

# Estadística Descriptiva con Datos Cualitativos

Ramon Ceballos

24/1/2021

## 1. Introducción al Análisis Estadístico

### 1.1 Tipos de análisis estadístico de los datos

Cuando tenemos una serie de datos que describen algunos aspectos de un conjunto de individuos queremos llevar a cabo un análisis estadístico. Estos análisis estadísticos se clasifican en:

- Análisis exploratorio, o descriptivo, si nuestro objetivo es resumir, representar y explicar los datos concretos de los que disponemos. La estadística descriptiva es el conjunto de técnicas que se usan con este fin (visualizar, representar, explicar y resumir esos datos).
- Análisis inferencial, si nuestro objetivo es deducir (inferir), a partir de estos datos, información significativa sobre el total de la población o las poblaciones de interés. Las técnicas que se usan en este caso forman la estadística inferencial. Engloba técnicas como los intervalos de confianza, el contraste de hipótesis...

A su vez, ambos análisis se encuentran interrelacionados.

Cualquier análisis estadístico inferencial empieza explorando los datos que se van a utilizar y muchas técnicas descriptivas son empleadas para estimar propiedades de una población.

Existe relación entre ambos. Cualquier análisis inferencial se suele empezar explorando los datos que se usarán así como también muchas técnicas descriptivas permiten estimar propiedades de la población de la que se ha extraído la muestra.

#### **Ejemplo**

La media aritmética de las alturas de una muestra de individuos nos da un valor representativo de esta muestra, pero también estima la media de las alturas del total de la población

#### **OBJETIVO DEL CURSO**

Nos centraremos en entender algunas técnicas básicas de la estadística descriptiva orientadas al análisis de datos.

Estas consistirán en una serie de medidas, gráficos y modelos descriptivos que nos permitirán resumir y explorar un conjunto de datos.

**Objetivo final:** entender los datos lo mejor posible.

## 1.2 Tipos de datos para el análisis estadístico descriptivo

Trabajamos con datos multidimensionales: observamos varias características de una serie de individuos (lugares u objetos).

Se registran en un archivo de ordenador con un formato preestablecido. Por ejemplo texto simple (codificado en diferentes formatos: ASCII, isolatin...), hojas de cálculo (archivos de Open Office o Excel), bases de datos, etc.

Una de las maneras básicas de almacenar datos es en forma de **tablas de datos**. En R hacemos uso de data frames.

En una tabla de datos cada columna expresa una variable, mientras que cada fila corresponde a las observaciones de estas variables para un individuo concreto.

- Los *datos* de una misma columna tienen que ser del mismo tipo, porque corresponden a observaciones de una misma propiedad.
- Las *filas* en principio son de naturaleza heterogénea, porque pueden contener datos de diferentes tipos.

Los tipos de datos que consideramos para el análisis descriptivo son los siguientes:

- Datos de tipo atributo, o cualitativos: Expresan una cualidad del individuo como puede ser el sexo, la especie, el país de origen... En R guardaremos las listas de datos cualitativos en vectores (habitualmente, de palabras), o en factores si vamos a usarlos para clasificar individuos.
- Datos ordinales: Similares a los cualitativos, con la única diferencia de que se pueden ordenar de manera natural. Por ejemplo, las calificaciones en un control (suspense, aprobado, notable, sobresaliente), la calidad de vida en un ecosistema (malo, regular, bueno)... En R guardaremos las listas de datos ordinales en **factores ordenados**.
- Datos cuantitativos: Se refieren a medidas, tales como edades, longitudes, etc. En R guardaremos las listas de datos cuantitativos en **vectores numéricos**.

### Ejemplo

```
#De las cinco columnas las 4 primeras son numericas y la ultima es un factor con tres niveles (factor no ordenado)  
head(iris,5)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species  
## 1         5.1         3.5         1.4         0.2   setosa  
## 2         4.9         3.0         1.4         0.2   setosa  
## 3         4.7         3.2         1.3         0.2   setosa  
## 4         4.6         3.1         1.5         0.2   setosa  
## 5         5.0         3.6         1.4         0.2   setosa
```

```
#Se dan 4 variables cuantitativas y una variable cualitativa (factor no ordenado)  
str(iris)
```

```
## 'data.frame':   150 obs. of  5 variables:  
## $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...  
## $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...  
## $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...  
## $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...  
## $ Species      : Factor w/ 3 levels "setosa","versicolor",...: 1 1 1 1 1 1 1 1 1 1 ...
```