Resumen Inferencia y Modelos Estadísticos

Introducción

* Población: conjunto de individuos o elementos de los cuales se busca una conclusión
* Muestra: Subconjunto de la población

Cap. 2: Conceptos básicos

* Nociones de datos
  + Los datos se almacenan en matrices
  + Existen tipos de variables, de las cuales no todas pueden tomar los mismos valores.
    - Numéricas: pueden tomar muchos valores numéricos
      * Continuas: cualquier valor en un intervalo del conjunto de Reales
      * Discretas: valores enteros no negativos
    - Categóricas: solo pueden tomar un valor en un conjunto acotado. Cada valor se denomina nivel.
      * Nominales: no existe un orden natural entre los niveles.
      * Ordinales: existe un orden natural entre los niveles.
  + Dos variables pueden ser:
    - Independenties
    - Dependientes
      * Asociación positiva: directamente proporcional
      * Asociación negativa: inversamente proporcional
  + Parámetro: cualquier número que describa una población en forma resumida (ej, promedio)
  + Estadístico: cualquier cantidad cuyo valor puede ser calculado a partir de datos muestrales (ej, media) (es una estimación del parámetro)
* Nociones de R
  + Importación de datos
    - Importar una matriz de datos (data frame) desde un txt o csv hay que considerar:
      * La primera fila son para nombres de las columnas o variables
      * La primera columna para nombres de las observaciones (únicos), solo se permite el uso de puntos y guiones bajos. Los nombres no deben empezar con un dígito.
    - Distingue mayúsculas
    - Sin filas en blanco
    - Sin comentarios
    - Si no hay valores, debe contener un NA
    - Formato fecha mm/dd/aaaa
    - setwd(“directorio”) permite establecer el directorio de trabajo de R
    - head(): muestra por consola las primeras 6 filas de la data frame
    - tail(): muestra por consola las últimas 6 filas de la data frame
  + Importación de paquetes:
    - Antes de utilizar un paquete, este debe ser instalado con install.packages(“nombre\_paquete”)
    - Para utilizar el paquete se debe colocar library(nombre\_paquete) o require(nombre\_paquete)
  + Construcción de una data frame:
    - As.Date() para dar formato de fecha
    - nombre <- c(“Pedro”, “Juan”, “Diego”) : crea un vector que contiene los nombres.
    - fecha\_nacimiento <- as.Date(c(" 2 0 0 8 -1 -2 5 " , " 20 0 6 -1 0 -4 ", " 2 0 0 8 -3 -2 7 ") crea un vector con fechas
    - Dataframe <- data.frame(nombre, fecha\_nacimiento) crea un dataframe

Cap. 3: Exploración de datos

* Datos numéricos
  + Estadísticas descriptivas
    - Cuando una medida se aplica a una muestra, corre=sponde a un estimador puntual.
    - Distribución de frecuencia: representa cuántas veces aparece cada valor para una variable.
    - Media (media aritmética o promedio)
      * Media muestral ()
      * Media poblacional (μ­x)
      * mean(nombre\_dataframe$variable) para calcular el promedio de una variable
      * sapply(nombre\_dataframe, mean) para calcular el promedio de todas las variables
    - Mediana: valor central de los valores ordenados
      * median()
    - Moda: valor que más se repite
      * Unimodal
      * Bimodal
      * Multimodal
      * Paquete modeest tiene la función mfv() para calcular la moda
    - Varianza y desviación estandar:
      * Se calcula en base a la desviación de las observaciones.
        + Distancia entre una observación y la media del conjunto de datos

QUEDE JUSTO DSP DE PAG 10