

SISTEMATIZACION Y MÉTODOS ESTADÍSTICOS

SAN JUAN BAUTISTA

UNIVERSIDAD PRIVADA



ESTUDIANTE:

- VASQUEZ ROJAS EVAZ
- OLGA LUISA BELLIDO ARONI
- HENRRY STEVEN JIBAJA COPARA

CARGAR PAQUETES Y IMPORTAR LA DATA



Cargar los paquetes

```
{r}  
library(tidyverse)  
library(here)  
library(rio)  
library(gtsummary)  
library(car)  
library(survival)
```

pacientes total
de : 1000

```
{r}  
dengue_5 <- import(here("dengue_5"))
```

Un vistazo a los datos

```
{r}  
head(dengue_5)
```

A tibble: 6 × 11

genero <chr>	edad <dbl>	ns1 <chr>	ig_g <chr>	ig_m <chr>	zona <chr>
Femenino	45	Negativo	Negativo	Negativo	Mirpur
Masculino	17	Negativo	Negativo	Positivo	Chawkbazar
Femenino	29	Negativo	Negativo	Negativo	Paltan
Femenino	63	Positivo	Positivo	Negativo	Motijheel
Masculino	22	Negativo	Negativo	Negativo	Gendaria
Femenino	36	Negativo	Negativo	Positivo	Dhanmondi

6 rows | 1-6 of 11 columns

El análisis univariado

```
{r}
dengue_6 <- dengue_5 |>
  mutate(hipert = relevel(as.factor(genero), ref = "Masculino"), edades = relevel(as.factor(edad), ref = "45"),
         desenlace = relevel(as.factor(tipo_zona), ref = "No desarrollada")) |>
  na.omit()
```

Para obtener la tabla con los resultados del análisis univariado, se utiliza la función `tbl_uvregression()`, que permite generar tablas con las estimaciones de regresión logística para cada variable incluida



En esta tabla, los resultados se expresan como odds ratios no ajustados (OR) con sus respectivos intervalos de confianza al 95% y valores p

Characteristic	OR no ajustado	95% CI	Valor P
Masculino/femenino			
Femenino	—	—	
Masculino	0.95	0.74, 1.21	0.660
años			
Negativo/Positivo	1.00	0.99, 1.01	0.567
Negativo/Positivo			
Negativo	—	—	
Positivo	0.85	0.66, 1.09	0.205
Negativo/Positivo			
Negativo	—	—	
Positivo	0.85	0.66, 1.09	0.204
Mirpur/Chawkbazar/Paltan/Motijheel/Gendaria/Dhanmondi			
Adabor	—	—	
Badda	0.60	0.20, 1.80	0.366
Banasree	1.03	0.33, 3.14	0.964
Bangshal	0.74	0.25, 2.17	0.585
Biman Bandar	1.44	0.47, 4.47	0.524
Bosila	0.98	0.31, 3.17	0.979

```
{r}
tabla_reg_log_univ <- dengue_6 |>
  tbl_uvregression(
    include = c(genero, edad
, ns1, ig_g, ig_m,zona, tipo_zona),
    y = desenlace,
    method = glm,
    method.args = list(family = binomial),
    exponentiate = TRUE,
    conf.int = TRUE,
    hide_n = TRUE,
    add_estimate_to_reference_rows = FALSE,
    pvalue_fun = ~ style_pvalue(.x, digits = 3),
    estimate_fun = ~ style_number(.x, digits = 2),
    label = list(
      genero ~ "Masculino/femenino",
      edad ~ "años",
      ns1 ~ "Negativo/Positivo",
      ig_g ~ "Negativo/Positivo",
      ig_m ~ "Negativo/Positivo",
      zona ~ "Mirpur/Chawkbazar/Paltan/Motijheel/Gendaria/Dhanmondi",
      tipo_zona ~ "Desarrollada/No desarrollada ")
    ) |>
  bold_labels() |>
  bold_p(t = 0.05) |>
  modify_header(estimate = "***OR no ajustado***", p.value = "***valor p***")
```

El análisis multivariado

Para el análisis de regresión logística multivariada, se aplicó una estrategia de selección automática de variables utilizando tres enfoques: eliminación hacia atrás (backward elimination), selección hacia adelante (forward selection) y selección paso a paso (stepwise selection).

1

Ajuste del modelo inicial Ajustamos un modelo de regresión logística binaria que incluya todas las variables candidatas

```
{r}
var_modelo = glm(desenlace ~ genero + edad + ns1 + ig_g + zona +
  tipo_zona + tipo_vivienda,
  data = dengue_6,
  family = binomial(link = "logit"))
```

2

Realizamos la selección de variables usando la técnica Eliminación hacia atrás (Backward elimination)

```
{r}
multi_backward <- var_modelo |>
  step(direction = "backward", trace = FALSE)
```

3

Realizamos la selección de variables usando la técnica Selección hacia adelante (Forward selection).

```
{r}
multi_forward <- var_modelo |>
  step(direction = "forward", trace = FALSE)
```

4

Realizamos la selección de variables usando la técnica Selección paso a paso (Stepwise selection).

```
{r}
multi_stepwise <- var_modelo |>
  step(direction = "both", trace = FALSE)
```



función o módulo relacionado con el entrenamiento de redes neuronales, especialmente si estás usando bibliotecas como PyTorch.

Estimados el AIC para los modelos.

Podemos visualizar el AIC y cuáles variables han sido seleccionadas en cada modelo, usando la función `summary`.

```
[r]
summary(multi_backward)

call:
glm(formula = desenlace ~ tipo_zona, family = binomial(link = "logit"),
     data = dengue_6)

Coefficients:
                                         Estimate Std. Error
(Intercept)                      26.57    15910.45
tipo_zonaNo desarrollada -53.13    22523.31
                                         z value Pr(>|z|)
(Intercept)                      0.002    0.999
tipo_zonaNo desarrollada -0.002    0.998

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1.3863e+03 on 999 degrees of freedom
Residual deviance: 5.8016e-09 on 998 degrees of freedom
AIC: 4

Number of Fisher Scoring iterations: 25
```



Estimados el AIC para los modelos.

Podemos visualizar el AIC y cuáles variables han sido seleccionadas en cada modelo, usando la función `summary`.



```
{r}
summary(multi_forward)

call:
glm(formula = desenlace ~ genero + edad + ns1 + ig_g + zona +
    tipo_zona + tipo_vivienda, family = binomial(link = "logit"),
    data = dengue_6)

Coefficients:
              Estimate Std. Error
(Intercept) 2.657e+01 8.369e+04
generoMasculino -9.445e-12 2.296e+04
edad          1.574e-13 6.984e+02
ns1Positivo   -2.017e-11 1.014e+05
ig_gPositivo  2.710e-12 1.020e+05
zonaBadda    -5.043e-11 9.956e+04
zonaBanasree -3.618e-11 1.011e+05
zonaBangshal -5.385e-11 9.777e+04
zonaBiman Bandar -4.381e-10 1.003e+05
zonaBosila   -4.292e-11 1.053e+05
zonaCantonment -3.455e-11 1.104e+05
zonachawkbazar -1.311e-10 1.075e+05
zonaDemra    -6.242e-11 9.549e+04
zonaDhanmondi 3.806e-11 9.779e+04
zonaGendaria -4.022e-11 1.076e+05
zonaGulshan   -4.319e-11 1.016e+05
zonaHazaribagh -5.885e-11 1.043e+05
zonaJatrabari -8.926e-11 9.597e+04
zonaKadamtali -1.754e-10 9.773e+04
zonaKafrul    -6.035e-11 1.036e+05
zonaKalabagan -6.007e-11 1.027e+05
zonaKamrangirchar -8.103e-11 1.035e+05
zonaKeraniganj -6.293e-11 9.853e+04
zonaKhilgaon  -2.754e-11 1.064e+05
zonaKhilkhet  -6.679e-11 1.002e+05
zonaLalbagh   -5.310e-11 1.157e+05
```

zonaRampura	0.000	1.000
zonaSabujbagh	0.000	1.000
zonaShahbagh	0.000	1.000
zonasher-e-Bangla Nagar	0.000	1.000
zonashyampur	0.000	1.000
zonaSutrapur	0.000	1.000
zonaTejgaon	0.000	1.000
tipo_zonaNo desarrollada	-0.002	0.998
tipo_viviendaotro	0.000	1.000
tipo_viviendaTinshed	0.000	1.000

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1.3863e+03 on 999 degrees of freedom
Residual deviance: 5.8016e-09 on 957 degrees of freedom
AIC: 86

Number of Fisher Scoring iterations: 25



Estimados el AIC para los modelos.

Podemos visualizar el AIC y cuáles variables han sido seleccionadas en cada modelo, usando la función `summary`.

```
{r}
summary(multi_stepwise)

Call:
glm(formula = desenlace ~ tipo_zona, family = binomial(link = "logit"),
     data = dengue_6)

Coefficients:
                                         Estimate Std. Error
(Intercept)                      26.57    15910.45
tipo_zonaNo desarrollada -53.13    22523.31
                                         z value Pr(>|z|)
(Intercept)                      0.002    0.999
tipo_zonaNo desarrollada -0.002    0.998

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1.3863e+03 on 999 degrees of freedom
Residual deviance: 5.8016e-09 on 998 degrees of freedom
AIC: 4

Number of Fisher scoring iterations: 25
```



Evaluación de colinealidad

Finalmente, evaluamos la colinealidad usando la función `check_collinearity()` del paquete `performance`.



```
{r}
performance::check_collinearity(multi_backward, ci = NULL)

{r}
performance::check_collinearity(multi_forward, ci = NULL)

# check for Multicollinearity

Low Correlation

  Term  VIF adj. VIF Tolerance
genero 1.04    1.02    0.96
edad   1.04    1.02    0.96
zona   1.40    1.18    0.72
tipo_zona 1.05    1.02    0.96
tipo_vivienda 1.07    1.04    0.93

High Correlation

  Term  VIF adj. VIF Tolerance
ns1   20.24    4.50    0.05
ig_g  20.44    4.52    0.05
```

```
{r}
performance::check_collinearity(multi_stepwise, ci = NULL)
```

Reporte del análisis univariado y multivariado



Como en las sesiones anteriores, reportaremos los resultados del modelo final de regresión logística.

Tabla para los resultados de la regresión univariado (no ajustado)

```
{r}
tabla_univ <- dengue_6 |>
 tbl_uvregression(
  include = c(genero, edad, ns1, ig_g, ig_m, zona, tipo_zona),
  y = desenlace,
  method = glm,
  method.args = list(family = binomial),
  exponentiate = TRUE,
  conf.int = TRUE,
  hide_n = TRUE,
  add_estimate_to_reference_rows = FALSE,
  pvalue_fun = ~ style_pvalue(.x, digits = 3),
  estimate_fun = ~ style_number(.x, digits = 2),
  label = list(
    genero ~ "Masculino/femenino",
    edad ~ "años",
    ns1 ~ "Negativo/Positivo",
    ig_g ~ "Negativo/Positivo",
    ig_m ~ "Negativo/Positivo",
    zona ~ "Mirpur/Chawkazar/Paltan/Motijheel/Gendaria/Dhanmondi",
    tipo_zona ~ "Desarrollada/No desarrollada")
) |>
bold_labels() |>
bold_p(t = 0.05) |>
modify_header(estimate = "***OR***", p.value = "***valor P***")
```

Tabla para los resultados de la regresión multivariable (ajustado)

```
{r}
tabla_multi <- glm(
  desenlace ~ edad + ig_g + ig_m + resultado_1,
  family = binomial(link = "logit"),
  data = dengue_6
) |>
tbl_regression(
  exponentiate = TRUE,
  conf.int = TRUE,
  pvalue_fun = ~ style_pvalue(.x, digits = 3),
  estimate_fun = ~ style_number(.x, digits = 2),
  label = list(
    edad ~ "Edad (años)",
    ig_g ~ "Positivo/Negativo",
    ig_m ~ "Positivo/Negativo",
    resultado_1 ~ "tiempo post salida"
  )
) |>
bold_labels() |>
bold_p(t = 0.05) |>
modify_header(estimate = "***OR***", p.value = "***valor P***")
```




MUCHAS GRACIAS

