

gaussian.R

Diego Jhoel Mamani Romero

2025-09-17

```
# 0. --- (Opcional) instalar paquetes la primera vez ---
# install.packages("readxl")
# install.packages("dplyr")
# install.packages("writexl") # para guardar .xlsx (alternativa: openxlsx)
```

```
# 1. Cargar librerías
library(readxl) # leer Excel
library(dplyr)  # manipulación de datos
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(writexl) # escribir Excel
```

```
## Warning: package 'writexl' was built under R version 4.4.3
```

```
# 2. Leer la base desde Excel (ajusta la ruta si es necesario)
base <- read_excel("D:/decimo/estadistica espacial/2024 - DESCOMPRIMIDO/973-Modulo1913/21_CAP1200B_ME.x")
```

```
# 3. Comprobaciones y limpieza básica (recomendado)
# Asegurarse de que REGION es numérica (evitar factores/strings)
if(!is.numeric(base$REGION)) {
  base$REGION <- as.numeric(as.character(base$REGION))
}
# Opcional: rellenar NA en REGION con la mediana (si hay NAs)
# base$REGION[is.na(base$REGION)] <- median(base$REGION, na.rm = TRUE)
```

```
# 4. Crear la variable REGION_GAUSS
set.seed(123) # semilla para reproducibilidad
base <- base %>%
  ungroup() %>% # evitar problemas si la tabla está "grouped"
```

```
mutate(REGION_GAUSS = rnorm(n = n(), mean = REGION, sd = 0.5))

# 5. Revisar resultados
head(base) # primeras filas con la nueva columna

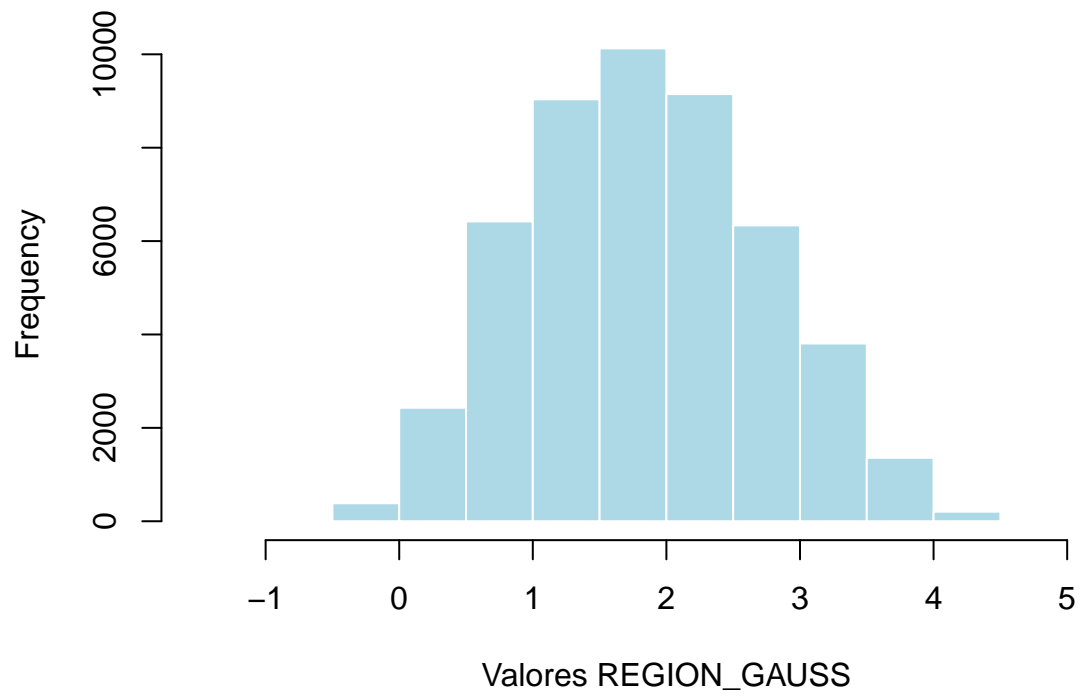
## # A tibble: 6 x 26
##   ANIO CCDD NOMBREDD CCpp NOMBREPV CCDI NOMBREDI NSEGM ID_PROD UA RESFIN
##   <dbl> <chr> <chr>    <chr> <chr>    <chr> <chr>    <chr> <chr> <chr> <dbl>
## 1  2024 25   UCAYALI  03   PADRE AB~ 05   ALEXAND~ 10002 01    01    1
## 2  2024 25   UCAYALI  03   PADRE AB~ 05   ALEXAND~ 10002 01    01    1
## 3  2024 25   UCAYALI  03   PADRE AB~ 05   ALEXAND~ 10002 02    02    1
## 4  2024 25   UCAYALI  03   PADRE AB~ 05   ALEXAND~ 10002 02    02    1
## 5  2024 25   UCAYALI  03   PADRE AB~ 05   ALEXAND~ 10002 03    03    1
## 6  2024 25   UCAYALI  03   PADRE AB~ 05   ALEXAND~ 10002 03    03    1
## # i 15 more variables: REGION <dbl>, ESTRATO <chr>, FACTOR_PRODUCTOR <dbl>,
## # CODIGO <dbl>, P101A <lgl>, P142_1 <dbl>, P142_2 <dbl>, P142_3 <dbl>,
## # P1207_N <dbl>, P1207_NOM <chr>, P1207_COD <dbl>, P1207_TIPO <dbl>,
## # P1208 <dbl>, OMICAP1200B_ME <lgl>, REGION_GAUSS <dbl>

summary(base$REGION_GAUSS) # estadísticos descriptivos

##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -1.065  1.183   1.819   1.842   2.461   5.161

# 6. Visualizar: histograma
hist(base$REGION_GAUSS,
      main = "Distribución Gaussiana simulada en REGION",
      xlab = "Valores REGION_GAUSS",
      col = "lightblue",
      border = "white")
```

Distribución Gaussiana simulada en REGION



```
# 7. Guardar la base con la nueva variable (xlsx y/o csv)
write_xlsx(base, "D:/decimo/estadistica espacial/2024 - DESCOMPRIMIDO/973-Modulo1913/21_CAP1200B_ME_GAUSS.xlsx")
write_csv(base, "D:/decimo/estadistica espacial/2024 - DESCOMPRIMIDO/973-Modulo1913/21_CAP1200B_ME_GAUSS.csv",
          row.names = FALSE)
```