

Modelos lineales

Armando Ocampo

Librerías de trabajo

Antes de comenzar a trabajar, te recomiendo abrir las siguientes librerías. Si no cuentas con alguna de ellas es posible instalarla mediante la función `install.packages()`

```
library(readr)
library(dplyr)
library(corrplot)
library(ggplot2)
library(scales)
library(ggpubr)
library(broom)
```

Asimismo, se agregan dos conjuntos de datos. Los cuales debes descargar y guardar en la carpeta `clean_data`.

Modelos lineales

Para este apartado utilizaremos el conjunto de datos de vacunación obtenido del portal **Our World in Data**, el cual se conforma por información de 171 países, agregando las variables esperanza de vida, producto interno bruto *per capita* y porcentaje de vacunación de 13 inmunizaciones. Este dataset tiene actualización al 31 de diciembre de 2019. El formato es un archivo de valores separados por coma (*comma separate values, csv*). Antes de comenzar, lo llamaremos a nuestro ambiente de trabajo.

```
vacunas_df <- read_csv('../data/dataset_vacunas.csv')
```

Para facilitar el uso de las variables del conjunto de datos utilizaremos la función `attach()` de la paquetería *base*. Esta permite que cada variable del dataset se vuelva un vector independiente, sin saturar la memoria del ambiente.

```
attach(vacunas_df)
```

Los modelos lineales describen la relación de la variable respuesta (dependiente) y la(s) variable(s) explicativa(s) (independiente). Esta relación puede ser positiva, negativa o estar ausente. En R, este tipo de modelos se realiza mediante la función `lm()` de la paquetería *stats*. Los argumentos que se colocan son los siguientes. *formula = y_x*, describe la relación de la variable “y” con la variable “x” (*y_x*, este elemento indica, en medida que x explica y). El siguiente argumento, *data=* detalla el conjunto de datos a utilizar. En el siguiente ejemplo se creará un modelo lineal que explica la relación de la esperanza de vida con el producto interno bruto en el conjunto de datos de vacunas.

```
gdp_vs_life <- lm(Life_expectancy~GDP, data = vacunas_df)
```

Para extraer la información del modelo se utilizan las funciones `print()`, `summary()` y `tidy()`.

```
summary(gdp_vs_life)
```

```
##
## Call:
## lm(formula = Life_expectancy ~ GDP, data = vacunas_df)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.901  -3.883   1.317   4.144   8.623
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 6.756e+01  5.798e-01  116.53  <2e-16 ***
## GDP         2.485e-04  1.962e-05   12.66  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.386 on 169 degrees of freedom
## Multiple R-squared:  0.4869, Adjusted R-squared:  0.4839
## F-statistic: 160.4 on 1 and 169 DF,  p-value: < 2.2e-16
confint(gdp_vs_life)

##              2.5 %      97.5 %
## (Intercept) 6.641799e+01 6.870703e+01
## GDP         2.097573e-04 2.872285e-04
tidy(gdp_vs_life)

## # A tibble: 2 x 5
##   term          estimate std.error statistic    p.value
##   <chr>          <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept) 67.6      0.580     117.    2.30e-163
## 2 GDP         0.000248 0.0000196    12.7 2.83e- 26
```