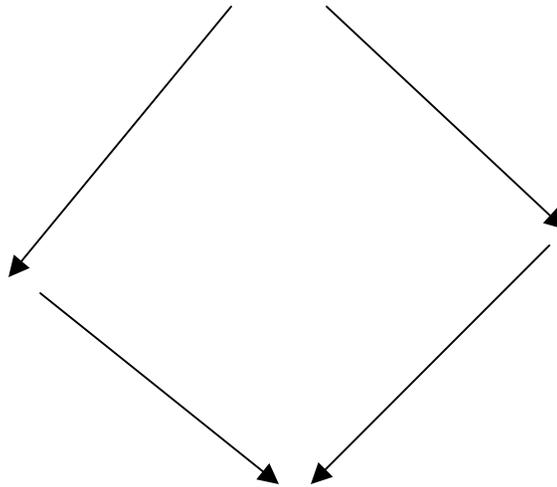
**El uso del R en la actualidad, y su desarrollo
frente a los paquetes comerciales**

Felipe de Mendiburu

Lenguaje S

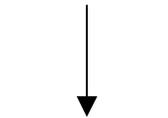


Robert Gentleman



Ross Ihaka

R & R



Projecto R

INICIO DE LA INVESTIGACION ESTADISTICA COMPUTACIONAL



AT&T Labs - Research



Investigacion estadistica en control de
calidad

Bell Labs

Investigacion para el lenguaje estadistico
S y el lenguaje de programacion C

Lucent Technologies
Bell Labs Innovations



Alcatel-Lucent



Iniciadores del S

"="", "<-"", "<<-"", "->"

La primavera de 1976, cinco personas de la Bell Labs: Rick Becker, John Chambers, Doug Dunn, Paul Tukey y Graham Wilkinson se juntaron para iniciar un proyecto de la creacion de un lenguaje para procesos estadisticos que llamaron S



John M. Chambers



**Rick
Becker**

John Chambers. Miembro del staff de laboratorios Bell desde 1966. En 1997 fue nombrado por Bell Labs como el primer estadístico, pionero de la computación en el campo de la estadística.

Rick Becker. había participado en un sistema estadístico en la Oficina Nacional de la Investigación Económica en Cambridge, Massachusetts antes de ir a Bell Labs.

Douglas Dunn. Particularmente interesado en la programación de series de tiempo.

Paul Tukey: con experiencia en el lenguaje APL (lenguaje de programación A)

Graham Wilkinson. 10 Años antes, junto con John Nelder estaban diseñando el sistema GenStat. Estuvo de visita y se quedó en el proyecto de Bell Labs.



Evolución del Lenguaje S

Version 1	1976-1980	GCOS de Honeywell, Fortran-lenguaje
Version 2	1980-1988	UNIX; Macros, el lenguaje y punto de contacto
	1981-1986	QPE (Ambiente de programación cuantitativa)
	1984-	Concesión de licencias, libros
Version 3	1988-1998	C-lenguaje; S funciones y objetos
	1991; rev. 1992	Modelos estadísticos; Clases informales y Métodos
	1993-	La licencia exclusiva a StatSci (posterior MathSoft) y para S-Plus.
Version 4	1998-	Clase-método formal modelo; Conexiones; Objetos grandes
	1999-	¿Puntos de contacto a Java, Corba?

APL

APL significa "A Programming Language". Se creó en el año 1960 por Ken Iverson y sus colegas en IBM. Matemáticamente inspirado, su objetivo principal era servir como un poderoso ejecutor de algoritmos con notación matemática.

APL es conocido por la utilización de símbolos no ASCII, incluidos algunos símbolos griegos, es un lenguaje matricial con alcance dinámico. En APL, todas las expresiones se evalúan de derecha a izquierda.

Ejemplo en APL

```
[0] Z←GAUSSRAND N;B;F;M;P;Q;R
[1] ⍝Returns ⍉ random numbers having a Gaussian normal distribution
[2] ⍝ (with mean 0 and variance 1) Uses the Box-Muller method.
[3] ⍝ See Numerical Recipes in C, pg. 289.
[4] ⍝
[5] Z←⊖0
[6] M←¯1+2*31 ⍝ largest integer
[7] L1:Q←N-ρZ ⍝ how many more we need
[8] →(Q≤0)/L2 ⍝ quit if none
[9] Q←⌈1.3×Q÷2 ⍝ approx num points needed
[10] P←¯1+(2÷M-1)×¯1+?(Q,2)ρM ⍝ random points in -1 to 1 square
[11] R←+/P×P ⍝ distance from origin squared
[12] B←(R≠0)∧R<1
[13] R←B/R ⋄ P←B÷P ⍝ points within unit circle
[14] F←(¯2×(⊙R)÷R)*.5
[15] Z←Z, ,P×F, [1.5]F
[16] →L1
[17] L2:Z←N↑Z
[18] ⍝ ArchDate: 12/16/1997 16:20:23.170
```

Programming in C

C es un propósito general, de bloque estructurado, de procedimiento, es un lenguaje de programación desarrollado en 1972 por Dennis Ritchie en los Laboratorios Bell Telephone para su uso con el sistema operativo Unix. Desde entonces se ha extendido a muchas otras plataformas. Aunque C se concibió como un sistema de aplicación de idiomas, también es utilizado ampliamente para las aplicaciones. C también ha influido en gran manera en muchos otros lenguajes populares especialmente C++, que fue originalmente diseñado como una extensión de la C.

Ejemplo de C

```
#include <stdio.h>
main(void)
{
    FILE *inp; /* apuntador archivo entrada */
    int sum=0, /* suma calculada hasta ese momento */
    score, /* resultado actual */
    input status; /* status regresado por fscanf() */
    inp = fopen("scores.dat","r");
    printf("Scores\n");
    for (input status = fscanf(inp, "%d", &score);
        input status != EOF;
        input status = fscanf(inp, "%d", &score)) {
        printf("%5dnn", score);
        sum += score;
    }
    printf("nn Suma de los resultados de los exámenes: %dnn", sum);
    fclose(inp);
    return (0);
}
```



Fortran Language

Un lenguaje para FORMula TRANslation

Esta maravillosa compilador FORTRAN fue diseñado y escrito desde Cero, en 1954-57 por un equipo dirigido por IBM John W. Backus

El primer compilador de FORTRAN fue un hito en la historia de la computación, en ese momento había muy pequeños ordenadores memorias (en el orden de 15 KB), eran lentos y tenían muy primitivos sistemas operativos

PROGRAM POLAR_CARTES

```
REAL X, Y, RHO, THETA  
READ (*,*) RHO, THETA  
CALL POLARES (X, Y, RHO, THETA)  
PRINT *, X, Y  
END
```

```
SUBROUTINE POLARES (X, Y, RHO, THETA)  
REAL X, Y, RHO, THETA  
X = RHO * COS (THETA)  
Y = RHO * SIN (THETA)  
END
```

Lucent Technologies Bell Labs To Host Statistical Computing Workshop; Event to honor Bell Labs' John Chambers, inventor of a key statistical language; speakers hail from Stanford, UCLA, Harvard, University of Wisconsin and AT&T.

M2 PRESSWIRE-29 April 2005-Lucent Technologies: Lucent Technologies Bell Labs To Host Statistical Computing Workshop; Event to honor Bell Labs' John Chambers, inventor of a key statistical language; speakers hail from Stanford, UCLA, Harvard, University of Wisconsin and AT&T(C)1994-2005 M2 COMMUNICATIONS LTD RDATE:26042005 MURRAY HILL, N.J.

Lucent Technologies' (NYSE: LU) Bell Labs will host an event in recognition of S system inventor John Chambers, entitled "40 Years of Statistical Computing." S, a breakthrough language and system for statistical computing, ...

Usuarios de S



El Proyecto R, Software estadístico libre



RStat ® es un ambiente de computación estadística que combina los puntos fuertes del sistema de software estadístico R del Proyecto R acoplado con apoyo y servicios a nivel de empresa “Random Technologies”. Actualmente utiliza más de 850 librerías de R



S-Plus, Comercial Utiliza el lenguaje S muchos años antes que R



S - PLUS es un paquete comercial de estadísticas avanzadas vendido por Insightful Corporation. Se trata de una aplicación comercial del lenguaje de programación S.

Su programación esta orientado a objetos y tiene capacidades avanzadas de análisis de algoritmos. Lentamente se esta convirtiendo en un feroz competidor del sistema SAS.

2007: S-PLUS 8 released. Nueva version compatible con toda la funcionalidad de los paquetes de R



GNU Operating System

¿Que es el GNU?

El Proyecto GNU se lanzó en 1984 para desarrollar un sistema operativo completo similar a UNIX que es el software gratuito: sistema de GNU. Variantes del sistema operativo de GNU que usan es llamado Linux, ahora están ampliamente usado; sin embargo estos sistemas a menudo se denominan “Linux”, que ellos se denominan con mayor precisión sistemas GNU/linux.

GNU es una sigla repetitiva para el GNU No UNIX.

¿Qué es el Software Gratuito?

El software gratuito es un asunto de la libertad, no es el precio. Para comprender el concepto, se debe pensar en “libre” como “libertad de expresión”.



El software gratuito es un asunto de la libertad de los usuarios para operar, copiar, distribuir, estudiar, cambiar y mejorar el software. Más precisamente, se refiere a cuatro clases de libertad, para los usuarios del software:

1. La libertad para administrar el programa, para cualquier fin.
2. La libertad para estudiar cómo trabaja el programa y adaptar a sus necesidades. El acceso al código fuente es una condición previa.
3. La libertad para redistribuir las copias para que usted puede ayudar a terceros.
4. La libertad para mejorar el programa y liberar sus mejoras al público, para que los beneficios sean para la comunidad. El acceso al código fuente es una condición previa.

<http://www.gnu.org/>

Iniciadores de R



Robert Gentleman: Estadístico Matemático

Nació en 1959 en Whitehorse, Canadá

1984 B.Sc. (Hons.) Mathematics University of British Columbia

1986 M.Sc. Statistics University of Washington

1988 Ph.D. Statistics University of Washington

1996–2000 Senior Lecturer, Department of Statistics, University of Auckland (Tenured) New Zealand



Ross Ihaka: Estadístico

Department of Statistics, University of Auckland

Private Bag 92019 Auckland

New Zealand

Welcome to the Department of Statistics

We are the largest Statistics Department in Australia and New Zealand and the birth place of .



DEPARTMENT OF STATISTICS
FACULTY OF SCIENCE
THE UNIVERSITY OF AUCKLAND

<http://www.stat.auckland.ac.nz/>



r-downunder

R Fest

DSC 2007

LDA 2008



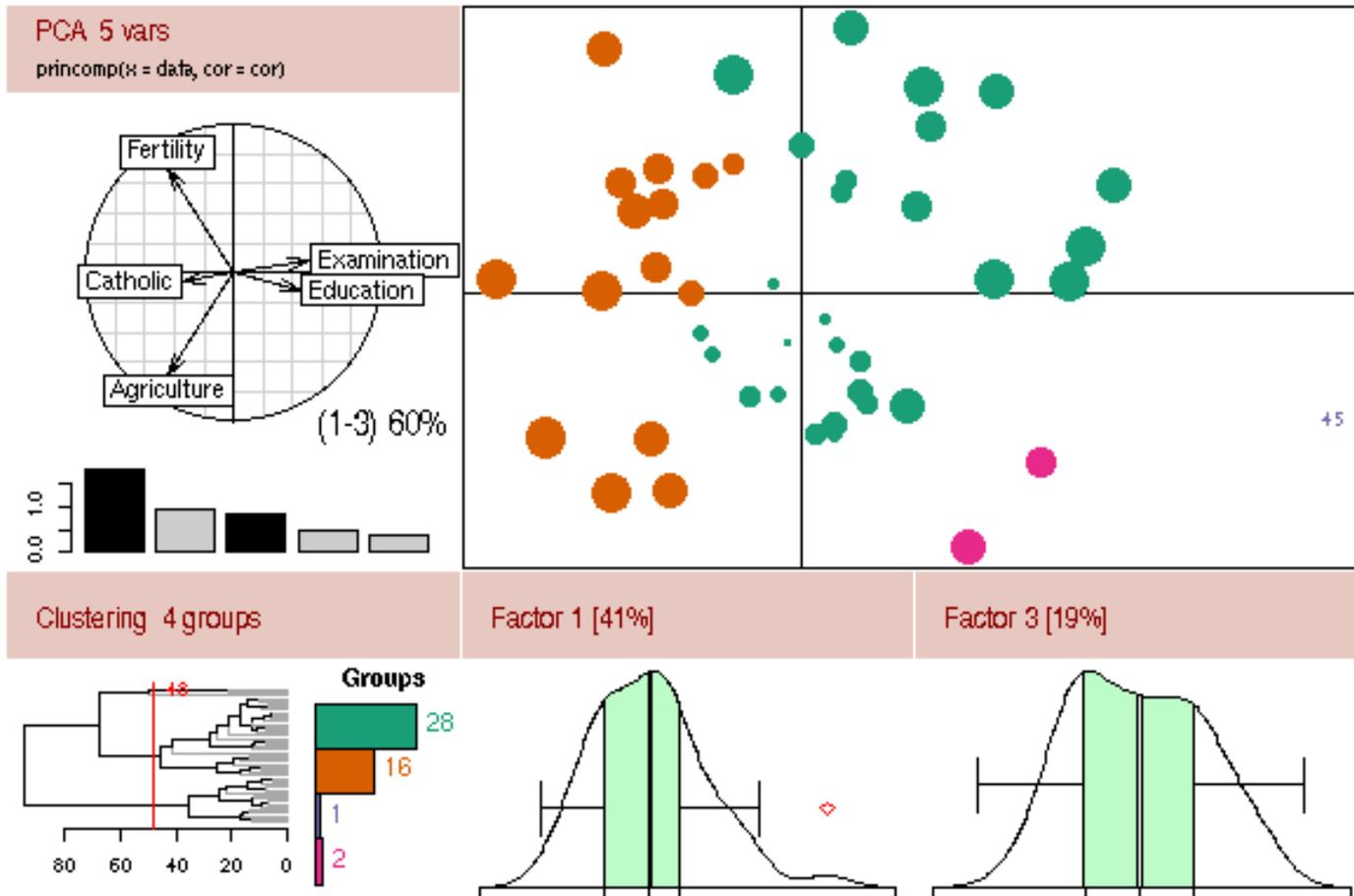
SRTL

IASS 55

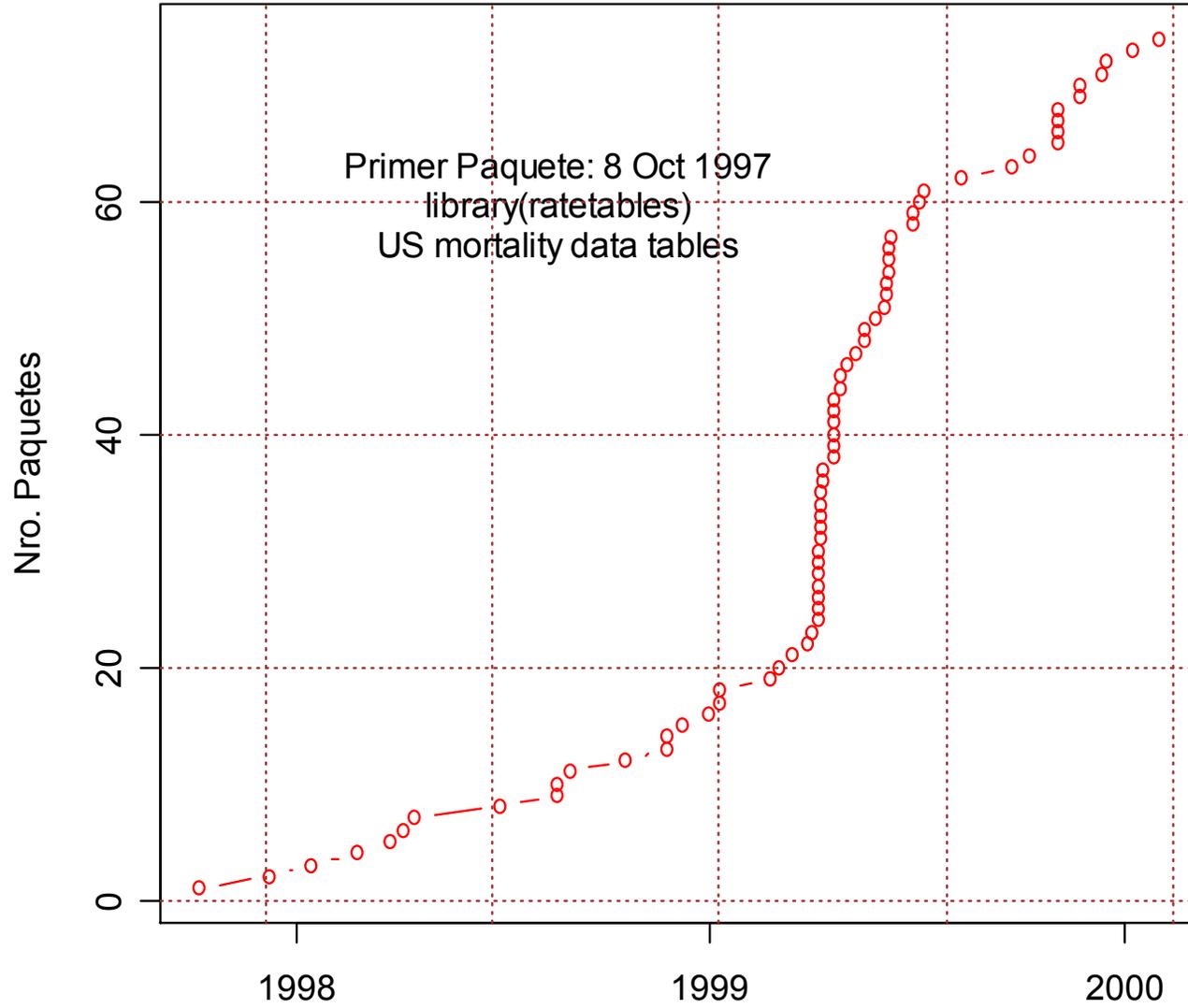


El Proyecto R para Estadística computacional

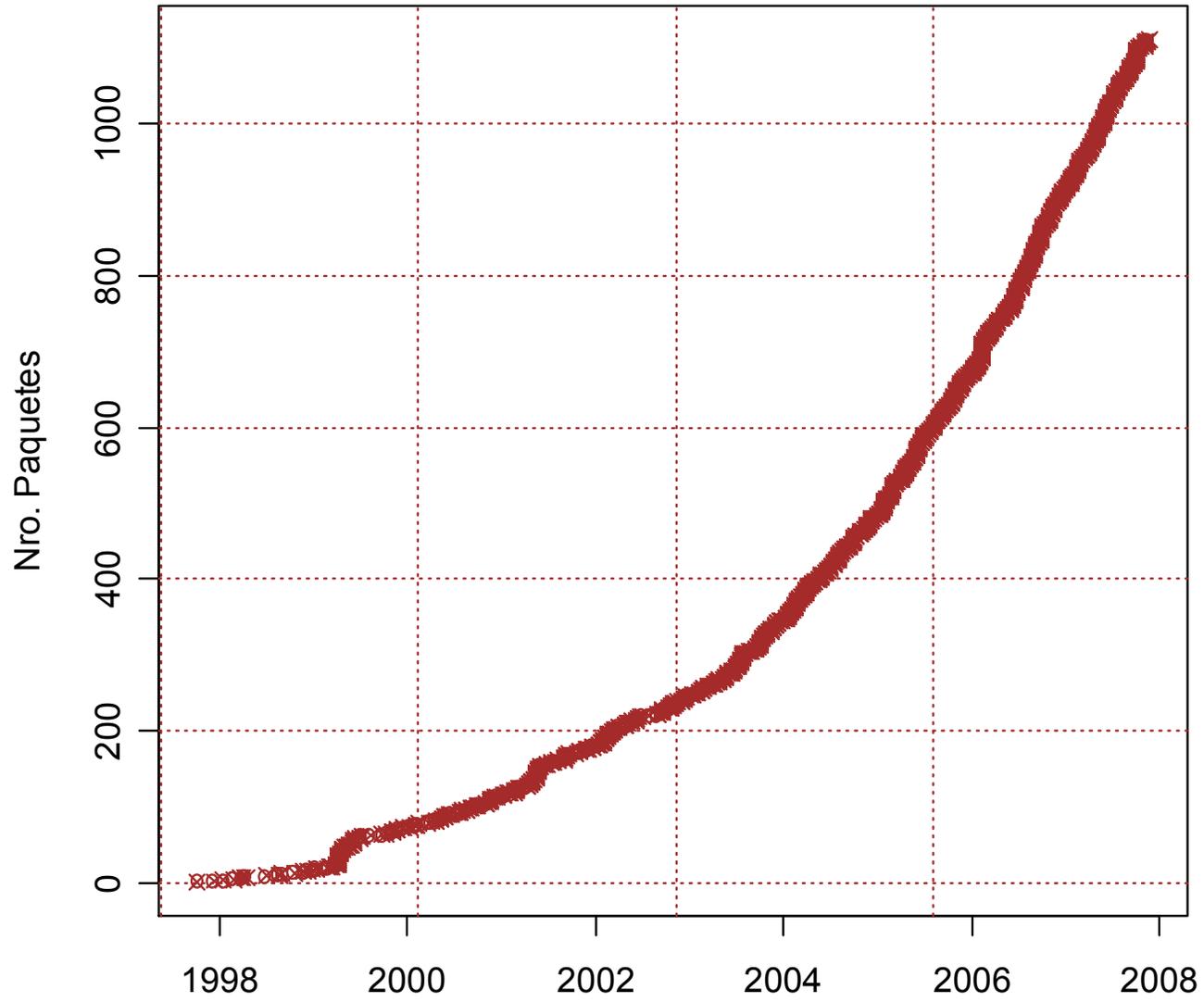
<http://www.r-project.org/>



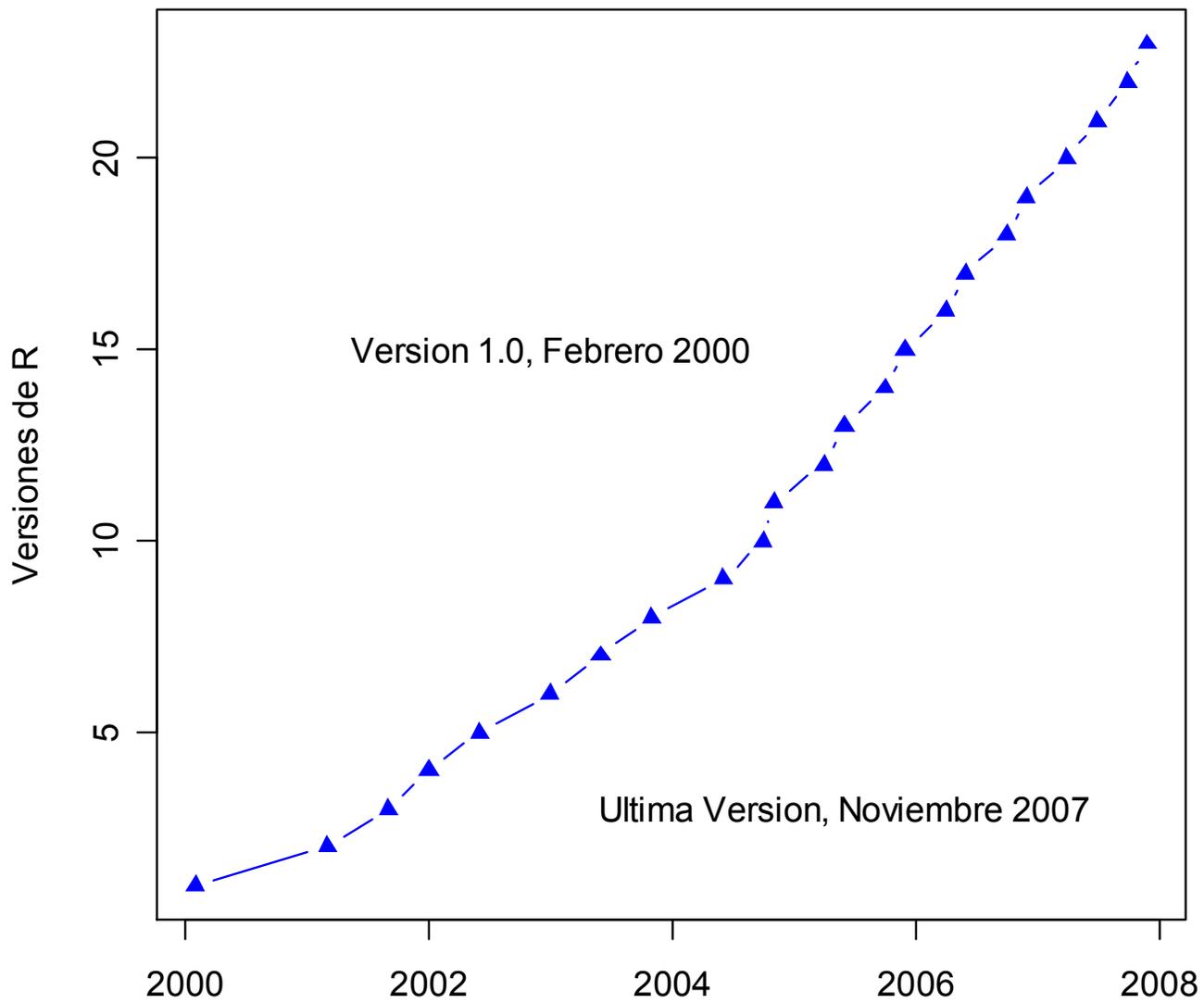
74 Paquetes antes de la version 1.0 de R



1111 Paquetes



Evolucion de R, Versión 23 en la actualidad



CRAN: The Comprehensive R Archive Network
Red amplia de archivos R

31 Países y 64 Universidades a nivel mundial

Argentina	Germany	Portugal
Australia	India	Slovenia
Austria	Ireland	South Africa
Belgium	Italy	Spain
Brazil	Japan	Sweden
Canada	Korea	Switzerland
Chile	Mexico	Taiwan
Croatia	Netherlands	Thailand
Czech Republic	Norway	UK
Denmark	Poland	USA
France		

12 Diciembre del 2007

CRAN Task Views

Bayesian	Bayesian Inference
Cluster	Cluster Analysis & Finite Mixture Models
Econometrics	Computational Econometrics
Environmetrics	Analysis of ecological and environmental data
Finance	Empirical Finance
Genetics	Statistical Genetics
Graphics	Graphic Displays & Dynamic Graphics & Graphic Devices & Visualization
gR	gRaphical models in R
MachineLearning	Machine Learning & Statistical Learning
Multivariate	Multivariate Statistics
Pharmacokinetics	Analysis of Pharmacokinetic Data
Psychometrics	Psychometric Models and Methods
SocialSciences	Statistics for the Social Sciences
Spatial	Analysis of Spatial Data

Procedimiento en R

```
setwd("g:/COLOQUIO-R")
teca <- read.table("teca.txt", header=T)
modelo2 <- lm(log(12-altura) ~ edad, teca)
anova(modelo2)
summary(modelo2)
coeficiente <- coef(modelo2)
tiempo <- seq(0, 30, 0.1)
f2 <- function(tiempo) {
  12-exp(coeficiente[1]+coeficiente[2]*tiempo)
}
y2 <- f2(tiempo)
attach(teca)
plot(tiempo, y2, type="l", ylim=c(0, 14), bty="l", xlab="", ylab="")
title(main="Crecimiento de Teca\n en 27 años", xlab="Edad",
      ylab="altura")
points(edad, altura, col="blue", pch=17)
grid(col="brown", lwd=1)
detach(teca)
```

```
> anova(modelo2)
```

```
Analysis of Variance Table
```

```
Response: log(12 - altura)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
edad	1	4.0348	4.0348	21.209	0.00999 **
Residuals	4	0.7610	0.1902		

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1  
' ' 1
```

```
> summary(modelo2)
```

```
Call:
```

```
lm(formula = log(12 - altura) ~ edad, data = teca)
```

```
Residuals:
```

1	2	3	4	5	6
0.3065	0.2200	0.3112	-0.4086	-0.5770	0.1479

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.15957	0.22975	9.400	0.000714 ***
edad	-0.09118	0.01980	-4.605	0.009991 **

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1  
' ' 1
```

```
Residual standard error: 0.4362 on 4 degrees of freedom
```

```
Multiple R-Squared: 0.8413,    Adjusted R-squared:  
0.8017
```

```
F-statistic: 21.21 on 1 and 4 DF,  p-value: 0.00999
```

Crecimiento de Teca en 27 años

$\log(12 - \text{altura}) \sim \text{edad}$

teca.txt

edad	altura
1	1.25
2	3
3	3
4	8
7	9.429
27	11.143

