KU CENTER FOR RESEARCH METHODS & DATA ANALYSIS

College of Liberal Arts & Sciences

Introducción a R

Mauricio Garnier Villarreal

2012

Instalación

- http://www.r-project.org/
- http://rstudio.org/
- http://notepad-plus-plus.org/

- 1983, GNU
- "Software Libre"
- R es un conjunto facilidades de software para la manipulación de datos, cálculos y gráficos
- Para julio 1 hay 3894 paquetes
- Capacidad de ver el código
- Comunidad de contribuyentes y depuradores

Operaciones básicas

- Programación en base a objetos
- Súper calculadora
  - 1+2
  - 1-2
  - 1*2
  - ½
  - sqrt(125)
- # comentarios

Objetos

- Data frame
- Vector
- Numeric
- Matrix
- Character
- Logical
- List
- Function

Objetos

- Cada objeto
  - Tiene un nombre
  - Tiene ciertas caracteristicas
  - Existe en el espacio de trabajo
- Nombres
  - Debe empezar con una letra
  - Puede incluir numeros (“a1”)
  - Sensible a mayusculas (“A” es diferente que “a”)
  - Puede incluir los caracteres “.” y “_”

Crear objetos

- Se crean objetos asignándolos con
  - `x <- 1`
  - `y <- 2`
- Podemos ver los objetos llamándolos
  - `x`
- Podemos ejecutar operaciones sobre los objetos
  - `x + y`
  - `z <- x + y`

Espacio de trabajo

- Como un escritorio
- Los objetos interactúan en el espacio de trabajo
- ls()
- Se puede guardar el espacio de trabajo o solo el código

Manipulando objetos

- `x <- 1`
- `y <- 2`
- `x`
- `x + y`
- `z <- x + y`

Vectores

- Vector numérico
  - x <- c(1, 2, 3, 4)
  - y <- c(5, 6, 7, 8)
  - x
- x + y
- x – 2.5

Funciones

- Las funciones se usan para llevar a cabo acciones
- Cada función requiere entradas especificas
- log(), tan(), cos(), sum(), prod(), min(), max(), range(), mean(), var(), sd(), length(), round()

Ejercicio 1

- Calcule los puntajes z para el siguiente vector
  - x <- c(4, 7, 3, 2, 9, 7, 6, 2, 4, 6, 9)
- Pasos:
  - Calcule la media de x
  - Reste le la media a x
  - Calcule la desviación estándar de x
  - Divida los puntajes centrados por la desviación estándar

Ayuda en R

- help.search("function") o help.search("palabra clave")
  - help.search("regression")
- ?function
  - ?sort

Objetos character

- Objetos de texto
  - x <- “pura vida”
  - x
- Un vector de characteres
  - x <- c(“lunes”, “martes”, “miércoles”, “jueves”)
  - x

Objetos lógicos

- x <- 4
- Hagamos preguntas sobre x
  - x == 4
  - x != 4
  - x < 4
  - x <= 5
- y <- c(95, 90, 85, 87, 62, 75)
  - y < 70
  - sum(y < 70)

Verificando los objetos

- Para verificar el tipo de objetos:
  - mode(x)
  - is(x)
- A veces se necesita forzar a un objeto a ser de otro tipo
  - x <- c(1, 2, 3, 4, 5, 6)
  - x[6]<- "NA"
  - mode(x)
  - is.numeric(x)
  - as.numeric(x)
- Ocasionalmente útiles: as.numeric(), is.numeric(), as.character(),is.character(),as.logical(), is.logical()

Notas

- NA es dato perdido, NaN es “not a number”
- Infinity es Inf
- pi
- R confía en ustedes de no nombrar objetos con nombres de objetos o funciones (como, pi, sum)

Paquetes

- Los desarrolladores suben paquetes llenos de objetos – funciones
- Los paquetes se instalan una vez, pero se deben cargar al espacio de trabajo.
- Instalar paquetes (coloca el paquete en la computadora)
  - install.packages("psych", dep=T)
- Cargar paquetes (trae el paquete al espacio de trabajo)
  - library(psych)

Working directory

- El working directory es donde R busca los datos y exporta, es recomendable tener un folder para cada proyecto y fijar este folder como working directory
- setwd("C:/Users/student/Desktop/combining chi suares/combining chi squares")

Importación y exportación de datos

- Primero: exporten la base de datos en formato .csv o .txt
- Desde excel o SPSS.
- read.csv() o read.table(), write.csv() o write.table()
- ?read.csv() ?read.table(), ?write.csv() o ?write.table()
- Se lee como un data frame: columnas son variables y filas son observaciones, base rectangular, combina diferentes tipos de variables

Ejemplo

- dat1 <- read.table(file="ex1.txt",header=T) ## el documento ex1.txt tiene que estar en el working directory
  - File="loquesea.txt", nombre del documento
  - header=T, especifica que la primer fila son los nombres de las columnas
  - sep= el separador (default es espacio, tab "\t", coma ",")
- head(dat1), dim(dat1)
- dat1$y1, dat1[,1], dat1[1,], dat1[1,1], dat1[,1:3], dat1[["y1"]], dat1["y1"]

Ejemplo

- dat1
- head(dat1)
- dat1$y1, dat1[,1], dat1[1,], dat1[1,1], dat1[,1:3], dat1[["y1"]], dat1["y1"]
- summary(dat1), str(dat1), colnames(dat1), attributes(dat1)

Estadísticas de grupo

- ?aggregate
  - aggregate(Temp ~ Month, data = airquality, mean)
  - aggregate(Temp ~ Month + FirstHalf, data = airquality2, mean)
- ?table

Pruebas de hipótesis

Independent t-test

- ?t.test
- Contra un valor
- Entre 2 variables
- Se asume varianzas iguales
- Pruebas pareadas
- Entre grupos (paired-sample t test)

Correlación

- ?cor
- cor(dat1)
- cov(dat1)
- cor(cbind(dat1$y1,dat1$y2,dat1$y3))
- cor(dat1$y1,dat1$y2)
- cor.test(dat1$y1,dat1$y2)

Regresión

- Regresión simple

$$y_i = b_0 + b_1 x_i + e_i$$

- ?lm
- lm(y1~y2,data=datsim)

$$y_1 = -0.06724 + 0.48913 * y2 + e_i$$

Regresión

- mod1<-lm(y1~y2,data=datsim)
- summary(mod1)
- plot(mod1)
- anova(mod1)
- vcov(mod1)
- confint(mod1)
- fitted(mod1)
- resid(mod1)

Regresión

- sort.dat<-datsim[order(datsim$y2),]
- modgraf<-lm(y1~y2,data=sort.dat)
- pc<-predict(modgraf,int="c")
- pp<-predict(modgraf,int="p")
- plot(y1~y2,data=sort.dat)
- matlines(sort.dat$y2,pc,lwd=1,lty=c(1,2,2),col="black")
- matlines(sort.dat$y2,pp,lwd=1,lty=c(1,3,3),col="red")

Regresión

- Regresión múltiple

$$y_i = b_0 + b_1 x_1 + b_2 x_2 + e_i$$

- lm(y1~y2+y3,data=datsim)

$$y_i = -0.05791 + 0.50117 * y2 + 0.26206 * y3 + e_i$$

Regresión

- mod2<-lm(y1~y2+y3,data=datsim)
- summary(mod2)
- anova(mod1,mod2)
- Variables independientes categoricas
- mod3<-lm(y1~y4,data=datsim)
- mod4<-lm(y1~y2+y3+y4,data=datsim)
- anova(mod2,mod4)

Modelos no lineales

- Cuadrático
  - mod5<-lm(y1~I(y2^2)+y3,data=datsim)
  - datsim$y22<-datsim$y2^2
  - mod52<-lm(y1~y22+y3,data=datsim)
  - anova(mod5,mod52)
- Interaccion
  - mod6<-lm(y1~y2*y3,data=datsim)
  - ps<-plotSlopes(mod6,plotx="y2",modx="y3")
  - testSlopes(ps)

Resumen

Que han aprendido?

Gracias