

```
---
title: "Análisis univariado y multivariado con gtsummary"
author: "Grupo 01"
format: html
---
```

GRUPO 01 - Integrantes

- [Bastidas Bendezu Ivan Luis](#)
- [Basurto Taype Marcelo Aaron](#)
- [Fonseca Moron José Kenneth](#)
- [Saravia Gutierrez Randy Esteban](#)
- [Talavera Ayllon Angel Ronaldo](#)

Cargar los paquetes

```
{r}
install.packages("performance")
```

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

— Attaching core tidyverse packages

tidyverse 2.0.0 —			
✓ dplyr	1.1.4	✓ readr	2.1.5
✓ forcats	1.0.0	✓ stringr	1.5.1
✓ ggplot2	3.5.2	✓ tibble	3.2.1
✓ lubridate	1.9.4	✓ tidyr	1.3.1
✓ purrr	1.0.4		

— Conflicts

```
tidyverse_conflicts() —
✗ dplyr::filter() masks stats::filter()
✗ dplyr::lag() masks stats::lag()
i Use the conflicted package to force all conflicts to become errors
```

here() starts at D:/Programas
Universidad/UPSJB_Practica_RStudio/estadistica_upsjb
Some optional R packages were not installed and therefore some file formats are not supported. Check file support with show_unsupported_formats()
Cargando paquete requerido: carData

Adjuntando el paquete: 'car'

The following object is masked from 'package:dplyr':

recode

The following object is masked from 'package:purrr':

some

1 Modelos univariados (no ajustados) vs. multivariados (ajustados)

Hasta ahora, hemos explorado modelos de regresión que evalúan un predictor a la vez. A estos se les denomina modelos univariados o no ajustados, ya que solo consideran una variable predictora. Sin embargo, datasets utilizados en estas sesiones, al igual que muchos datos que probablemente recolectes, provienen de estudios observacionales. Es decir, no existe un control estricto sobre qué individuos se incluyen en el análisis y cuáles no. Esto implica que múltiples factores pueden influir en el desenlace de interés de manera simultánea.

Por esta razón, no es adecuado extraer conclusiones definitivas a partir de modelos no ajustados, ya que estos ignoran el efecto de posibles variables de confusión. En su lugar, es necesario realizar un análisis multivariado o ajustado, que permita considerar de manera simultánea varios predictores potenciales.

Por ejemplo, es poco probable que solo el tipo de accidente cerebrovascular (ACV) —isquémico o hemorrágico— determine la probabilidad de fallecer tras un evento de este tipo. Factores como la edad, el sexo, las comorbilidades preexistentes y los hábitos de vida también pueden afectar de manera importante este riesgo. Ignorar estas variables podría conducir a estimaciones sesgadas o erróneas.

1.1 Interpretación general del modelo ajustado

Cuando se incluyen varias covariables en un modelo de regresión, se obtienen medidas de efecto ajustadas, como el Odds Ratio ajustado (OR ajustado) en la regresión logística, o el riesgo relativo ajustado (RR ajustado) en la regresión de Cox. Estas medidas estiman la asociación entre una variable específica y el desenlace de interés, mientras se controla el efecto de las demás covariables incluidas en el modelo.

Por ejemplo, el OR ajustado para fallecer tras un ACV isquémico indica la fuerza de esta asociación independientemente de otros factores como la edad, el sexo o las comorbilidades del paciente.

En esta sesión aplicaremos tanto modelos univariados (no ajustados) como multivariados (ajustados), utilizando el dataset previamente analizados en sesión de regresión logística.

1.2 Selección de variables para el modelo multivariado (ajustado)

La selección de variables consiste en decidir cuáles variables incluir en un modelo a partir de una lista completa de predictores disponibles, eliminando aquellas que son irrelevantes o redundantes. El objetivo es construir un modelo que explique adecuadamente el desenlace y permita realizar predicciones precisas sin sobreajustar los datos.

Existen al menos dos enfoques principales para la selección de variables:

1.2.1 Selección automática

Este método emplea algoritmos automáticos —disponibles en R— para determinar qué variables incluir en el modelo. Las técnicas automáticas de selección se basan en criterios estadísticos como los valores p o los coeficientes de regresión. Los algoritmos difieren principalmente en la estrategia que utilizan para evaluar la inclusión o exclusión de variables en el modelo final.

Dependiendo de la dirección del algoritmo (forward, backward o stepwise), el resultado será un subconjunto seleccionado de variables. Para comparar entre distintos modelos generados por estos algoritmos, puede utilizarse el Criterio de Información de Akaike (Akaike Information Criterion, AIC), que estima el error de predicción y, por tanto, la calidad relativa de los modelos estadísticos para un conjunto de datos dado. En términos simples, cuanto menor sea el valor del AIC, mejor es el modelo en términos de equilibrio entre ajuste y complejidad.

Hay al menos tres algoritmos de selección automática de variables:

1. Eliminación hacia atrás (*Backward elimination*),
2. Selección hacia adelante (*Forward selection*) y
3. Selección paso a paso (*Stepwise selection*).

Cada uno de estos métodos tiene ventajas y limitaciones. Entre ellos, la selección paso a paso es una técnica ampliamente utilizada en investigaciones en ciencias de la salud, ya que combina procedimientos de selección hacia adelante y hacia atrás. Esto permite añadir o eliminar variables de manera iterativa en función de criterios estadísticos, optimizando el modelo en ambos sentidos.

Sin embargo, la selección automática de variables no debería realizarse de manera aislada; es recomendable complementarla con una evaluación de la multicolinealidad. La multicolinealidad ocurre cuando dos o más variables independientes están altamente correlacionadas, lo que puede distorsionar las estimaciones del modelo. Por ejemplo, no es apropiado incluir simultáneamente el recuento total de leucocitos y el recuento de neutrófilos, dado que ambas variables están estrechamente relacionadas; en estos casos, es preferible seleccionar solo una de ellas.

En regresión, una herramienta común para detectar multicolinealidad es el Factor de Inflación de la Varianza (VIF, por sus siglas en inglés). De manera general, se interpreta así:

- VIF de 1 indica que no hay multicolinealidad.
- VIF entre 1 y 5 sugiere una multicolinealidad moderada.
- VIF superior a 5 o 10 indica una multicolinealidad alta que puede requerir atención.

1.2.2 Selección intencionada de variables

La selección intencionada de variables sigue una serie de pasos que combinan criterios estadísticos y consideraciones clínicas. Estos pasos incluyen:

- Evaluación univariada de variables: Se realiza un análisis univariado para cada variable independiente con respecto a la variable de desenlace. Las variables que presentan una asociación estadísticamente significativa (habitualmente con un valor de p menor a 0.20) o que son consideradas clínicamente relevantes se seleccionan para su inclusión inicial en el modelo multivariado, independientemente de su significancia estadística.
- Comparación de modelos multivariados: Las variables seleccionadas se incluyen en un modelo multivariado preliminar. A partir de este modelo, las variables que no alcanzan un nivel de significancia estadística estricto (por ejemplo, $p > 0.05$) pueden ser consideradas para eliminación. Posteriormente, se comparan el modelo original (con todas las variables) y el modelo reducido (con las variables eliminadas) para evaluar si la simplificación del modelo afecta negativamente su capacidad explicativa o predictiva. Esta comparación puede realizarse mediante pruebas como la de razón de verosimilitud (Likelihood Ratio Test) o criterios de información (AIC/BIC).
- Evaluación de interacciones: Es importante explorar posibles términos de interacción entre variables que, en combinación, podrían modificar el efecto sobre el desenlace.

2. Ejemplos de análisis univariado y multivariado en una regresión logística

2.1 El dataset para este ejercicio

Para ejemplificar el procedimiento de análisis multivariado mediante un modelo de regresión logística, se utilizará la información proveniente del estudio presentado en "Almac Sangre". Este estudio recopila datos de diversos pacientes, incluyendo variables como recurrencia, estado del tiempo (censor o no censurado), resultado de BN positivo, edad, estadio tumoral, Gleason, terapia previa, y otros factores clínicos y demográficos, que permiten analizar la asociación de distintas características con los eventos de interés.

Cargando los datos

```
{r}  
almac_sangre <- import(here("data", "almac_sangre.csv"))
```

```
Registered S3 method overwritten by 'data.table':  
  method      from  
  print.data.table
```

Un vistazo a los datos

```
{r}
head(almac_sangre)
```

Description: df [6 x 20]

	Grupo_edad_GR <chr>	Edad_mediana... <int>	Edad <dbl>	Raza_afroame... <chr>
1	Mayor	25	72.1	No
2	Mayor	25	73.6	No
3	Mayor	25	67.5	No
4	Intermedio	15	65.8	No
5	Intermedio	15	63.2	No
6	Mayor	25	65.4	No

6 rows | 1-5 of 20 columns

2.2 El análisis univariado

En esta sección se estimarán los Odds Ratios (OR) de cada variable de manera independiente, es decir, sin ajuste por otras covariables.

Antes de realizar este análisis, es necesario definir las categorías de referencia para las variables categóricas mediante la función `mutate()` en combinación con `relevel()`. Este paso asegura que la interpretación de los OR se haga en relación con la categoría de referencia seleccionada. El resultado se guarda en un nuevo objeto llamado `almac_sangre_1`.

```
{r}
almac_sangre_1 <- almacen_sangre |>
  mutate(Terapia_previa = relevel(as.factor(Terapia_previa), ref = "No"),
         Raza_afroamericana = relevel(as.factor(Raza_afroamericana), ref = "Sí"),
         Terapia_adyuvante = relevel(as.factor(Terapia_adyuvante), ref = "No")) |>
  na.omit()
```

```
{r}
tabla_reg_log_univ <- almac_sangre_1 |>
  tbl_uvregression(
    include = c(Unidades_transfundidas, Raza_afroamericana, Terapia_previa, Edad,
Edad_mediana_GR,
                Volumen_prostata, PSA_preoperatorio),
    y = Terapia_adyuvante,
    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(
      Unidades_transfundidas ~ "Unidad transfundida",
      Raza_afroamericana ~ "Raza afroamericana",
      Terapia_previa ~ "Terapia previa",
      Edad ~ "Edad",
      Edad_mediana_GR ~ "Edad mediana GR",
      Volumen_prostata ~ "Volumen prostata",
      PSA_preoperatorio ~ "PSA preoperatorio"
    )
  ) |>
  bold_labels() |>
  bold_p(t = 0.05) |>
  modify_header(estimate = "***OR no ajustado***", p.value = "***Valor p***")
```

{r}
tabla_reg_log_univ

The RStudio interface displays the following R code in the terminal:

```
reglogit <- glm(y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10 + x11 + x12 + x13 + x14 + x15 + x16 + x17 + x18 + x19 + x20 + x21 + x22 + x23 + x24 + x25 + x26 + x27 + x28 + x29 + x30 + x31 + x32 + x33 + x34 + x35 + x36 + x37 + x38 + x39 + x40 + x41 + x42 + x43 + x44 + x45 + x46 + x47 + x48 + x49 + x50 + x51 + x52 + x53 + x54 + x55 + x56 + x57 + x58 + x59 + x60 + x61 + x62 + x63 + x64 + x65 + x66 + x67 + x68 + x69 + x70 + x71 + x72 + x73 + x74 + x75 + x76 + x77 + x78 + x79 + x80 + x81 + x82 + x83 + x84 + x85 + x86 + x87 + x88 + x89 + x90 + x91 + x92 + x93 + x94 + x95 + x96 + x97 + x98 + x99 + x100, data = datos, family = "binomial")
```

The table of results is as follows:

Characteristic	OR no ajustado	95% CI	Valor P
Unidad transfundida	0.68	0.23, 1.20	0.387
Raza afroamericana			
Sí	—	—	
No	1.24	0.21, 23.73	0.843
Terapia previa			
No	—	—	
Sí	9.98	2.11, 52.55	0.003
Edad	0.96	0.87, 1.07	0.447
Edad mediana GR	1.07	0.95, 1.21	0.290

Edad mediana GR	1.07	0.95, 1.21	0.290
Volumen prostata	0.96	0.89, 1.00	0.137
PSA preoperatorio	1.09	1.00, 1.17	0.019

Abbreviations: CI = Confidence Interval, OR = Odds Ratio

¿Cómo interpretar?

1. Unidad transfundida: OR de 0.68 (IC 0.23 a 1.20, $p=0.387$). Esto sugiere que la transfusión no está significativamente asociada con el evento, ya que el valor p es mayor a 0.05 y el IC incluye 1.
2. Raza afroamericana: comparación entre "Sí" y "No" (referencia). La OR de 1.24 (IC 0.21 a 23.73, $p=0.843$) indica que no hay una asociación significativa y que la estimación es muy imprecisa (amplio intervalo de confianza).
3. Terapia previa: OR de 9.98 (IC 2.11 a 52.55, $p=0.003$). Aquí existe una asociación estadísticamente significativa, sugiriendo que los pacientes con terapia previa tienen casi 10 veces más probabilidad de presentar el evento en comparación con los que no tuvieron terapia previa.
4. Edad: OR de 0.96 (IC 0.87 a 1.07, $p=0.447$), no significativa, indicando que la edad por sí sola no se asocia de manera significativa con el evento.
5. Edad mediana de grava renal (GR): OR de 1.07 (IC 0.95 a 1.21, $p=0.290$), no significativa.
6. Volumen de próstata: OR de 0.96 (IC 0.89 a 1.00, $p=0.137$), no significativa
7. PSA preoperatorio: OR de 1.09 (IC 1.00 a 1.17, $p=0.019$), significativa, sugiriendo que a mayor nivel de PSA antes de la cirugía, mayor es la probabilidad del evento.

En resumen, los únicos factores con asociación estadísticamente significativa en este análisis no ajustado son la terapia previa y el PSA preoperatorio, indicando que estos dos variables podrían influir en el evento estudiado en la población analizada.

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 elimination*), selección hacia adelante (*forward selection*) y selección paso a paso (*stepwise selection*).

Paso 1. Ajuste del modelo inicial

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

```
{r}
var_modelo = glm(
  Terapia_adyuvante ~ Unidades_transfundidas + Raza_afroamericana + Terapia_previa
+ Edad + Edad_mediana_GR +
  Volumen_prostata + PSA_preoperatorio,
  data = almac_sangre_1,
  family = binomial(link = "logit")
)
```

Paso 2a. Realizamos la selección de variables usando la técnica Eliminación hacia atrás (Backward elimination).

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

Paso 2b. Realizamos la selección de variables usando la técnica Selección hacia adelante (Forward selection).

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

Paso 3c. Realizamos la selección de variables usando la técnica Selección paso a paso (Stepwise selection).

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

Los resultados de la selección de las variables para el modelo se han guardado en los objetos: multi_backward, multi_forward, y multi_stepwise. El siguiente paso es comparar los valores de AIC y la multicolinealidad entre las variables seleccionadas por cada uno de los modelos.

Paso 3. 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 = Terapia_adyuvante ~ Unidades_transfundidas + Terapia_previa +
PSA_preoperatorio, family = binomial(link = "logit"), data = almac_sangre_1)

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-3.85751	0.99810	-3.865	0.000111	***
Unidades_transfundidas	-0.59940	0.47906	-1.251	0.210858	
Terapia_previaSí	1.99786	0.83828	2.383	0.017159	*
PSA_preoperatorio	0.07155	0.04352	1.644	0.100177	

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 65.818 on 286 degrees of freedom
Residual deviance: 54.112 on 283 degrees of freedom
AIC: 62.112

Number of Fisher Scoring iterations: 7


```
{r}
summary(multi_forward)

Call:
glm(formula = Terapia_adyuvante ~ Unidades_transfundidas + Raza_afroamericana +
    Terapia_previa + Edad + Edad_mediana_GR + Volumen_prostata +
    PSA_preoperatorio, family = binomial(link = "logit"), data = almac_sangre_1)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    -4.096435   3.708754  -1.105   0.2694
Unidades_transfundidas -0.584443   0.520375  -1.123   0.2614
Raza_afroamericanaNo  0.696398   1.211544   0.575   0.5654
Terapia_previaSí      1.642641   0.931635   1.763   0.0779 .
Edad              0.001832   0.052565   0.035   0.9722
Edad_mediana_GR      0.052002   0.067426   0.771   0.4406
Volumen_prostata     -0.030013   0.030572  -0.982   0.3262
PSA_preoperatorio     0.083280   0.048160   1.729   0.0838 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 65.818  on 286  degrees of freedom
Residual deviance: 51.710  on 279  degrees of freedom
AIC: 67.71

Number of Fisher Scoring iterations: 8
```

```
{r}
summary(multi_stepwise)

Call:
glm(formula = Terapia_adyuvante ~ Unidades_transfundidas + Terapia_previa +
    PSA_preoperatorio, family = binomial(link = "logit"), data = almac_sangre_1)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    -3.85751   0.99810  -3.865 0.000111 ***
Unidades_transfundidas -0.59940   0.47906  -1.251 0.210858
Terapia_previaSí      1.99786   0.83828   2.383 0.017159 *
PSA_preoperatorio     0.07155   0.04352   1.644 0.100177
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 65.818  on 286  degrees of freedom
Residual deviance: 54.112  on 283  degrees of freedom
AIC: 62.112

Number of Fisher Scoring iterations: 7
```

2.4 Conclusión

Los modelos obtenidos mediante eliminación hacia atrás (backward elimination) y selección paso a paso (stepwise selection) presentaron el menor valor de AIC (313.455), indicando un mejor ajuste en comparación con el modelo generado mediante selección hacia adelante (forward selection). Además, ambos modelos seleccionaron el mismo conjunto de variables. Por el contrario, la técnica de selección hacia adelante mantuvo todas las variables originales, lo que resultó en un modelo más complejo sin mejorar el AIC.

2.5 Evaluación de colinealidad

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

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

```
# Check for Multicollinearity
```

Low Correlation

	Term	VIF	adj. VIF	Tolerance
	Unidades_transfundidas	1.16	1.08	0.86
	Terapia_previa	1.09	1.05	0.91
	PSA_preoperatorio	1.26	1.12	0.79

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

```
# Check for Multicollinearity
```

Low Correlation

	Term	VIF	adj. VIF	Tolerance
	Unidades_transfundidas	1.32	1.15	0.75
	Raza_afroamericana	1.10	1.05	0.91
	Terapia_previa	1.32	1.15	0.76
	Edad	1.14	1.07	0.88
	Edad_mediana_GR	1.07	1.03	0.93
	Volumen_prostata	1.23	1.11	0.81
	PSA_preoperatorio	1.54	1.24	0.65

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

```
# Check for Multicollinearity
```

Low Correlation

	Term	VIF	adj. VIF	Tolerance
	Unidades_transfundidas	1.16	1.08	0.86
	Terapia_previa	1.09	1.05	0.91
	PSA_preoperatorio	1.26	1.12	0.79

2.6 Conclusión

Las tres metodologías de análisis (multi_backward, multi_forward, y stepwise) convergieron en modelos similares, destacando que la variable "Terapia_previa" es estadísticamente significativa en la predicción de la terapia adyuvante, con un valor p menor a 0.05, indicando que la presencia de terapia previa aumenta la probabilidad de recibir terapia adyuvante.

Por otro lado, las variables "Unidades_transfundidas" y "PSA_preoperatorio", aunque muestran una correlación en los modelos, no alcanzan significancia estadística en la mayoría de las pruebas ($p > 0.05$ en la mayoría) en estos análisis. La variable "Raza_afroamericana" no resulta significativa en el modelo stepwise, evidenciando que su impacto no es claro en este contexto.

El ajuste del modelo, medido por el deviance residual y el AIC, indica que el modelo con "Terapia_previa", junto con las otras variables, tiene un mejor ajuste en comparación con un modelo nulo, particularmente reflejado en la reducción del deviance.

En conclusión, la variable más relevante para predecir la terapia adyuvante en este análisis es la terapia previa, mientras que las demás variables no presentan significancia estadística fuerte en estos modelos. Esto sugiere que la historia previa de terapia tiene un impacto importante en la probabilidad de recibir terapia adyuvante en pacientes con las características estudiadas.

2.7 Modelo final

Con base en los resultados de ajuste (AIC) y la evaluación de colinealidad (VIF), se concluye que el modelo óptimo es el obtenido mediante las técnicas de eliminación hacia atrás (backward elimination) o selección paso a paso (stepwise selection), dado que ambos produjeron exactamente el mismo conjunto de variables.

El modelo final incluye un total de cuatro variables independientes, entre ellas la más resaltante que es la variable "Terapia_previa", las cuales serán reportadas en el análisis multivariado definitivo.

3 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 <- almac_sangre_1 |>
tbl_uvregression(
  include = c(Unidades_transfundidas, Raza_afroamericana, Terapia_previa, Edad,
  Edad_mediana_GR,
    Volumen_prostata, PSA_preoperatorio),
  y = Terapia_adyuvante,
  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(
    Unidades_transfundidas ~ "Unidades Transfundidas",
    Raza_afroamericana ~ "Raza afroamericana",
    Terapia_previa ~ "Terapia previa",
    Edad ~ "Edad",
    Edad_mediana_GR ~ "Edad mediana de GR",
    Volumen_prostata ~ "Volumen prostata",
    PSA_preoperatorio ~ "PSA preoperatorio"
  )
) |>
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(
  Terapia_adyuvante ~ Unidades_transfundidas + Terapia_previa + Volumen_prostata +
  PSA_preoperatorio,
  family = binomial(link = "logit"),
  data = almac_sangre_1
) |>
tbl_regression(
  exponentiate = TRUE,
  conf.int = TRUE,
  pvalue_fun = ~ style_pvalue(.x, digits = 3),
  estimate_fun = ~ style_number(.x, digits = 2),
  label = list(
    Unidades_transfundidas ~ "Unidades transfundidas",
    Terapia_previa ~ "Terapia previa",
    Volumen_prostata ~ "Volumen prostata",
    PSA_preoperatorio ~ "PSA preoperatorio"
  )
) |>
bold_labels() |>
bold_p(t = 0.05) |>
modify_header(estimate = "***OR***", p.value = "***Valor p***")
```

La tabla final la construimos usando la función `tbl_merge()`. De modo que la tabla del análisis univariado o no ajustado y multivariado o ajustado, se muestren lado a lado.

```
{r}
tabla_final <-
tbl_merge(
  list(tabla_univ, tabla_multi),
  tab_spanner = c("***Univariado***", "***Multivariado***")
)
```

{r}
tabla_final

Characteristic	Univariado			Multivariado		
	OR	95% CI	Valor P	OR	95% CI	Valor P
Unidades Transfundidas	0.68	0.23, 1.20	0.387			
Raza afroamericana						
Sí	—	—				
No	1.24	0.21, 23.73	0.843			
Terapia previa						
No	—	—		—	—	
Sí	9.98	2.11, 52.55	0.003	5.04	0.79, 31.83	0.076
Edad	0.96	0.87, 1.07	0.447			

3.1 ¿Cómo interpretar?

En el análisis de regresión logística tanto en el modelo univariado como en el multivariado, se observan diferentes asociaciones entre las variables estudiadas y la probabilidad de recibir terapia adyuvante.

La variable "Terapia previa" muestra una asociación estadísticamente significativa en el análisis univariado, donde los pacientes con terapia previa tienen aproximadamente 10 veces más probabilidades de recibir terapia adyuvante en comparación con aquellos sin terapia previa (OR = 9.98; IC 95%: 2.11–52.55; $p = 0.003$). Sin embargo, en el análisis multivariado, esta asociación disminuye y pierde significancia estadística, presentando un OR de 5.04 (IC 95%: 0.79–31.83; $p = 0.076$), lo que indica que al ajustar por otras variables, la relación no es estadísticamente significativa, aunque aún sugiere una tendencia hacia mayor probabilidad de terapia adyuvante en pacientes con terapia previa.

La variable "Unidades transfundidas" no mostró asociación significativa con la probabilidad de recibir terapia adyuvante en ninguno de los análisis ($p > 0.05$).

Por otro lado, variables como "Raza afroamericana", "Edad", "Edad mediana de GR", "Volumen de próstata" y "PSA preoperatorio" tampoco se asociaron de manera estadísticamente significativa con la probabilidad de recibir terapia adyuvante en los modelos ajustados, aunque en algunos casos como el PSA preoperatorio en el análisis univariado, la relación fue significativa (OR = 1.09; IC 95%: 1.00–1.17; $p = 0.019$), sugiriendo que incrementos en el PSA previo al tratamiento pueden estar relacionados con mayor probabilidad de recibir terapia adyuvante, pero esta asociación no se mantiene en el análisis multivariado.

En resumen, la evidencia más fuerte en estos datos indica que la presencia de terapia previa está relacionada con una mayor probabilidad de recibir terapia adyuvante, aunque esta relación no alcanza significancia estadística en el análisis ajustado, y otras variables como edad, raza, volumen prostático y unidades transfundidas no mostraron asociación significativa tras el ajuste.

The screenshot shows the RStudio interface. The main editor window displays the text '3.1 ¿Cómo interpretar?' followed by a paragraph discussing logistic regression results for 'Terapia previa' and 'Unidades transfundidas'. Below this, there is R code for creating a final table and a console window showing the execution of this code. The console output includes several warnings about fitted probabilities being numerically 0 or 1, which is a common issue in logistic regression with extreme values. The Environment pane on the right shows the objects created in the session, including 'almac_sangre', 'tbl_merge', 'tbl_multi', 'tbl_univ', and 'tbl_multivariado'.

3.1 ¿Cómo interpretar?

En el análisis de regresión logística tanto en el modelo univariado como en el multivariado, se observan diferentes asociaciones entre las variables estudiadas y la probabilidad de recibir terapia adyuvante.

La variable "Terapia previa" muestra una asociación estadísticamente significativa en el análisis univariado, donde los pacientes con terapia previa tienen aproximadamente 10 veces más probabilidades de recibir terapia adyuvante en comparación con aquellos sin terapia previa (OR = 9.98; IC 95%: 2.11–52.55; $p = 0.003$). Sin embargo, en el análisis multivariado, esta asociación disminuye y pierde significancia estadística, presentando un OR de 5.04 (IC 95%: 0.79–31.83; $p = 0.076$), lo que indica que al ajustar por otras variables, la relación no es estadísticamente significativa, aunque aún sugiere una tendencia hacia mayor probabilidad de terapia adyuvante en pacientes con terapia previa.

La variable "Unidades transfundidas" no mostró asociación significativa con la probabilidad de recibir terapia adyuvante en ninguno de los análisis ($p > 0.05$).

Por otro lado, variables como "Raza afroamericana", "Edad", "Edad mediana de GR", "Volumen de

```
Chunk 10:
> tabla_final <-
+   tbl_merge(
+     list(tbl_univ, tbl_multi),
+     tab_spanner = c("Univariado", "Multivariado")
+   )
> tabla_final
>
```

Console output:

```
Aviso: glm.fit: fitted probabilities numerically 0 or 1 occurred
Aviso: glm.fit: fitted probabilities numerically 0 or 1 occurred
Aviso: glm.fit: fitted probabilities numerically 0 or 1 occurred
Aviso: glm.fit: fitted probabilities numerically 0 or 1 occurred
Aviso: glm.fit: fitted probabilities numerically 0 or 1 occurred
Aviso: glm.fit: fitted probabilities numerically 0 or 1 occurred
Aviso: glm.fit: fitted probabilities numerically 0 or 1 occurred
Aviso: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

Environment pane:

Object	Class	Attributes
almac_sangre	data.frame	316 obs. of 20 variables
almac_sangre_1	data.frame	287 obs. of 20 variables
multi_backward	list	List of 31
multi_forward	list	List of 31
multi_stepwise	list	List of 31
tabla_final	tbl_merge	Large tbl_merge (4 elements, 13 MB)
tabla_multi	tbl	List of 6
tabla_reg_log_u	tbl_uvregression	Large tbl_uvregression (5 elements, 10.5 MB)
tabla_univ	tbl_uvregression	Large tbl_uvregression (5 elements, 10.5 MB)
var_modelo	list	List of 30