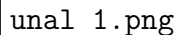

Universidad Nacional de Colombia

Sede La Paz

Daniel David Torres Martinez
Jorge Palacios Rico

Taller de programación en lenguaje estadístico



unal 1.png

Trabajo dirigido por
Prof. Jose Francisco Ruiz Munoz

Programacion en lenguaje estadistico, 2022-1

Resumen

R es un lenguaje y un entorno potentes para estadística y gráficas. es un bien público principal (un llamado "GNU") proyecto que es similar al lenguaje S comercial y al entorno que fue desarrollado en Bell Laboratories (fueron anteriormente ATT, ahora Lucent Technologies) por John Chambers y colegas. R puede ser considerado como una implementación diferente de S, y se usa mucho como lenguaje educativo y herramienta de investigación. Las principales ventajas de R son el hecho de que R es freeware y que hay mucha ayuda disponible en línea. Es bastante similar a otra programación. paquetes como MatLab (no freeware), pero más fácil de usar que los lenguajes de programación como C++ o Fortran. Puede usar R tal como está, pero para propósitos educativos, preferimos usar R en combinación con la interfaz RStudio (también freeware), que tiene un diseño organizado y varias opciones. Este documento contiene explicaciones, ejemplos y ejercicios, que también se pueden entender (con suerte) por personas sin ninguna programación experiencia. Repasando todo el texto y los ejercicios. toma alrededor de 1 o 2 horas. Ejemplos de frecuentemente los comandos usados y los mensajes de error se enumeran en las dos últimas páginas de este documento y puede ser utilizado como referencia durante la programación.

R is a powerful language and environment for statistical computing and graphics. It is a public domain (a so called "GNU") project which is similar to the commercial S language and environment which was developed at Bell Laboratories

(for-merly ATT, now Lucent Technologies) by John Chambers and colleagues. R can be considered as a different implementation of S, and is much used in as an educational language and research tool. The main advantages of R are the fact that R is freeware and that there is a lot of help available online. It is quite similar to other programming packages such as MatLab (not freeware), but more user-friendly than programming languages such as C++ or Fortran. You can use R as it is, but for educational purposes we prefer to use R in combination with the RStudio interface (also freeware), which has an organized layout and several extra options. This document contains explanations, examples and exercises, which can also be understood (hopefully) by people without any programming experience. Going through all text and exercises takes about 1 or 2 hours. Examples of frequently used commands and error messages are listed on the last two pages of this document and can be used as a reference while programming.

1. Como instalar R

Para instalar R en su computadora ir al sitio web de inicio de R y haga lo siguiente asumiendo que trabaja en un computadora con windows:

- haga clic en descargar CRAN en la barra izquierda
- elegir un sitio de descarga
- elegir Windows como sistema operativo de destino
- haga clic en la base
- elija Descargar R 3.0.3 para Windows

elegir respuestas predeterminadas para todas las preguntas También es posible ejecutar R y RStudio desde una memoria USB en lugar de instalarlos. Este

podría ser útil cuando no tiene administra derechos de autor en su computadora. Vea nuestra nota separada “Cómo usar versiones portátiles de R y RStudio” para obtener ayuda sobre este tema. Después de terminar esta configuración, debería ver una Ricono en su escritorio. Al hacer clic en esto se iniciaría

la interfaz estándar. Recomendamos, como- nunca, para usar la interfaz de RStudio Instalar RStudio, vaya a: <http://www.rstudio.org/> y haz lo siguiente asumiendo que trabajas en un windows:

- haga clic en Descargar RStudio
- haga clic en Descargar RStudio Desktop
- haga clic en Recomendado para su sistema
- descargue el archivo .exe y ejecútelo elija predeterminado respuestas a todas las preguntas

Descripcion y ejemplos de las estructuras de datos basicas de R (vectores, matrices y data frames) y su manipulacion

Como en muchos otros programas, R organiza números en escalares un solo número - 0-dimensional vectores (una fila de números, también llamados arreglos 1-dimensional) y matrices como una tabla bidimensional. El a que definiste antes era un escalar. Definir un vector con los números 3, 4 y 5, necesitas el función c, que es la abreviatura de concatenar (pegar juntar) $b=c(3,4,5)$ Matrices y otras estructuras bidimensionales Algunas personas prefieren usar $<$ – en lugar de $=$ (hacen la misma cosa). $<$ – consta de dos caracteres, $<y$ -, y representa una flecha apuntando al objeto que recibe el valor de la expresión.

2. Vectores

Si desea calcular la media de todos los elementos en el vector b del ejemplo anterior, podrías escribir $> (3 + 4 + 5)/3$ Pero cuando el vector es muy largo, esto es muy aburrido y el trabajo que consume mucho tiempo. Por eso las cosas a menudo automatizan en las llamadas funciones. Algunas funciones son estándar en R o en

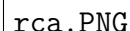
uno de los paquetes. También puede programar su propia función. Cuando usas una función para calcular una media, escribirás: `> mean(x=b)`

Dentro de los corchetes se especifican los argumentos. Los argumentos dan información adicional a la función. En este caso, el argumento `x` dice de qué conjunto de números (vector) la media debe calcularse (es decir, de `b`). A veces, el nombre del argumento `mean` no es necesario: `mean(b)` también funciona.

3. Matrices

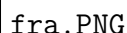
Las matrices no son más que bidimensionales.

Vectores Para definir una matriz, utilice la función `matriz`:



rca.PNG

El argumento `data` especifica qué números debe estar en la matriz. Utilice `ncol` para especificar el número de columnas o `nrow` para especificar el número de filas. Coloque los números del 31 al 60 en un vector llamado `P` y en una matriz de 6 filas y 5 columnas llamado `Q`. Sugerencia: use la función `seq`. Mirar las diferentes formas escalares, vectores y matrices se indican en la ventana del área de trabajo. Las operaciones matriciales son similares a las operaciones vectoriales.



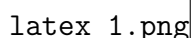
fra.PNG

Los elementos de una matriz se pueden abordar en el forma habitual: fila, columna línea 1.

- Línea 3: cuando desee seleccionar una fila completa, deja el lugar para el número de columna vacío (al revés para las columnas, por supuesto).

4. Dataframes

Dataframe Las series de tiempo a menudo se ordenan en dataframes. A los dataframes, es una matriz con nombres por encima de la columnas. Esto es bueno, porque puedes llamar y usar una de las columnas sin saber en cual puesto esta



latex 1.png

- En las líneas 1-2 se muestra un dataframe típico llamado `t` construido. Las columnas tienen los nombres `x`, `y` y `z`.
- Las líneas 8-11 muestran dos formas de cómo puede seleccionar la columna llamada `z` del dataframe llamado `t`.

5. Opciones y ejemplos de visualizacion de datos

5.1. plot

La funcion, plot es usada de manera general para crear gráficos. Esta función tiene un comportamiento especial, pues dependiendo del tipo de dato que le demos como argumento, generará diferentes tipos de gráfica. por ejemplo para hacer un plot de una base de datos llamado salon de clases y queremos graficar cuantos estudiantes hay hacemos lo siguiente: `plot(salon_de_clases)`

5.2. hist

Un histograma en R es el gráfico más habitual para representar datos continuos. Este tipo de gráfico es similar a un diagrama de barras, pero que representa las frecuencias en las que aparecen las mediciones agrupadas en ciertos intervalos y cuenta cuántas observaciones caen en cada intervalo. Además, la altura está determinada por el ratio entre la frecuencia y el ancho del intervalo. La funcion hist, crea una histrograma con los datos que nosotros queramos graficar, por ejemplo, si queremos graficar los años de un dataframe llamado poblacion, hacemos lo siguiente: `hist(poblacion$ años)` y la grafica nos saldra en el apartado de plots

5.3. barplot

Cuando una variable toma pocos valores, es común resumir la información con una tabla de frecuencias que se puede representar con un gráfico de barras en R. Este tipo de gráficos se suelen utilizar, por ejemplo, para representar precipitaciones y temperaturas (si añadimos una curva por encima), en lo que se conoce como climogramas. En este artículo vamos a explicar los conceptos básicos de la creación de diagramas de barras en R

6. Bibliografía

- Moon, K. W. (2016). Labelling a Bar Plot (I). In Learn ggplot2 Using Shiny App (pp. 129-139). Springer, Cham.
- Ramasamy, R. K., Ramasamy, S., Bindroo, B. B., Naik, V. G. (2014). STRUCTURE PLOT: a program for drawing elegant STRUCTURE bar plots in user friendly interface. SpringerPlus, 3(1), 1-3..
- Phillips, N. D. (2017). Yarrrr! The pirate's guide to R. APS Observer, 30(3).