November 1, 2023

The results below are generated from an R script.

```r
# Librería para leer XLSX
library(readxl)
library(dplyr)

# leemos los datos previamente salvados
df <- read_excel("Dry_Bean_Dataset.xlsx")

# Número de observaciones y variables
dim(df)
```

```
## [1] 13611    17
```

```r
# Particiones

# mediante una semilla conseguimos que el ejercicio sea reproducible
set.seed(12321)

# creamos índices
ntotal <- dim(df)[1]
indices <- 1:ntotal
ntrain <- ntotal * .6
ntest <- ntotal *.2
indices.train <- sample(indices, ntrain, replace = FALSE)
indices.test  <- sample(indices[-indices.train],ntest,replace=FALSE)
indices.valid <- indices[-c(indices.train,indices.test)]

# Usamos el 60% de la base de datos como conjunto de entrenamiento, 20% como conjunto de test y 20% como
train  <- df[indices.train, ]
test   <- df[indices.test, ]
valid  <- df[indices.valid,]

dim(train)
```

```
## [1] 8166    17
```

```r
dim(test)
```

```
## [1] 2722    17
```

```r
dim(valid)
```

```
## [1] 2723    17
```

```r
# Media de la variable AREA
media.train=mean(train$Area)
sd.train=sqrt(var(train$Area))
mean(test$Area)
```

```
## [1] 52377.31

mean(valid$Area)

## [1] 53244.92

# Escalamos la variable.

train=
  train %>%
  mutate(Area_Scale=scale(Area))

mean(train$Area_Scale)

## [1] 1.944833e-17

var(train$Area_Scale)

##       [,1]
## [1,]    1

# Aplicamos la misma transformación en los datos de test y validación

test=
  test %>%
  mutate(Area_Scale=(Area-media.train)/sd.train)

mean(test$Area_Scale)

## [1] -0.02785735

var(test$Area_Scale)

## [1] 0.8295407

valid=
  valid %>%
  mutate(Area_Scale=(Area-media.train)/sd.train)

mean(valid$Area_Scale)

## [1] 0.001295361

var(valid$Area_Scale)

## [1] 1.024713
```

The R session information (including the OS info, R version and all packages used):

```
sessionInfo()

## R version 4.3.1 (2023-06-16)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 20.04.6 LTS
##
## Matrix products: default
## BLAS:   /usr/lib/x86_64-linux-gnu/atlas/libblas.so.3.10.3
```

```
## LAPACK: /usr/lib/x86_64-linux-gnu/atlas/liblapack.so.3.10.3;  LAPACK version 3.9.0
##
## locale:
##  [1] LC_CTYPE=es_ES.UTF-8       LC_NUMERIC=C              LC_TIME=es_ES.UTF-8
##  [4] LC_COLLATE=es_ES.UTF-8     LC_MONETARY=es_ES.UTF-8   LC_MESSAGES=es_ES.UTF-8
##  [7] LC_PAPER=es_ES.UTF-8       LC_NAME=C                 LC_ADDRESS=C
## [10] LC_TELEPHONE=C             LC_MEASUREMENT=es_ES.UTF-8 LC_IDENTIFICATION=C
##
## time zone: Europe/Madrid
## tzcode source: system (glibc)
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] readxl_1.4.3    caret_6.0-94    lattice_0.21-9   ggplot2_3.4.3    rpart.plot_3.1.1
## [6] rpart_4.1.19    caTools_1.18.2  dplyr_1.1.3      ISLR2_1.3-2
##
## loaded via a namespace (and not attached):
##  [1] gtable_0.3.4         xfun_0.40            recipes_1.0.8       tzdb_0.4.0
##  [5] vctrs_0.6.3          tools_4.3.1          bitops_1.0-7        generics_0.1.3
##  [9] stats4_4.3.1         parallel_4.3.1       proxy_0.4-27        tibble_3.2.1
## [13] fansi_1.0.5          highr_0.10           ModelMetrics_1.2.2.2 pkgconfig_2.0.3
## [17] Matrix_1.6-1.1       data.table_1.14.8    lifecycle_1.0.3     stringr_1.5.0
## [21] compiler_4.3.1       farver_2.1.1         tinytex_0.47        munsell_0.5.0
## [25] codetools_0.2-19     htmltools_0.5.6.1    class_7.3-22        yaml_2.3.7
## [29] prodlim_2023.08.28   pillar_1.9.0         MASS_7.3-60         gower_1.0.1
## [33] iterators_1.0.14     foreach_1.5.2        nlme_3.1-163        parallelly_1.36.0
## [37] lava_1.7.2.1         tidyselect_1.2.0     digest_0.6.33       stringi_1.7.12
## [41] future_1.33.0        reshape2_1.4.4       purrr_1.0.2         listenv_0.9.0
## [45] labeling_0.4.3       splines_4.3.1        fastmap_1.1.1       grid_4.3.1
## [49] colorspace_2.1-0     cli_3.6.1            magrittr_2.0.3      survival_3.5-7
## [53] utf8_1.2.3           e1071_1.7-13         future.apply_1.11.0 readr_2.1.4
## [57] withr_2.5.1          scales_1.2.1         lubridate_1.9.3     timechange_0.2.0
## [61] rmarkdown_2.25       globals_0.16.2       nnet_7.3-19         timeDate_4022.108
## [65] cellranger_1.1.0     hms_1.1.3            evaluate_0.22       knitr_1.44
## [69] hardhat_1.3.0        rlang_1.1.1          Rcpp_1.0.11         glue_1.6.2
## [73] pROC_1.18.4          ipred_0.9-14         rstudioapi_0.15.0   R6_2.5.1
## [77] plyr_1.8.9

Sys.time()

## [1] "2023-11-01 20:39:27 CET"
```