

12.-Visualización_matriz_correlacion_y_regresion_lineal

Abel_Sánchez

2024-06-28

```
knitr::opts_chunk$set(echo = TRUE)
# Configurar el directorio de trabajo
if (.Platform$OS.type == "windows") {
  setwd("C:/Users/FX506/Documents/Portafolio")
} else {
  setwd("~/Portafolio")
}

# Limpiar el entorno
rm(list = ls())
gc()
```

```
##          used (Mb) gc trigger (Mb) max used (Mb)
## Ncells 492827 26.4   1064847 56.9   686460 36.7
## Vcells 921266  7.1    8388608 64.0   1876651 14.4
```

```
# Limpiar el entorno
rm(list = ls())
gc()
```

```
##          used (Mb) gc trigger (Mb) max used (Mb)
## Ncells 493068 26.4   1064847 56.9   686460 36.7
## Vcells 921749  7.1    8388608 64.0   1876651 14.4
```

```
# Instalar y cargar paquetes necesarios
if(!require(corrplot)) install.packages("corrplot")
```

```
## Cargando paquete requerido: corrplot
```

```
## corrplot 0.92 loaded
```

```
if(!require(car)) install.packages("car")
```

```
## Cargando paquete requerido: car
```

```
## Cargando paquete requerido: carData
```

```
if(!require(ggplot2)) install.packages("ggplot2")
```

```
## Cargando paquete requerido: ggplot2
```

```
library(readxl)
library(dplyr)
```

```
##
```

```
## Adjuntando el paquete: 'dplyr'
```

```
## The following object is masked from 'package:car':
```

```
##
```

```
##     recode
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##     filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##     intersect, setdiff, setequal, union
```

```
library(stringr)
library(janitor)
```

```
##
```

```
## Adjuntando el paquete: 'janitor'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##     chisq.test, fisher.test
```

```
library(car)
library(ggplot2)
library(corrplot)
```

```
# Definir la ruta del archivo
```

```
file_path <- "data/minsa/raw/datos_ense2017_resumen.xls"
```

```
output_folder <- "output/minsa"
```

```
# Leer y limpiar los datos
```

```
percepcion_salud <- read_excel(file_path, sheet = "T1001_estado_de_salud", range = "A1:F9", col_types =
```

```
percepcion_salud_clean <- percepcion_salud %>%
```

```
  janitor::clean_names() %>%
```

```
  mutate(across(muy_bueno:muy_malo, ~ as.numeric(str_replace_all(., ",", "."))))
```

```
colnames(percepcion_salud_clean) <- c("grupo_etareo", "muy_bueno", "bueno", "regular", "malo", "muy_malo")
```

```
enfermedades_cronicas <- read_excel(file_path, sheet = "T1025_enfermedades_cronicas", range = "A1:AF9",
```

```
enfermedades_cronicas <- enfermedades_cronicas %>%
```

```
  select(mujeres, tensión_alta, artrosis, dolor_espalda_cervical, dolor_espalda_lumbar, diabetes, coles
```

```

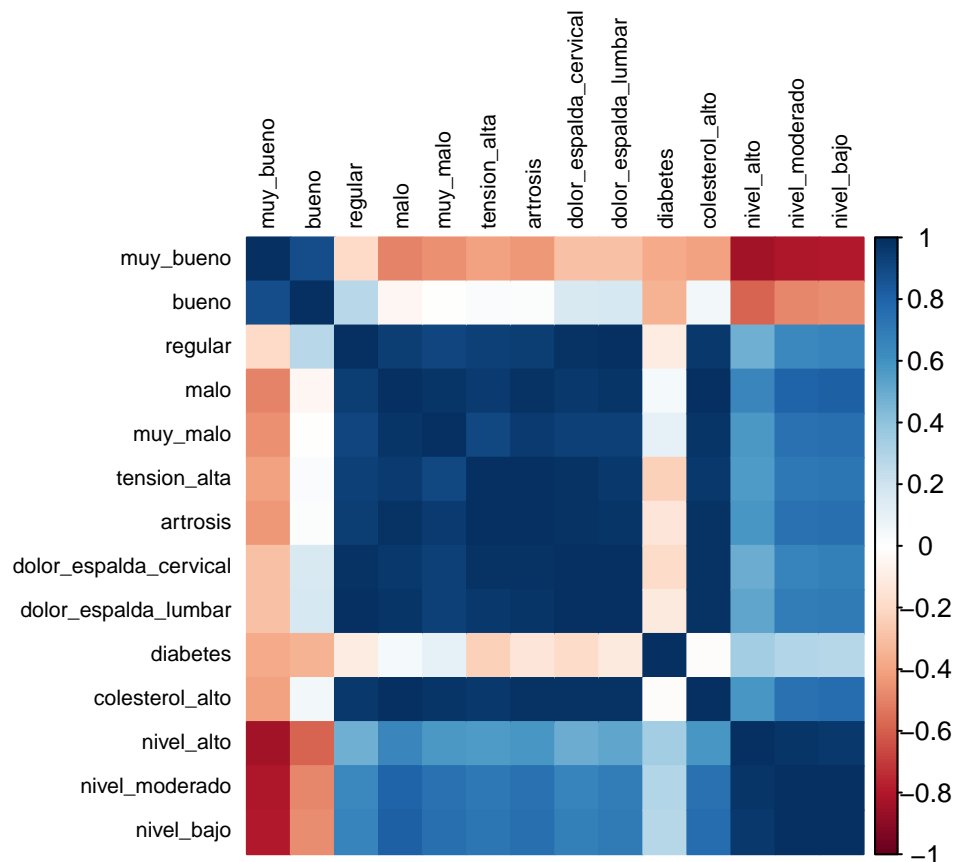
enfermedades_cronicas_clean <- enfermedades_cronicas %>%
  janitor::clean_names() %>%
  mutate(across(-mujeres, ~ as.numeric(str_replace_all(., ",", "."))))
colnames(enfermedades_cronicas_clean)[1] <- "grupo_etareo"

actividad_fisica <- read_excel(file_path, sheet = "T3066_nivel_actividad_fisica", range = "A1:D9", col_
actividad_fisica_clean <- actividad_fisica %>%
  janitor::clean_names() %>%
  mutate(across(nivel_alto:nivel_bajo, ~ as.numeric(str_replace_all(., ",", "."))))
colnames(actividad_fisica_clean)[1] <- "grupo_etareo"

# Combinar los datos
datos_combinados <- percepcion_salud_clean %>%
  inner_join(enfermedades_cronicas_clean, by = "grupo_etareo") %>%
  inner_join(actividad_fisica_clean, by = "grupo_etareo")

# Calcular y visualizar la matriz de correlación
datos_numericos <- datos_combinados %>% select(-grupo_etareo)
matriz_correlacion <- cor(datos_numericos, use = "complete.obs")
corrplot(matriz_correlacion, method = "color", tl.cex = 0.7, tl.col = "black")

```



```

# Guardar la matriz de correlación en un archivo .rds
saveRDS(matriz_correlacion, file.path(output_folder, "matriz_correlacion.rds"))

# Ajustar el modelo de regresión excluyendo variables altamente correlacionadas

```

```

modelo <- lm(nivel_bajo ~ muy_bueno + bueno + malo + muy_malo + tension_alta + artrosis, data = datos_c
# Guardar el resumen del modelo de regresión en un archivo .rds
saveRDS(summary(modelo), file.path(output_folder, "modelo_regresion.rds"))

# Detectar multicolinealidad
vif_valores <- vif(modelo)
print(vif_valores)

```

```

##      muy_bueno      bueno      malo      muy_malo tension_alta      artrosis
##      873.2114      630.4872  1155.0576  483.2242    4867.1429  11919.8469

```

```

# Resumen del modelo de regresión
summary(modelo)

```

```

##
## Call:
## lm(formula = nivel_bajo ~ muy_bueno + bueno + malo + muy_malo +
##      tension_alta + artrosis, data = datos_combinados)
##
## Residuals:
##      1      2      3      4      5      6      7      8
## -0.004828  0.038008 -0.007757 -0.537685  0.909511 -0.312031 -0.114947  0.029729
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   19.6479     11.7911   1.666   0.344
## muy_bueno     -4.3336      2.2582  -1.919   0.306
## bueno         10.9282      6.0078   1.819   0.320
## malo          -2.2765      1.6514  -1.379   0.400
## muy_malo      -0.6656      0.4109  -1.620   0.352
## tension_alta  -2.2562      1.5565  -1.450   0.384
## artrosis       3.9122      2.6638   1.469   0.381
##
## Residual standard error: 1.109 on 1 degrees of freedom
## Multiple R-squared:  0.9768, Adjusted R-squared:  0.8378
## F-statistic: 7.025 on 6 and 1 DF,  p-value: 0.2811

```

```

# Generar predicciones y residuos
predicciones <- predict(modelo, datos_combinados)
residuos <- residuals(modelo)
summary(cars)

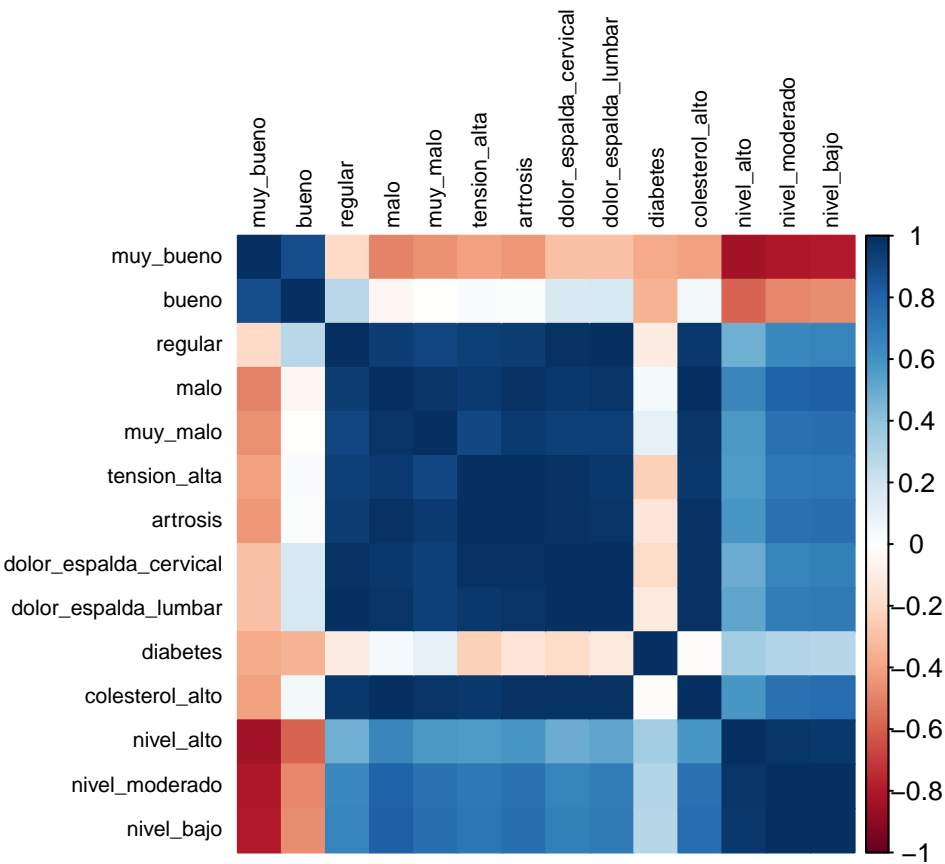
```

```

##      speed      dist
## Min.   : 4.0    Min.   : 2.00
## 1st Qu.:12.0    1st Qu.: 26.00
## Median :15.0    Median : 36.00
## Mean   :15.4    Mean   : 42.98
## 3rd Qu.:19.0    3rd Qu.: 56.00
## Max.   :25.0    Max.   :120.00

```

Including Plots



Residuos vs. Valores ajustados

