



## COMPUTACIONAL ESTADÍSTICA

### LABORATORIO 3

**Base de datos:** *ds\_empleados.xlsx*

**Tamaño muestral:** 300 registros (trabajadores)

**Descripción general:** Contiene información simulada sobre características demográficas, contractuales y salariales de empleados de una empresa ficticia.

#### Variables

1. **Sexo** (*categorica*) — Género del trabajador: "Femenino" o "Masculino".
2. **Sueldo** (*numérica, entera*) — Salario mensual en soles, entre 1 150 y 20 580. Presenta distribución bimodal para simular distintos niveles de remuneración.
3. **Edad** (*numérica, entera*) — Edad del empleado (22 a 65 años).
4. **Estado\_Civil** (*categorica*) — Estado civil: "Soltero", "Casado", "Divorciado", "Viudo".
5. **Condicion** (*categorica*) — Tipo de contrato: "Estable" o "Contratado".
6. **Servicio** (*numérica, entera*) — Años de servicio en la empresa (0 a 15 años).
7. **AFP** (*categorica*) — Administradora de fondos de pensiones: "Horizonte", "Integra", "Prima", "Profuturo".

**Base de datos:** *ds\_ventas\_retail.xlsx*

**Tamaño muestral:** 240 registros (combinaciones de meses, regiones y categorías de producto).

**Descripción general:** Contiene datos de ventas y márgenes de utilidad para diferentes categorías de producto en diversas regiones geográficas, registrados mensualmente.

#### Variables

1. **Mes** (*numérica, entera*) — Mes del año (1 a 12).
2. **Region** (*categorica*) — Ubicación geográfica: "Costa", "Sierra", "Selva", "Norte Chico", "Sur Andino".
3. **Categoria** (*categorica*) — Tipo de producto: "Alimentos", "Bebidas", "Hogar", "Electrónica".
4. **Ventas** (*numérica, decimal*) — Monto total de ventas mensuales (en soles). Se simula mayor volumen a partir de mitad de año y para productos de electrónica.



5. **Margen** (*numérica, decimal*) — Margen de utilidad sobre ventas (proporción entre 0 y 1).

**Base de datos:** *ds\_alumnos.xlsx*

**Tamaño muestral:** 200 registros (estudiantes de pregrado).

**Descripción general:** Incluye información académica y de hábitos de estudio de estudiantes pertenecientes a distintos programas.

**VARIABLES:**

1. **Programa** (*categorica*) — Carrera o especialidad: "Administración", "Marketing", "Economía", "Ingeniería".
2. **Sexo** (*categorica*) — Género del estudiante: "Femenino" o "Masculino".
3. **Edad** (*numérica, entera*) — Edad del estudiante (18 a 35 años).
4. **HorasEstudio** (*numérica, decimal*) — Número de horas promedio de estudio por semana.
5. **Nota** (*numérica, decimal*) — Calificación promedio en una escala de 0 a 20. Se simuló relación positiva con las horas de estudio.
6. **GrupoEstudio** (*categorica*) — Modalidad de estudio: "Individual", "Parejas", "Grupo".

Con las Bases de datos mencionadas, se pide:

**Gráficos tradicionales**

Usa **base R** y, cuando aplique, cualquiera de los 3 datasets.

1. **Tallos y hojas (stem):** sobre Servicio en **ds\_empleados**. Interpreta la concentración de valores y outliers.
2. **Boxplot univariado:** Edad en **ds\_empleados** con main y ylab. Identifica el rango intercuartílico.
3. **Boxplot bivariado:** Edad ~ Sexo en **ds\_empleados** (horizontal y vertical). Compara medianas y dispersión.
4. **Histograma simple:** Sueldo en **ds\_empleados** con breaks = "Sturges" y labels = TRUE. Añade ylim adecuado.
5. **Histogramas por subgrupos:** dos paneles (par(mfrow=c(1,2))) separando Sueldo por Sexo en **ds\_empleados**.
6. **Pie chart:** distribución de AFP en **ds\_empleados**; añade etiquetas con nombre y porcentaje redondeado a 1 decimal.



7. **Barras simples:** conteo de Estado\_Civil en **ds\_empleados** con `legend = TRUE` y color personalizado.
8. **Barras apiladas y agrupadas:** cruza AFP y Estado\_Civil (usa `table()`), primero apiladas y luego `beside=TRUE`.
9. **Dispersiones (pairs):** HorasEstudio, Edad, Nota de **ds\_alumnos** con `pairs()`; colorea por Sexo.
10. **Dispersión 2D con anotaciones:** Nota vs HorasEstudio en **ds\_alumnos**; agrega `abline(lm(...))` y comenta la tendencia.

## Gráficos 3D y superficies

Puedes usar `scatterplot3d` y funciones base como `persp()`.

1. **scatterplot3d (Iris-like):** Con **ds\_alumnos**, grafica Nota (z), HorasEstudio (x), Edad (y).
2. **Color por grupo:** en el 3D anterior, colorea por GrupoEstudio.
3. **Etiquetas y pch:** personaliza `pch`, `xlab`, `ylab`, `zlab` y `main`.
4. **Plano de regresión (opcional):** añade plano ajustado  $\text{Nota} \sim \text{HorasEstudio} + \text{Edad}$  sobre el 3D si la función lo permite.
5. **Superficie con persp():** define una función  $f(x,y)=\sin(x)*\cos(y)$  en una malla  $x,y \in [-4,4]$ . Graficala.
6. **Superficies comparadas:** cambia parámetros (amplitud/frecuencia) y muestra dos `persp()` en par(`mfrow=c(1,2)`).
7. **Mapa de contornos (extra):** genera `contour(x,y,f)` para la superficie del ejercicio 5 y compárala con el `persp()`.
8. **Rotación de vista:** modifica `theta` y `phi` en `persp()` y comenta cambios de percepción.
9. **Interpolación con outer:** replica el esquema de **normal bivariada** del archivo para otra `rho` (p.e. 0.3) y comenta.
10. **Gráfico interactivo (si tienes rgl):** `plot3d(Nota, HorasEstudio, Edad)` en **ds\_alumnos** con color por Programa.

## Gráficos Lattice

Usa `lattice` con los 3 datasets (o el de iris si prefieres).

1. **bwplot univariante:** `bwplot(~ Sueldo, data = empleados)`; interpreta mediana y colas.



2. **bwplot condicionado:** `bwplot(Sueldo ~ Sexo | Condicion, data = empleados);` comenta diferencias entre paneles.
3. **histogram condicionado:** `histogram(~ Sueldo | AFP, data = empleados)` con `type="count"`; compara frecuencias.
4. **histogram con cortes:** usa `histogram(~ Ventas | Categoria, data = ventas_retail, breaks=20)`; ¿hay sesgos?
5. **xyplot simple:** `xyplot(Nota ~ HorasEstudio, data = alumnos);` añade `type=c("p","r")` para línea de regresión.
6. **xyplot condicionado:** `xyplot(Nota ~ HorasEstudio | Programa, groups = Sexo, auto.key = TRUE, data = alumnos).`
7. **densityplot por grupos:** `densityplot(~ Edad, groups = Sexo, auto.key=TRUE, data = empleados);` compara modos.
8. **barchart de conteos:** `barchart(table(empleados$Estado_Civil, empleados$AFP))` o usa `as.data.frame(table(...))`.
9. **pairs con splom:** `splom(~ alumnos[, c("Edad","HorasEstudio","Nota")], groups=Sexo, auto.key=TRUE).`
10. **Temas y paneles:** aplica un tema (`trellis.par.set`) y personaliza un panel (ej. `panel.abline`) en un `xyplot`.