

# Exposición Estadística Descriptiva

Bioestadística 1 - UNAL Medellín

*Edimer David Jaramillo*

*22 de marzo de 2018*

## Contents

<b>Pautas generales</b>	<b>2</b>
Fecha . . . . .	2
Base de datos . . . . .	2
Objetivo . . . . .	3
Consideraciones . . . . .	3
<b>Bases de datos</b>	<b>3</b>
Base de datos Grupo 1 . . . . .	3
Descripción . . . . .	3
Variables . . . . .	3
Base de datos Grupo 2 . . . . .	4
Descripción . . . . .	4
Variables . . . . .	4
Base de datos Grupo 3 . . . . .	4
Descripción . . . . .	4
Variables . . . . .	4
Base de datos Grupo 4 . . . . .	5
Descripción . . . . .	5
Variables . . . . .	5
Base de datos Grupo 5 . . . . .	5
Descripción . . . . .	5
Variables . . . . .	5
Base de datos Grupo 6 . . . . .	5
Descripción . . . . .	5
Variables . . . . .	6
Base de datos Grupo 7 . . . . .	6
Descripción . . . . .	6
Variables . . . . .	6
Base de datos Grupo 8 . . . . .	6
Descripción . . . . .	6
Variables . . . . .	6
Base de datos Grupo 9 . . . . .	7
Descripción . . . . .	7
Variables . . . . .	7
Base de datos Grupo 10 . . . . .	7
Descripción . . . . .	7
Variables . . . . .	7
Base de datos Grupo 11 . . . . .	8
Descripción . . . . .	8
Variables . . . . .	8
Base de datos Grupo 12 . . . . .	8
Descripción . . . . .	8
Variables . . . . .	8

Base de datos Grupo 14 . . . . .	8
Variables . . . . .	9
Base de datos Grupo 15 . . . . .	9
Descripción . . . . .	9
Variables . . . . .	9
Base de datos Grupo 16 . . . . .	9
Descripción . . . . .	9
Variables . . . . .	10
Base de datos Grupo 17 . . . . .	10
Descripción . . . . .	10
Variables . . . . .	10
Base de datos Grupo 18 . . . . .	10
Descripción . . . . .	10
Variables . . . . .	10
Base de datos Grupo 19 . . . . .	11
Descripción . . . . .	11
Variables . . . . .	11
Base de datos Grupo 20 . . . . .	11
Descripción . . . . .	11
Variables . . . . .	11
Base de datos Grupo 21 . . . . .	12
Descripción . . . . .	12
Variables . . . . .	12
Base de datos Grupo 22 . . . . .	12
Descripción . . . . .	12
Variables . . . . .	12
Base de datos Grupo 23 . . . . .	12
Descripción . . . . .	12
Variables . . . . .	13

## Pautas generales

A continuación se describen las “normas” generales para la exposición de estadística descriptiva.

### Fecha

Las exposiciones serán los días **3, 4, 5 y 6** de abril de 2018.

### Base de datos

Cada grupo tiene asignada una base de datos con la cual desarrollará la exposición de estadística descriptiva, la exposición de inferencia estadística y redactará el documento final; este último, producto de las dos exposiciones.

Todas las bases de datos están contenidas en un archivo *.zip* que puede ser descargado desde la página web en el ítem “Datos” de exposición 1 (Trabajos).

## Objetivo

El objetivo principal de la exposición es utilizar las herramientas de estadística descriptiva para caracterizar un fenómeno a partir de la información registrada. En ese orden de ideas, cada grupo deberá plantear **8 consultas o más** a la base de datos, con la finalidad de describir **objetivamente** el problema planteado. Las consultas deben **obligatoriamente** cumplir con:

- Objetividad
- Relación con el fenómeno de interés
- Capacidad de generar hipótesis a partir de ellas

## Consideraciones

- El día de la exposición cada grupo deberá presentar únicamente **4 de las 8 consultas formuladas**. Se sugiere que las escogidas para la exposición cumplan con ser las de mayor importancia.
- El tiempo de exposición será máximo de **15 minutos** por grupo.
- En la exposición se espera como mínimo:
  - **Título** (conciso, objetivo e insesgado)
  - **Contextualización** (revisión de literatura o marco teórico)
  - **Objetivo** (conciso y diciente)
  - **Resultados y discusión** (consultas formuladas)
- Esta primera exposición es la base del trabajo final
- Se tendrá en cuenta la ortografía de la presentación.

## Bases de datos

Cada grupo tiene asignada una base de datos, que contiene información acerca de fenómenos de interés en Ciencias Biológicas. Cada base de datos representa procesos naturales diferentes e importantes en la experimentación, con dos características fundamentales:

- Datos reales
- Importancia práctica

Cada base de datos se describe a continuación de forma general, junto a las variables que la conforman.

### Base de datos Grupo 1

#### Descripción

Comportamiento productivo del maíz en ensayo multi-ambiente. Se registraron datos de la producción de maíz, cultivados en ambientes diferentes y con distintos niveles de nitrógeno en la fertilización.

La base de datos está en el archivo “G1\_MaizNitro.csv”

#### Variables

La tabla de datos tiene 5 variables que se describen a continuación:

- **Sitio:** 5 sitios diferentes de cultivo
- **Localidad:** 5 localidades diferentes
- **Rep:** repetición
- **Nitro:** dosis de nitrógeno (kg/hectárea)

- **Producción:** producción en Mg/hectárea

## Base de datos Grupo 2

### Descripción

Incidencia de Vitamina E y Cobre en el desempeño de los cerdos. Se registran pesos durante 12 semanas para cerdos de 21 camadas, sometidos a dietas con diferentes niveles de Vitamina E y Cobre.

Los datos están registrados en el archivo “G2\_Porcinos.csv”.

### Variables

Hay un total de 6 variables que se describen a continuación:

- **Peso:** peso promedio en kilogramos
- **Semana:** semana
- **Id\_cerdo:** identificación del animal
- **Vit\_e:** nivel de vitamina E (mg/kg de alimento):
  - 1: 0
  - 2: 100
  - 3: 200
- **Cu:** nivel de cobre (mg/kg de alimento):
  - 1: 0
  - 2: 35
  - 3: 175
- **Camada:** número de camada

## Base de datos Grupo 3

### Descripción

Crecimiento dental en cuyes. Se registra la longitud de odontoblastos (células encargadas de la dentinogénesis) en 60 conejillos de indias o cuyes (*Cavia porcellus*) con diferentes niveles de vitamina C en la ingesta a través de dos métodos de suplementación.

Los datos están registrados en el archivo “G3\_Dental.csv”.

### Variables

Hay un total de 3 variables que se describen a continuación:

- **Longitud:** longitud de odontoblastos (mm)
- **Suplemento:** suplemento de vitamina C:
  - AA: ácido ascórbico
  - JN: jugo de naranja
- **Dosis:** dosis de suplemento (mg/día):
  - 0.5
  - 1
  - 2

## Base de datos Grupo 4

### Descripción

Efecto de diferentes niveles de proteína en la dieta sobre el peso de pollos de engorde, desde el nacimiento hasta los 21 días de edad; los pesos fueron registrados cada 2 días hasta el día 20 más el registro del día 21.

Los datos están registrados en el archivo “G4\_Pollos.csv”.

### Variables

Hay un total de 4 variables que se describen a continuación:

- **Peso:** peso en gramos
- **Edad:** edad del animal en días
- **Id:** identificación del animal
- **Dieta:** nivel de proteína en la dieta (1, 2, 3 y 4). Siendo 1 el de menor cantidad proteína y el 4 el de mayor cantidad.

## Base de datos Grupo 5

### Descripción

Evaluación de la toxicidad de agentes contaminantes en sistemas acuáticos. Se registró el número de crías vivas para tres crianzas diferentes de *Ceriodaphnia dubia* sometidas a diferentes dosis de Nitrofen (herbicida utilizado en Estados Unidos hasta hace algunos años).

Los datos están en el archivo “G5\_Nitrofen.csv”.

### Variables

La base de datos tiene 5 variables que se describen a continuación:

- **conc:** concentración de Nitrofen en la solución (taza/litro):
  - 0
  - 80
  - 160
  - 235
  - 310
- **C1:** número de crías vivas en la primera etapa reproductiva
- **C2:** número de crías vivas en la segunda etapa reproductiva
- **C3:** número de crías vivas en la tercera etapa reproductiva
- **Total:** número de crías vivas total

## Base de datos Grupo 6

### Descripción

Evaluación del comportamiento de la proteína en leche en el primer tercio de lactancia. Se registraron datos para el porcentaje de proteína en leche en 79 animales durante algunas semanas de producción, los animales fueron sometidos a tres tipos de dieta.

Los datos están en el archivo “G6\_Leche.csv”

## Variables

La base de datos tiene 3 variables que se describen a continuación:

- **Dieta:** tipo de dieta (3 en total):
  - Semi: semiestabulado (medio tiempo en pastoreo y medio tiempo en estabulación)
  - Pastoreo
  - Estabulado
- **Id:** identificación del animal
- **Semana:** semana de lactancia
- **Proteína:** proteína en leche (%)

## Base de datos Grupo 7

### Descripción

Producción y otras características de interés en Soya. Se registraron datos de producción, altura de la planta, tamaño de la semilla, porcentaje de grasa y porcentaje de proteína. El experimento fue desarrollado en dos años, en 4 localidades diferentes y 9 ambientes para cada una de estas, con un total de 58 genotipos. El experimento fue desarrollado en Australia.

La base de datos está en el archivo “G7\_Soya.csv”.

## Variables

La base de datos tiene un total de 9 variables que se describen a continuación:

- **Ambiente:** 9 ambientes (categorización basada en características ambientales)
- **Localidad:** localidad donde se realizó el experimento (4 localidades)
- **Año:** años 1970 y 1971
- **Gen:** genotipo (58 genotipos)
- **Producción:** producción en toneladas por hectárea
- **Altura:** altura de la planta en metros
- **Semilla:** tamaño de la semilla (milímetros)
- **Proteína:** porcentaje de proteína
- **Grasa:** porcentaje de grasa

## Base de datos Grupo 8

### Descripción

Efecto de la reproducción aumentada sobre la longevidad en machos de *Drosophila melanogaster*. Se registran datos para la longitud del tórax y el porcentaje de tiempo que dedican a dormir, cuando se someten a convivencia con diferentes tipos de hembras.

Los datos están en el archivo “G8\_Moscas.csv”

## Variables

La base de datos tiene 7 variables que se describen a continuación:

- **ID:** identificación del animal

- **Parejas:** número de parejas hembras:
  - 0
  - 1
  - 8
- **TipoP:** tipo de pareja:
  - 0: embarazada
  - 1: virgen
  - 9: ninguna
- **Longevidad:** tiempo de vida (en días)
- **Tórax:** longitud del tórax (mm)
- **Sueño:** porcentaje del día durmiendo
- **Tratamiento:** grupos experimentales

## Base de datos Grupo 9

### Descripción

Análisis de datos longitudinales (crecimiento en árboles) en *Picea sitchensis* cultivados en cámaras enriquecidas con ozono y sin ozono. Se registraron datos para el tamaño del tronco en ocho tiempos diferentes.

Los datos están en el archivo “G9\_Arboles.csv”

### Variables

Las base de datos tiene 4 variables que se describen a continuación:

- **Tamaño:** tamaño del árbol (altura por diámetro cuadrado)
- **Tiempo:** 8 tiempos de medición (días)
- **Árbol:** identificación del árbol
- **Tratamiento:** con o sin ozono (control)

## Base de datos Grupo 10

### Descripción

Rendimiento productivo en patatas. Se registró el rendimiento en kilogramos por hectárea de papa, para 15 variedades diferentes, sometidas a un diseño experimental con bloques.

Los datos están registrados en el archivo “G10\_Patatas.csv”

### Variables

Las base de datos tiene 6 variables que se describen a continuación:

- **Bloque:** bloque
- **Variedad:** variedad (14 variedades)
- **Fila:** fila del diseño
- **Columna:** columna del diseño
- **Producción:** rendimiento en kilogramos por hectárea
- **Maduración:** grupo de maduración (M1, M2, M3), siendo M1 el de menor grado de maduración y M3 el de mayor grado.

## Base de datos Grupo 11

### Descripción

Evaluación del rendimiento en guisantes bajo experimentación. Se registraron datos en un experimento factorial clásico N, P, K (nitrógeno, fósforo, potasio) con 6 bloques.

Los datos están en “G11\_Guisantes.csv”

### Variables

La tabla de datos tiene 5 variables que se describen a continuación:

- **Bloque:** bloque del experimento (cada bloque es una parcela)
- **N:** presencia (1) o ausencia (0) de nitrógeno
- **P:** presencia (1) o ausencia (0) de fósforo
- **K:** presencia (1) o ausencia (0) de potasio
- **Producción:** producción en libras por parcela

## Base de datos Grupo 12

### Descripción

Evaluación del sistema de riego por exudación con cuatro variedades diferentes de melón. Se registraron datos de producción en cuatro variedades de melón, cuya finalidad es comparar el rendimiento productivo de los híbridos (variedades).

La base de datos está en el archivo “G12\_Melon.csv”

### Variables

La tabla de datos tiene 4 variables que se describen a continuación:

- **Fila:** fila del experimento
- **Columna:** columna del experimento
- **Variedad:** variedad de melón:
  - V1: Híbrido Mission
  - V2: Híbrido Mark
  - V3: Híbrido Topfligh
  - V4: Híbrido Hales Best Jumbo
- **Producción:** producción en kilogramos

## Base de datos Grupo 14

Altura promedio para 15 genotipos de cebada en cada uno de 9 años (1974 - 1982) en Noruega. Con 19 registros de variables ambientales en cada uno de los 9 años. La etapa de crecimiento de cada año fue dividida en ocho periodos desde la siembra hasta la cosecha. Sólo los primeros 6 periodos se incluyen aquí.

Los datos están registrados en dos archivos diferentes. Los datos de variables ambientales están en el archivo “G14\_Ambientales.csv” y los de altura están en “G14\_Altura.csv”.



## Variables

Hay un total de 22 variables que se describen a continuación:

- **year:** Año
- **R1:** precipitación media (mm / día) en el período 1
- **R2:** precipitación media (mm / día) en el período 2
- **R3:** precipitación media (mm / día) en el período 3
- **R4:** precipitación media (mm / día) en el período 4
- **R5:** precipitación media (mm / día) en el período 5
- **R6:** precipitación media (mm / día) en el período 6
- **S1:** radiación solar diaria ( $W/cm^2$ ) en el período 1
- **S2:** radiación solar diaria ( $W/cm^2$ ) en el período 2
- **S3:** radiación solar diaria ( $W/cm^2$ ) en el período 3
- **S4:** radiación solar diaria ( $W/cm^2$ ) en el período 4
- **S5:** radiación solar diaria ( $W/cm^2$ ) en el período 5
- **S6:** radiación solar diaria ( $W/cm^2$ ) en el período 6
- **T1:** temperatura promedio (grados Celsius) en el período 1
- **T2:** temperatura promedio (grados Celsius) en el período 2
- **T3:** temperatura promedio (grados Celsius) en el período 3
- **T4:** temperatura promedio (grados Celsius) en el período 4
- **T5:** temperatura promedio (grados Celsius) en el período 5
- **T6:** temperatura promedio (grados Celsius) en el período 6
- **Genotipo:** genotipo
- **Altura:** altura (cm)

## Base de datos Grupo 15

### Descripción

Efecto de la tuberculosis y la dosificación de hierro sobre el peso de las vacas. Se registran datos de peso en vacas con y sin dosificación de hierro e infectadas o no con el microorganismo.

Los datos están registrados en “G15\_Vacas.csv”

## Variables

La tabla de datos tiene 5 variables que se describen a continuación:

- **Animal:** identificación del animal
- **Hierro:** dosificación de hierro (hierro y sin hierro)
- **Infectado:** infectado con el microorganismo (infectado y no infectado)
- **Peso:** peso en kilogramos
- **Día:** día de pesaje

## Base de datos Grupo 16

### Descripción

Producción y número de plantas en remolacha bajo experimentación. Se registraron datos para la producción y el número de plantas de remolacha sometidas a diferentes tipos de fertilización en un experimento con bloques.

Los datos están en el archivo “G16\_Remolacha.csv”

## Variables

La tabla de datos tiene 4 variables que se describen a continuación:

- **Fertilizante:** tipo de fertilizante utilizado (7 en total)
- **Bloque:** bloque del experimento (6 en total)
- **Producción:** producción en toneladas por hectárea
- **nPlantas:** número de plantas

## Base de datos Grupo 17

### Descripción

Peso al nacimiento y al destete en corderos Dorper x Red Maasi. Se registraron pesos durante 6 años (1991 - 1996) en corderos de ambos sexos, puros y cruzados.

La base de datos está en el archivo “G17\_Corderos.csv”

## Variables

La tabla de datos tiene 6 variables que se describen a continuación:

- **Año:** años del experimento (1991 - 1996)
- **Sexo:** sexo (hembras: F y Machos: M)
- **Raza:** Raza (puras: DD y RR, cruces: DR y RD)
- **PesoNac:** peso al nacimiento en kilogramos
- **PesoDes:** peso al destete en kilogramos
- **EdadDes:** edad al destete en días

## Base de datos Grupo 18

### Descripción

Medidas repetidas del peso en terneros bajo experimentación. Se registraron pesos (11 pesajes para cada animal) para 60 animales, sometidos a dos tipos de tratamiento para controlar larvas ingeridas en pastoreo.

Los datos están en el archivo “G18\_Terneros.csv”

## Variables

La tabla de datos tiene 4 variables que se describen a continuación:

- **Animal:** 30 animales para cada tratamiento
- **Tto:** tratamiento para control de patógenos:
  - A
  - B
- **Día:** día de pesaje (desde el nacimiento hasta el día 133)
- **Peso:** peso del animal en kilogramos

## Base de datos Grupo 19

### Descripción

Altura de árboles de Eucalipto en el sur de Brasil. Se registraron datos para la altura de árboles de Eucalipto en 70 genotipos de origen diferente y sembrados en localidades distintas del sur de Brasil.

Los datos están en el archivo “G19\_Eucalyptus.csv”

### Variables

La tabla de datos tiene 4 variables que se describen a continuación:

- **Gen:** genotipos
- **Origen:** los genotipos tienen diferente origen cada uno (3 en total)
- **Localidad:** localidad de siembra (en Brasil):
  - L1: Barra Ribeiro
  - L2: Telemaco Borba
  - L3: Esperança de Sul
  - L4: Guanhaes
  - L5: Ipatinga
  - L6: Aracruz
  - L7: Cacapua
- **Altura:** altura del árbol en metros

## Base de datos Grupo 20

### Descripción

Heterosis y características agronómicas en patatas del Perú. Se registraron datos para heterosis y algunas características agronómicas en cultivo de papa en familias TPS. El experimento se desarrolló en dos lugares distintos, en hembras y machos con varias réplicas y bajo diferentes tratamientos.

La base de datos está en el archivo “G20\_Heterosis.csv”

### Variables

La tabla de datos tiene 11 variables que se describen a continuación:

- **Lugar:** La Molina (1) y Huancayo (2)
- **Replica:** número de replicación del experimento
- **Tratamiento:** 36 tratamientos
- **Factor:** factor que indica la progenie, el progenitor o el control donde fueron aplicados los tratamientos
- **Hembra:** 8 hembras
- **Macho:** 3 machos
- **Producción:** kg/planta
- **Azúcares:** azúcares reducidos (1 y 5, nivel más bajo y alto, respectivamente)
- **MS:** materia seca del tubérculo (%)
- **Tubérculos:** número de tubérculos por planta
- **pesoTub:** peso promedio (g) del tubérculo

## Base de datos Grupo 21

### Descripción

Evaluación de la producción en Sorgo mediante ensayo multi-ambiente en Etiopía. Se registraron datos de producción durante 5 años (2001 - 2005), en 28 genotipos de Sorgo, con ensayos diferentes y en localidades distintas.

La base de datos está en el archivo “G21\_Sorgo.csv”

### Variables

La tabla de datos tiene 6 variables que se describen a continuación:

- **Gen:** genotipos de sorgo (28 genotios)
- **Ensayo:** ensayo:
  - T1: híbridos y polinización libre
  - T2: líneas experimentales
- **Ambiente:** correspondiente a épocas (13 ambientes)
- **Producción:** producción en kilogramos por hectárea
- **Año:** años 2001 a 2005
- **Localidad:** tres localidades diferentes

## Base de datos Grupo 22

### Descripción

Comportamiento productivo en ovejas. Se registraron datos para peso de animales de diferente raza y bajo cuatro tipos de alimentación o dieta.

La base de datos están en el archivo “G22\_Ovejas.csv”

### Variables

La tabla de datos tiene 4 variables que se describen a continuación:

- **Dieta:** cuatro tipos de dieta
- **Raza:** animales de 4 razas diferentes
- **Peso:** peso en libras
- **Mes:** mes en el cual se realizó el pesaje (el mes 0 corresponde al peso de nacimiento)

## Base de datos Grupo 23

### Descripción

Producción para 17 genotipos de maíz en Louisiana. Se registraron datos de producción en toneladas por hectarrea, durante tres años, en cuatro localidades diferentes, conformando un total de 12 ambientes o épocas distintas.

La base de datos está en el archivo “G23\_Maiz.csv”

## Variables

La tabla de datos tiene 5 variables que se describen a continuación:

- **Genotipo:** 17 genotipos diferentes
- **Ambiente:** 12 condiciones ambientales diferentes
- **Producción:** producción en toneladas por hectárea
- **Año:** año de registro (1985-1987)
- **Localidad:** cuatro localidades diferentes