

# Econometría I

## Reto 01

Carlos A. Yanes Guerra

Universidad del Norte | Departamento de Economía

### Contenido

Antes de empezar . . . . .	1
Brief de consultoría . . . . .	1
Objetivo . . . . .	1
Base simulada . . . . .	1
Preguntas del reto . . . . .	2
1) Tienda top/bottom y umbral de mediana . . . . .	2
2) Norte vs. Sur: nivel y dispersión . . . . .	3
3) Variable “Productividad” con <code>mutate</code> . . . . .	3
4) Momentos de “Productividad” . . . . .	4
5) Histograma + densidad (listo para comité) . . . . .	4
Tabla ejecutiva (para el memo) . . . . .	5
Mini-memo ( $\leq 150$ palabras) <code>echo=FALSE</code> . . . . .	5
Criterios de evaluación (rúbrica breve) . . . . .	5

### Antes de empezar

Recuerde que los retos se desarrollan en clase en **1h:20 min.** Trabaje solo o en pareja (máx. 2).  
Si el enunciado **solo pide resultado**, use `echo=FALSE` en ese *chunk*.

---

### Brief de consultoría

**Cliente:** *RetailCo* (110 tiendas).

**Pedido:** 1) foto de ventas y productividad; 2) comparación **Zona Norte vs. Zona Sur**; 3) un gráfico listo para comité.

**Entregables al final del Rmd:** - **Mini-memo** ( $\leq 150$  palabras): 3 hallazgos + 1 recomendación. - **1 tabla** de métricas clave. - **1 gráfico** (histograma+density) con línea en media o mediana.

---

### Objetivo

Realizar un **análisis descriptivo y de momentos** para **ventas y productividad**, comparando **Norte vs. Sur** (apoyarse en la *Guía de Econometría*).

---

### Base simulada

Puedes **mantener tu versión** original con `runif`. Aquí se ofrece una **opción más realista** (ventas asimétricas y leve prima en Norte). Las preguntas no cambian.

```

set.seed(228908)

tiendas <- 110

data_ventas <- tibble(
  tienda = 1:tiendas,
  zona = sample(c("Zona Norte", "Zona Sur"), tiendas, replace = TRUE),
  empleados = sample(5:50, tiendas, replace = TRUE)
) >
mutate(
  ventas_millones = round(
    rlnorm(n(), meanlog = 3.6, sdlog = 0.5) * if_else(zona == "Zona Norte", 1.07, 1), 2
  )
)

head(data_ventas)

```

```

## # A tibble: 6 x 4
##   tienda zona      empleados ventas_millones
##   <int> <chr>         <int>         <dbl>
## 1     1 1 Zona Sur          37          27.6
## 2     2 2 Zona Sur          40          50.2
## 3     3 3 Zona Norte         44          40.6
## 4     4 4 Zona Norte         29          36.5
## 5     5 5 Zona Sur          21          62.9
## 6     6 6 Zona Sur          13          67.0

```

Si prefieres tu versión original:

```

set.seed(228908)
tiendas <- 110
data_ventas <- data.frame(
  tienda = 1:tiendas,
  ventas_millones = round(runif(tiendas, min = 10, max = 100), 2),
  empleados = sample(5:50, tiendas, replace = TRUE),
  zona = sample(c("Zona Norte", "Zona Sur"), tiendas, replace = TRUE)
)
head(data_ventas)

```

## Preguntas del reto

### 1) Tienda top/bottom y umbral de mediana

¿Qué tienda vende más y cuánto? ¿Cuál vende menos y cuánto?

¿Cuál es el mínimo para estar por encima de la mediana de ventas?

> Muestre código y justifique en 1-2 líneas.

```

resumen_top_bottom <- data_ventas >
summarise(
  mediana = median(ventas_millones),
  max_venta = max(ventas_millones),
  tienda_max = tienda[which.max(ventas_millones)],
  min_venta = min(ventas_millones),
  tienda_min = tienda[which.min(ventas_millones)]
)

```

```
)
resumen_top_bottom

## # A tibble: 1 x 5
##   mediana max_venta tienda_max min_venta tienda_min
##   <dbl>   <dbl>   <int>   <dbl>   <int>
## 1    38.8    111.     67    12.8     71
```

## 2) Norte vs. Sur: nivel y dispersión

¿Qué tan distinto es el promedio de ventas entre zonas?

¿La desviación estándar difiere mucho? (compare numéricamente)

> Entregue una tabla por zona con: media, sd, p25, p50, p75.

```
tabla_zonas <- data_ventas %>%
  group_by(zona) %>%
  summarise(
    mean = mean(ventas_millones),
    sd = sd(ventas_millones),
    p25 = quantile(ventas_millones, .25),
    p50 = median(ventas_millones),
    p75 = quantile(ventas_millones, .75),
    .groups = "drop"
  )
knitr::kable(tabla_zonas, caption = "Ventas por zona: nivel y dispersión")
```

Table 1: Ventas por zona: nivel y dispersión

zona	mean	sd	p25	p50	p75
Zona Norte	40.61265	17.79879	30.14	37.50	45.50
Zona Sur	44.67869	20.75750	28.40	40.63	56.05

## 3) Variable “Productividad” con mutate

Defina:

$$\text{Productividad} = \frac{\text{Ventas} \times 1,000,000}{\text{No. empleados}}$$

```
data_ventas <- data_ventas %>%
  mutate(productividad = ventas_millones*1e6 / empleados)

data_ventas %>%
  select(tienda, zona, empleados, ventas_millones, productividad) %>%
  slice_head(n=5)
```

```
## # A tibble: 5 x 5
##   tienda zona      empleados ventas_millones productividad
##   <int> <chr>         <int>         <dbl>         <dbl>
## 1     1 1 Zona Sur         37          27.6         745946.
## 2     2 2 Zona Sur         40          50.2        1253750
## 3     3 3 Zona Norte        44          40.6        922955.
## 4     4 4 Zona Norte        29          36.5        1258621.
## 5     5 5 Zona Sur         21          62.9        2993810.
```

#### 4) Momentos de “Productividad”

Reporte **media**, **asimetría** y **curtosis (exceso)**.

Interprete: sesgo (derecha/izquierda) y colas (pesadas/ligeras). Conecte con el negocio (p.ej., “pocas tiendas ultra-productivas”).

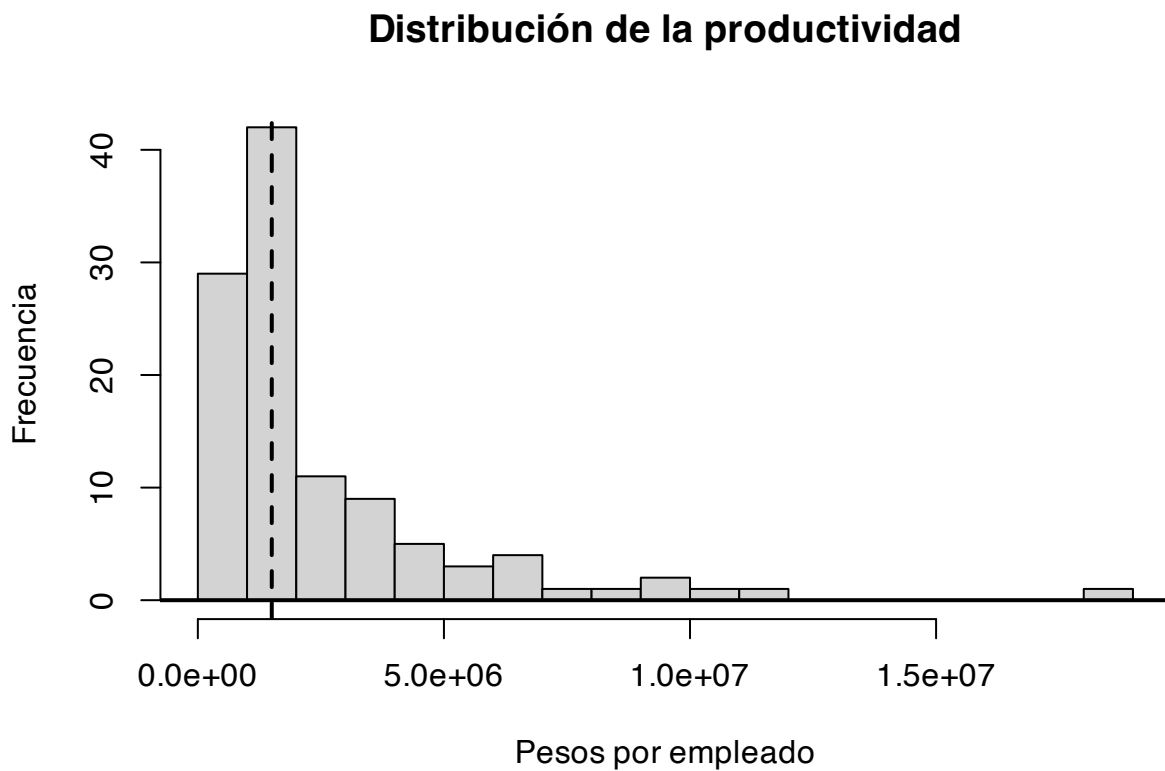
```
momentos_prod <- tibble(  
  media = mean(data_ventas$productividad),  
  asimetria = moments::skewness(data_ventas$productividad),  
  curtosis_exceso = moments::kurtosis(data_ventas$productividad) - 3  
)  
momentos_prod
```

```
## # A tibble: 1 x 3  
##   media asimetria curtosis_exceso  
##   <dbl>   <dbl>         <dbl>  
## 1 2513780.     2.89         10.9
```

#### 5) Histograma + densidad (listo para comité)

Incluya **histograma** con **línea de densidad** y una **línea vertical** en **media o mediana** (justifique su elección).

```
x <- data_ventas$productividad  
hist(x, breaks = 20,  
  main = "Distribución de la productividad",  
  xlab = "Pesos por empleado", ylab = "Frecuencia")  
lines(density(x), lwd = 2)  
abline(v = median(x), lwd = 2, lty = 2)
```



## Tabla ejecutiva (para el memo)

```
tabla_retail <- data_ventas >
group_by(zona) >
summarise(
  ventas_media = mean(ventas_millones),
  ventas_mediana = median(ventas_millones),
  prod_media = mean(productividad),
  prod_mediana = median(productividad),
  .groups = "drop"
)
knitr::kable(tabla_retail, caption = "Métricas clave por zona")
```

Table 2: Métricas clave por zona

zona	ventas_media	ventas_mediana	prod_media	prod_mediana
Zona Norte	40.61265	37.50	2782366	1583333
Zona Sur	44.67869	40.63	2298031	1489565

## Mini-memo (≤150 palabras) *echo=FALSE*

**Hallazgos:** (1) La Zona \_\_\_\_ presenta ventas medias \_\_\_\_% superiores; (2) La dispersión en \_\_\_\_ es mayor/menor, sugiriendo \_\_\_\_; (3) La productividad exhibe asimetría \_\_\_\_ (colas \_\_\_\_), consistente con \_\_\_\_.

**Recomendación:** Priorizar \_\_\_\_ en Zona \_\_\_\_ y ajustar dotación/procesos en tiendas con \_\_\_\_.

## Criterios de evaluación (rúbrica breve)

- **Correctitud numérica (40%)**
- **Evidencia (25%):** tabla por zona + histograma/densidad claro
- **Interpretación econométrica (25%)**
- **Presentación consultiva (10%):** mini-memo, títulos, unidades