

Funciones básicas para el análisis estadístico



PROCESO	FUNCIÓN	DESCRIPCIÓN	EJEMPLO
Preparar el entorno de trabajo	rm(list=ls())	Quita todos los elementos del environment	
	graphics.off()	Quita todas las gráficas	
	options(warn=-1)	Evita que aparezcan los warnings	
Abrir una base de datos y explorar su contenido	setwd()	Establecer la ruta de trabajo	setwd("C:/Documentos/")
	library()	Carga los paquetes que se van a utilizar	library(foreign)
	read.csv()	Carga una base con formato csv	read.csv("Datos.csv", header=T)
	read.dbf()	Carga una base con formato .dbf	read.dbf("Datos.dbf")
	read.dta	Carga una base con formato .dta	read.dta("Datos.dta")
	rm()	Quita bases de datos y variables	rm("base")
	data.frame()	Convierte las tablas a DataFrame	read.dta("Datos.dta")%>%data.frame()
Exploración inicial de la base de datos	names()	Muestra todas las variables de la base de datos	names(Datos)
	head()	Muestra los primeros registros de la tabla	head(Datos, 2)
	dim()	Tamaño de la base de datos	dim(Datos)

PROCESO	FUNCIÓN	DESCRIPCIÓN	EJEMPLO
Exploración del contenido de las variables	class()	Tipo de dato (integer, character, numeric)	class(Datos\$variable)
	as.character()	Cambiar el tipo de dato a carácter	Datos\$variable%>%as.character()
	as.numeric()	Cambiar el tipo de dato a numérico	Datos\$variable%>%as.numeric()
	factor()	Cambiar el tipo de dato a factor	factor(Datos\$variable)
	table()	Tabulado de una o dos variables	table(Datos\$variable)
	wtd.table()	Tabulado de una o dos variables (permite ponderar)	wtd.table(Datos\$variable)
	aggregate()	Agrega los datos de acuerdo con una variable	aggregate(Datos\$variable, by = list(Datos\$lista), min)
	min()	Estadísticos básicos	min(Datos\$variable)
	max()		max(Datos\$variable)
	mean()		mean(Datos\$variable)
Subconjuntos	select()	Seleccionar variables	D1=Datos%>%select(variable1, variable2)
	filter()	Seleccionar casos	D1=Datos%>%filter(variable1==1)
	drop_na()	Quita los registros que tengan missings en una variable	drop_na(variable)
	reshape()	Cambia de formato la tabla	reshape(idvar = "var1", timevar = "var2", direction = "wide")
	mutate()	Genera una nueva variable	D1=Datos%>%mutate(Y=mean(variable))

PROCESO	FUNCIÓN	DESCRIPCIÓN	EJEMPLO
Inferencia estadística	set.seed()	Fija la semilla en un número en particular	set.seed(123)
	shapiro.test()	Valida la normalidad de una variable	shapiro.test(Datos\$variable)
	var.test()	Valida si dos muestras están relacionadas	var.test(Datos\$variable)
	t.test()	Prueba t-student	t.test(Datos\$var1, mu=0.40)
	binom.test()	Prueba binomial	binom.test(x,n, p=0.5, alternative=c("two.sided"))
	wilcox.test()	Prueba Wilcox	wilcox.test(Datos\$var1, Datos\$var2)
	chisq.test()	Prueba Chi cuadrada	chisq.test(tabla)**

PROCESO	FUNCIÓN	DESCRIPCIÓN	EJEMPLO
Regresión lineal	lm()	Regresión lineal	lm(Y ~ X1 + X2, data = Datos)
		Conviene guardar los resultados en una lista (r1)	r1=lm(Y ~ X1 + X2, data = Datos)
	summary()	Muestra los resultados de la regresión	summary(r1)
Regresión logística	glm()	Regresión logística	glm(Y ~ X1 + X2, data = Datos, family="binomial")
		Conviene guardar los resultados en una lista (r2)	r2=glm(Y ~ X1 + X2, data = Datos, family="binomial")
	summary()	Muestra los resultados de la regresión	summary(r2)
	exp(coefficients())	Genera los momios de la regresión	exp(coefficients(resultados2))
Análisis de componentes principales	prcomp()	Desarrolla el análisis de componentes principales	prcomp(Datos, scale=TRUE)
		Conviene guardar los resultados en una lista (r3)	r3=prcomp(Datos, scale=TRUE)
	plot()	Genera una línea, si se grafica los resultados de los componentes principales se convierte en gráfica de sedimentación	plot(r3, type="lines")
K-medias	kmeans()	Genera los grupos (variables continuas), se debe especificar el número de grupos	kmeans(Datos, 3)
		Conviene guardar los resultados en una lista (r4)	r4=kmeans(Datos, 3)
	data.frame(Datos, r4\$cluster)	Pega los grupos en la base de datos	D1=data.frame(Datos, r4\$cluster)