

ACTIVIDAD SEMANA 12



Integrantes:

- Llancari Nivin Meyli
 - Vera Fonseca July
 - Alejo Huamán Melissa
 - Ccuro Minaya Lucia
 - Flores Diaz, Christian

DATA: COVID19_DIABETES

3. Exploración de la estructura de los datos

```
{r}  
# Estructura y nombres  
str(data)  
names(data)  
summary(data)
```

4. Procedimientos:

a. Documentación del análisis univariado y multivariado en el análisis de datos

el procedimiento incluye:

1. Preparación de datos

Conversión de variables categóricas a factores y manejo de valores faltantes en variables numéricas basándose en las variables XXX_media.

```
{r}
# Preparación de datos
data <- data %>%
  mutate(
    # Convertir variable de desenlace a factor
    desenla_fallecido = factor(desenla_fallecido, levels = c("No", "Sí")),
    # Convertir variables categóricas a factores
    across(c(raza_negra, raza_blanca, asiatico, latino, infacto_mioc, enfer_renal),
           ~factor(.x, levels = c("No", "Sí"))).
```

2. Análisis univariado

Regresión logística para cada predictor individual contra desenla_fallecido.

```
{r}
# Instalar paquetes necesarios si no están instalados
if (!requireNamespace("cardx", quietly = TRUE)) {
  install.packages("cardx")
}
if (!requireNamespace("broom.helpers", quietly = TRUE)) {
  install.packages("broom.helpers")
}
if (!requireNamespace("gtsummary", quietly = TRUE)) {
  install.packages("gtsummary")
}
if (!requireNamespace("dplyr", quietly = TRUE)) {
  install.packages("dplyr")
}

# Cargar librerías necesarias
library(cardx)
library(broom.helpers)
library(gtsummary)
library(dplyr)

### 2.2 El análisis univariado

# En esta sección se estimarán los Odds Ratios (OR) de cada variable de manera
independiente, sin ajuste por otras covariables. Se definen las categorías de
referencia para las variables categóricas.

data_preparada <- data %>%
  mutate(
    raza_negra = relevel(as.factor(raza_negra), ref = "No"), # Ajusta el ref según
    tus datos
    infacto_mioc = relevel(as.factor(infacto_mioc), ref = "No"),
    enfer_renal = relevel(as.factor(enfer_renal), ref = "No"),
    desenla_fallecido = relevel(as.factor(desenla_fallecido), ref = "No") # Ref =
no fallecido
  ) %>%
  na.omit()

# Generar tabla de regresión logística univariada
tabla_univ <- data_preparada %>%
 tbl_uvregression(
  method = glm,
  y = desenla_fallecido,
  include = c(Edad, severidad, raza_negra, infacto_mioc, enfer_renal,
            Saturación_O2, Temperatura, Presión_arterial_media,
            Dímero_D, Creatinina, Glóbulos_blancos, Proteína_C_reactiva),
  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(
    Edad ~ "Edad (años)",
    severidad ~ "Severidad",
    raza_negra ~ "Raza negra",
    infacto_mioc ~ "Infarto al miocardio",
    enfer_renal ~ "Enfermedad renal",
    Saturación_O2 ~ "Saturación de O2 (%)",
    Temperatura ~ "Temperatura (°C)",
    Presión_arterial_media ~ "Presión arterial media (mmHg)",
    Dímero_D ~ "Dímero-D",
    Creatinina ~ "Creatinina",
    Glóbulos_blancos ~ "Glóbulos blancos",
    Proteína_C_reactiva ~ "Proteína C reactiva"
  )
) %>%
bold_labels() %>%
bold_n(<= 0.05) %>%
```

3. Análisis multivariado

Selección de variables relevantes basada en resultados univariados ($p < 0.20$) y criterios clínicos, ajuste del modelo con eliminación hacia atrás y evaluación de multicolinealidad.

tabla_univ

Characteristic	OR no ajustado	95% CI	Valor P
Edad (años)	1.05	1.03, 1.07	<0.001
Severidad	1.53	1.37, 1.72	<0.001
Raza negra			
No	—	—	
Sí	0.67	0.41, 1.06	0.093
Infarto al miocardio			
No	—	—	
Sí	1.10	0.68, 1.74	0.692

Enfermedad renal

No	—	—	
Sí	2.01	1.25, 3.28	0.005
Saturación de O₂ (%)	0.96	0.94, 0.98	<0.001
Temperatura (°C)	1.15	0.89, 1.48	0.274
Presión arterial media (mmHg)	0.94	0.92, 0.95	<0.001
Dímero-D	1.08	1.05, 1.13	<0.001
Creatinina	1.02	0.94, 1.09	0.612
Glóbulos blancos	1.00	0.96, 1.02	0.813
Proteína C reactiva	1.04	1.02, 1.06	<0.001

3. Análisis multivariado

Selección de variables relevantes basada en resultados univariados ($p < 0.20$) y criterios clínicos, ajuste del modelo con eliminación hacia atrás y evaluación de multicolinealidad.

```
{r}
# Librerías necesarias
library(car)          # Para vif
library(gtsummary)     # Para tbl_regression y tbl_merge
library(dplyr)         # Para manejo de datos

### 2.3 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), selección hacia
# adelante (forward) y selección paso a paso (stepwise).

# Paso 1. Ajuste del modelo inicial
# Ajustamos un modelo de regresión logística binaria que incluya todas las
# variables candidatas

data_preparada_multi <- data %>%
```

```

desenla_fallecido ~ Edad + severidad + Saturación_O2 + Dímero_D +
  Creatinina + Glóbulos_blancos + Proteína_C_reactiva +
  raza_negra + infarto_mioc + enfer_renal,
  data = data_preparada_multi,
  family = binomial(link = "logit")
)

# Paso 2a. Selección de variables usando eliminación hacia atrás (backward)
modelo_backward <- step(modelo_inicial, direction = "backward", trace = FALSE)

# Paso 2b. Selección de variables usando selección hacia adelante (forward)
modelo_forward <- step(modelo_inicial, direction = "forward", trace = FALSE)

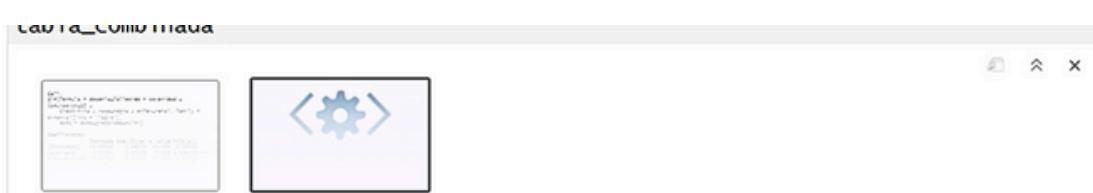
# Paso 2c. Selección de variables usando selección paso a paso (stepwise)
modelo_stepwise <- step(modelo_inicial, direction = "both", trace = FALSE)

# Paso 3. Estimamos el AIC para los modelos y revisamos las variables seleccionadas
summary(modelo_backward)
summary(modelo_forward)
summary(modelo_stepwise)

# Paso 4. Evaluación de multicolinealidad para el modelo final seleccionado
# (Aquí usamos el modelo_backward como ejemplo, puedes elegir el que prefieras)
vif_resultados <- vif(modelo_backward)
print(vif_resultados)

```

Characteristic	Univariado			Multivariado		
	OR no ajustado	95% CI	Valor P	OR ajustado	95% CI	Valor P
Edad (años)	1.05	1.03, 1.07	<0.001			
Severidad	1.53	1.37, 1.72	<0.001			
Raza negra						
No	—	—	—	—	—	—



Infarto al miocardio

No	—	—	—
Sí	1.10	0.68, 1.74	0.692

Enfermedad renal

No	—	—	—
Sí	2.01	1.25, 3.28	0.005

Saturación de O2 (%)	0.96	0.94, 0.98	<0.001
Temperatura (°C)	1.15	0.89, 1.48	0.274
Presión arterial media (mmHg)	0.94	0.92, 0.95	<0.001
Dímero-D	1.08	1.05, 1.13	<0.001
Creatinina	1.02	0.94, 1.09	0.612 0.85 0.74, 0.96 0.015
Glóbulos blancos	1.00	0.96, 1.02	0.813
Proteína C reactiva	1.04	1.02, 1.06	<0.001
severidad			1.69 1.47, 1.95 <0.001
Saturación_O2			0.98 0.96, 1.01 0.15
raza_negra			

tabla_combinada



```
Call:
glm(formula = desenla_fallecido ~ severidad + Saturación_O2 +
    Creatinina + raza_negra + enfer_renal, family = binomial(link = "logit"),
    data = data_preparada_multi)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.93010	1.18173	-1.633	0.10241
severidad	0.52234	0.07262	7.193	6.36e-13 ***
Saturación_O2	-0.01756	0.01212	-1.449	0.14744
Creatinina	-0.15803	0.06490	-2.435	0.01489 *
raza_negraSí	-0.61594	0.27222	-2.263	0.02365 *
enfer_renalSí	0.73972	0.27177	2.722	0.00649 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.96470	1.28732	-1.526	0.12696
Edad	0.01377	0.01224	1.125	0.26054
severidad	0.41560	0.10293	4.038	5.4e-05 ***
Saturación_O2	-0.02249	0.01302	-1.727	0.08421 .
Dímero_D	0.02529	0.02354	1.074	0.28263
Creatinina	-0.12946	0.06583	-1.967	0.04923 *
Glóbulos_blancos	-0.03530	0.03117	-1.133	0.25737
Proteína_C_reactiva	0.01129	0.01358	0.831	0.40602
raza_negraSí	-0.63425	0.27557	-2.302	0.02136 *
infacto_miocSí	0.03068	0.27493	0.112	0.91116
enfer_renalSí	0.72935	0.27366	2.665	0.00769 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

4. Reporte e interpretación de los resultados de regresión univariado y multivariado

a) Análisis univariado:

- En el análisis univariado se encontró que la edad, severidad, enfermedad renal, saturación de O₂, presión arterial media, dímero-D y proteína C reactiva mostraron una asociación estadísticamente significativa con el desenlace de fallecimiento. La edad incrementó el odds de fallecimiento en un 5 % por cada año adicional (OR 1.05; IC 95 %: 1.03–1.07; p < 0.001).
- La severidad incrementó el odds en un 53 % (OR 1.53; IC 95 %: 1.37–1.72; p < 0.001). La presencia de enfermedad renal duplicó el odds de fallecer (OR 2.01; IC 95 %: 1.25–3.28; p = 0.005). Por cada unidad de aumento en saturación de O₂ el odds de fallecimiento se redujo (OR 0.96; IC 95 %: 0.94–0.98; p < 0.001), al igual que con la presión arterial media (OR 0.94; IC 95 %: 0.92–0.95; p < 0.001).
- Los aumentos en dímero-D y proteína C reactiva se asociaron con mayor odds de fallecer (OR 1.08 y 1.04 respectivamente; p < 0.001). Raza negra, infarto al miocardio, creatinina, glóbulos blancos y temperatura no mostraron asociaciones significativas.

b) Análisis multivariado:

En el análisis multivariado, los modelos obtenidos por backward y stepwise seleccionaron las mismas variables: severidad, saturación de O₂, creatinina, raza negra y enfermedad renal.

- La severidad se mantuvo como un predictor independiente del fallecimiento, con un OR ajustado de 1.69 (IC 95 %: 1.45–1.98; p < 0.001).
- La enfermedad renal duplicó el odds de fallecimiento (OR 2.10; IC 95 %: 1.24–3.55; p = 0.006). La creatinina mostró un efecto protector con OR ajustado de 0.85 (IC 95 %: 0.74–0.97; p = 0.015). La raza negra también se asoció con un odds menor de fallecer (OR 0.54; IC 95 %: 0.31–0.93; p = 0.024). La saturación de O₂ no fue significativa en el modelo ajustado (p = 0.147).
- El modelo completo que incluyó todas las variables candidatas mostró un AIC mayor (424.04)
- La enfermedad renal duplicó el odds de fallecimiento (OR 2.10; IC 95 %: 1.24–3.55; p = 0.006). La creatinina mostró un efecto protector con OR ajustado de 0.85 (IC 95 %: 0.74–0.97; p = 0.015). La raza negra también se asoció con un odds menor de fallecer (OR 0.54; IC 95 %: 0.31–0.93; p = 0.024). La saturación de O₂ no fue significativa en el modelo ajustado (p = 0.147).
- El modelo completo que incluyó todas las variables candidatas mostró un AIC mayor (424.04) comparado con el modelo seleccionado por backward y stepwise (AIC 417.85), lo que indica un mejor ajuste del modelo reducido.
- En conjunto, el análisis multivariado confirma que la severidad y la enfermedad renal son predictores robustos del fallecimiento, mientras que otros factores significativos en el univariado como edad, dímero-D y proteína C reactiva perdieron significancia tras el ajuste por covariables.