

# Contenido Sistemático de Análisis Descriptivo y Exploratorio de datos con R

*Estudio R*

2017

## Contents

<b>Análisis Exploratorio de Datos con R</b>	<b>1</b>
Descripción . . . . .	1
Objetivos . . . . .	1
Metodología . . . . .	2
Temas . . . . .	2
1. Introducción a R . . . . .	2
2. Datos en R . . . . .	2
3. Funciones en R . . . . .	3
4. Estadística descriptiva: conceptos . . . . .	3
5. R Commander . . . . .	3
6. Población - Muestra . . . . .	4
7. Manejo de datos con R . . . . .	4
8. Sistemas de gráficas . . . . .	4
9. Análisis exploratorio de datos . . . . .	5
10. Análisis exploratorio de datos (ejemplo práctico) - CONSULTAS ICFES 2016-02 . . . . .	5
11. Análisis de Componentes Principales ( <i>ACP - PCA</i> ) . . . . .	6
12. Análisis de Cluster . . . . .	6

## Análisis Exploratorio de Datos con R

### Descripción

En este curso se presentan los aspectos básicos del manejo, visualización e interpretación de análisis descriptivo de datos, utilizando el sistema estadístico R; se destaca la importancia de los métodos exploratorios y sus aplicaciones concretas desde la perspectiva científica y organizacional.

### Objetivos

En este curso el usuario podrá:

- Entender la necesidad del uso de herramientas libres (como R) para la implementación de técnicas de estadística descriptiva y exploratoria.
- Conocer los conceptos básicos de Estadística Descriptiva y análisis exploratorio.
- Reconocer adecuadamente la aplicación de cada uno de los métodos exploratorios sobre un conjunto de datos reales.
- Aprender tópicos básicos de visualización y representación de la información.
- Identificar atipicidades en los datos.
- Conocer elementos de *Ciencia Reproducible* e identificar ventajas respecto al método tradicional.
- Adquirir algunas habilidades y destrezas en cuanto a la implementación de algoritmos para ejecutar análisis exploratorio de datos.

- Realizar consultas sobre conjuntos de datos, cuyo desarrollo será con R y R-Studio.

## Metodología

Las secciones de estudio son desarrolladas con base en la teoría y la aplicación directa de conceptos aprendidos, mediante reuniones presenciales; cuyo espacio de estudio está dirigido a la discusión sobre temas de estadística descriptiva y exploratoria, aplicaciones mediante software libre, reconocimiento de patrones o tendencias en los datos, graficación y representación de la información, entre otros.

## Temas

### 1. Introducción a R

- ¿Que es R?
- Historia de R
- ¿Por qué R?
- Ventajas de R
- Recursos de información
- Repositorio de instalación (CRAN)
- Interfaz RStudio
- Repositorio de instalación RStudio
- Bibliotecas en R
- Documentación en R

### 2. Datos en R

- Lenguaje de programación: concepto
- Estructura de datos
  - Vector
  - Matriz
  - Factor
  - Listas
  - Marco de datos
- Clases atómicas
  - Numéricos
  - Enteros
  - Carácteres
  - Lógicos
  - Complejos
- Operadores básicos
  - Suma
  - Resta
  - Producto
  - División
  - Operaciones vectorizadas
  - Operaciones con matrices
  - ...
- Estructura de una función
  - Nombre
  - Argumentos
  - Cuerpo

### 3. Funciones en R

- Función: concepto
- Flujogramas
- Estructura de una función
  - Cuerpo
  - Parámetros formales
  - Entorno
- Estructuras de control
  - Estructuras de decisión
    - \* If, else
  - Estructuras de repetición
    - \* For
    - \* While

### 4. Estadística descriptiva: conceptos

- Estadística: concepto
- Población: concepto
- Muestra: concepto
- Variable: concepto
- Variabilidad: concepto
- Tipos de variables
  - Cualitativas
    - \* Nominal
    - \* Ordinal
  - Cuantitativas
    - \* Discretas
    - \* Continuas
- Herramientas para análisis descriptivo de datos
  - Tablas
  - Gráficos
  - Resumen numérico

### 5. R Commander

- Datos
  - Importar
    - \* Desde R
    - \* Archivos .csv, .txt, xlsx, entre otros.
- Exploración de datos
  - Editar
  - Eliminar columnas
  - Eliminar filas
  - Gráficos de variables cualitativas
    - \* Diagrama de barras
    - \* Diagrama circular
  - Gráficos de variables cuantitativas
    - \* Dispersión
    - \* Boxplot
    - \* Histogramas

## 6. Población - Muestra

- Población: concepto
- Muestra: concepto
- Proporción: concepto
- Simulación
  - Ejercicio práctico
  - Población real
  - Simulación del tamaño de muestra
  - Distribución (histogramas)
- Tamaño de muestra
- Intervalo de confianza
- Error
- Histogramas

## 7. Manejo de datos con R

- Estadística descriptiva o exploratoria
- Organización de la información
- Bases de datos
  - Modelos
  - Matriz de datos
- Manipulación de datos en R
  - Importando datos
    - \* Valores separados por comas
    - \* Texto plano
    - \* Excel
  - Exportando datos
  - Bibliotecas
  - Resumen de datos
  - Tidyverse
    - \* Verbos
    - \* Gramática
  - Ejemplos
  - Gráficos

## 8. Sistemas de gráficas

- Líneas, símbolos y colores en R
- Biblioteca `base`
  - Lectura de datos (ejemplo práctico)
  - Manipulación de datos
    - \* Depuración
    - \* Cambio de formatos
    - \* Concatenando datos
  - Histograma
  - Densidad
  - Boxplot
  - Diagrama de barras
  - Diagrama de dispersión
  - Personalización de gráficos
- Biblioteca `lattice`

- Lectura de datos
- Histograma
- Densidad
- Boxplot
- Diagrama de barras
- Diagrama de dispersión
- Biblioteca `ggplot2`
  - Representación gráfica
    - \* Positiva
    - \* Negativa
  - Gramática de gráficas
  - Componentes de gráficas `ggplot2`
    - \* Datos
    - \* Mapeo estético
    - \* Objetos geométricos
    - \* Transformaciones estadísticas
    - \* Escalas
    - \* Coordenadas
    - \* Posición
  - Lectura de datos (ejemplo práctico)
  - Depuración
  - Histograma
  - Densidad
  - Boxplot
  - Diagrama de barras
  - Diagrama de dispersión
  - Personalización de gráficos

## 9. Análisis exploratorio de datos

- Bibliotecas
- Lectura y conversión de datos
- Consultas a la base de datos
  - ¿Qué se quiere responder?
  - ¿Cuál es la pregunta correcta?
  - ¿Cómo?
- Respuesta a consultas
  - Tablas
  - Gráficos
  - Resumen numérico
- Personalización de gráficos y tablas.

## 10. Análisis exploratorio de datos (ejemplo práctico) - CONSULTAS ICFES 2016-02

- 50 consultas orientadas a la resolución de problemas específicos en un ejemplo práctico de manejo de información, cuyo volumen (tamaño) es considerable; ligados a la representación y visualización de datos mediante las herramientas de graficación y resumen estadístico de R.
- Respuesta a consultas
  - Gráficos
  - Tablas
  - Estadísticos descriptivos
  - Moda
  - Mediana

- Rango
- Dispersión
- Media
- Visualización e interpretación de resultados

## 11. Análisis de Componentes Principales (*ACP - PCA*)

- Generalidades
- Objetivos del ACP
- Aplicaciones
- Interpretación geométrica
- Interpretación algebraica
- Ejemplo e implementación en R
  - Datos de gorriones (ejemplo de *selección natural*)
- Matriz de correlación
- Estandarización de variables
- Graficación Biplot
- Interpretación de las componentes

## 12. Análisis de Cluster

- Generalidades
- Objetivos del Cluster
- Aplicaciones
- Medidas de disimilitud
- Métodos de agrupamiento
- Ejemplo e implementación en R
  - Datos de gorriones (ejemplo de *selección natural*)
- Dendrogramas
- Componentes Principales + Cluster
- Graficación: proyecciones en el plano (ACP + Cluster)