

```
---
title: "Análisis univariado y multivariado con gtsummary"
author: "GRUPO 10"
format: html
---
```

Grupo 10

- ANNALISA CATERINA GUTIERREZ UTRILLA
- CARLOS RAMIRO HUARCAYA ANTEZANA
- EDWIN ADRIAN PECEROS ARENAS
- MILUSKA SARAI ZAMBRANO MOTTA
- YEZIT KATERIN QUISPE MONROY

Cargar los paquetes

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

```
{r}
install.packages("broom.helpers")
```

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

```
{r}
library(broom.helpers)
```

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.

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 empleará el dataset `tiroide_s`. Este estudio recopila datos de diversos pacientes (383), incluyendo variables como el tipo de lesión (unifocal o multifocal), el tamaño tumoral (T2, T3a), el estado linfático (N0 o N1b), la presencia de metástasis, y características clínicas como antecedentes de fumador, presencia de bocio multinodular, y alteraciones en la función tiroidea (hipertiroidismo, hipotiroidismo, eutiroidea). Estas variables permiten analizar la asociación de diferentes características clínicas y demográficas con la presencia y gravedad de las afecciones tiroideas, así como la probabilidad de recurrencia o avance de la enfermedad.

Cargando los datos

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

Un vistazo a los datos

```
{r}
head(tiroide_s)
```

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

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 `tiroide_s_1`.

```
{r}
tiroide_s_1 <- tiroide_s |>
  mutate(Fumador = relevel(as.factor(Fumador), ref = "No"),
         Genero = relevel(as.factor(Genero), ref = "Femenino"),
         Recurrencia = relevel(as.factor(Recurrencia), ref = "Sí")) |>
  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. Entre sus argumentos se especifican el método de regresión, las variables a analizar, la familia de distribución (binomial para modelos logísticos), y opciones de presentación de los resultados como los intervalos de confianza, valores p y formato de los estimadores.

```
{r}
tabla_reg_log_univ <- tiroide_s_1 |>
tbl_uvregression(
  include = c(Edad, Genero, Fumador, Historia_Fumador, Historia_Radioterapia,
              Focalidad, M),
  y = Recurrencia,
  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(
    Edad ~ "Edad (años)",
    Genero ~ "Sexo del paciente",
    Fumador ~ "Consumo actual de tabaco",
    Historia_Fumador ~ "Historia previa de tabaquismo",
    Historia_Radioterapia ~ "Exposición previa a radioterapia en región cervical"
  ),
  Focalidad ~ "Numero de focos tumorales",
  M ~ "Metastasis"
) |>
bold_labels() |>
bold_p(t = 0.05) |>
modify_header(estimate = "**OR no ajustado**", p.value = "**Valor P**")
```

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

```
{r}
tabla_reg_log_univ
```

¿Cómo interpretar?

La interpretación de estos resultados indica que varias características asociadas con los pacientes están relacionadas con una menor probabilidad de presentar un determinado evento o condición de interés (por ejemplo, una complicación, un tipo de respuesta, o un hallazgo clínico), según los análisis no ajustados.

- **Edad (años):** Por cada año adicional de edad, la odds de presentar el evento disminuyen en un 4% (OR = 0.96; IC 95%: 0.95–0.98; $p < 0.001$). Esto sugiere que la edad avanzada está asociada con menor probabilidad del evento estudiado.
- **Sexo del paciente:** Los pacientes masculinos muestran una odds significativamente menor para el evento en comparación con los femeninos, con una odds de aproximadamente 0.19 (IC 95%: 0.11–0.32; $p < 0.001$), lo que significa que los hombres tienen mucho menos riesgo en comparación con las mujeres, en el contexto analizado.
- **Consumo actual de tabaco:** Los pacientes que consumen tabaco actualmente tienen una probabilidad significativamente menor de presentar el evento (OR = 0.14; IC 95%: 0.07–0.26; $p < 0.001$) en comparación con quienes no lo hacen, lo cual puede parecer contraintuitivo y puede requerir interpretación adicional en su contexto clínico.

- **Historia previa de tabaquismo:** Similarmente, los pacientes con historia previa de tabaquismo presentan una odds menor (OR = 0.36; IC 95%: 0.16–0.79; $p = 0.010$), sugiriendo que tienen menor probabilidad que aquellos sin historia previa.
- **Exposición previa a radioterapia en región cervical:** La exposición previa a radioterapia en esta región también se asocia con una reducción significativa en la probabilidad del evento (OR = 0.06; IC 95%: 0.00–0.37; $p = 0.010$), indicando una relación fuerte en este sentido.
- **Número de focos tumorales:** Los pacientes con un foco tumoral (unifocal) tienen una odds considerablemente mayor de presentar el evento en comparación con los multifocales, con un OR de 5.83 (IC 95%: 3.63–9.53; $p < 0.001$). Esto indica que la presencia de un solo foco tumoral aumenta significativamente la probabilidad en comparación con múltiples focos.
- **Metástasis:** La presencia de metástasis (M1) no mostró una asociación estadísticamente significativa con el evento, dado que el OR es casi cero y el valor de p no es significativo ($p = 0.975$).

En resumen, factores como la menor edad, ser masculino, consumir tabaco actualmente o tener historia previa de tabaquismo, y haber sido expuesto previamente a radioterapia en la región cervical se asocian con una menor probabilidad del evento estudiado. Por otro lado, presentar un foco tumoral único aumenta considerablemente esa probabilidad. Sin embargo, algunas de estas asociaciones pueden requerir una interpretación contextual adicional, especialmente en aspectos que parecen contradecir expectativas clínicas o epidemiológicas.

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(
  Recurrencia ~ Edad + Genero + Fumador + Historia_Fumador + Historia_Radioterapia
+   Focalidad + M,
  data = tiroide_s_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 = Recurrencia ~ Edad + Genero + Focalidad + M, family = binomial(link = "logit"),
    data = tiroide_s_1)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    1.432456   0.443720   3.228  0.00125 **
Edad           -0.020813   0.009019  -2.308  0.02102 *
GeneroMasculino -1.282409   0.319061  -4.019 5.84e-05 ***
FocalidadUnifocal 1.400263   0.270529   5.176 2.27e-07 ***
MM1            -17.692179 843.753759  -0.021  0.98327
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 455.63  on 382  degrees of freedom
Residual deviance: 343.75  on 378  degrees of freedom
AIC: 353.75

Number of Fisher Scoring iterations: 16
```

```
{r}
summary(multi_forward)

Call:
glm(formula = Recurrencia ~ Edad + Genero + Fumador + Historia_Fumador +
    Historia_Radioterapia + Focalidad + M, family = binomial(link = "logit"),
    data = tiroide_s_1)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    1.374123   0.448538   3.064  0.00219 **
Edad           -0.018576   0.009278  -2.002  0.04528 *
GeneroMasculino -1.075307   0.381642  -2.818  0.00484 **
FumadorSí      -0.359778   0.485295  -0.741  0.45848
Historia_FumadorSí -0.649634   0.553367  -1.174  0.24041
Historia_RadioterapiaSí 0.082811   2.066366   0.040  0.96803
FocalidadUnifocal 1.418002   0.275103   5.154 2.54e-07 ***
MM1            -17.630603 837.826721  -0.021  0.98321
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 455.63  on 382  degrees of freedom
Residual deviance: 341.38  on 375  degrees of freedom
AIC: 357.38

Number of Fisher Scoring iterations: 16
```



```
{r}
summary(multi_stepwise)

Call:
glm(formula = Recurrencia ~ Edad + Genero + Focalidad + M, family = binomial(link = "logit"),
    data = tiroide_s_1)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    1.432456   0.443720   3.228  0.00125 **
Edad           -0.020813   0.009019  -2.308  0.02102 *
GeneroMasculino -1.282409   0.319061  -4.019  5.84e-05 ***
FocalidadUnifocal 1.400263   0.270529   5.176  2.27e-07 ***
MM1            -17.692179  843.753759  -0.021  0.98327
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 455.63  on 382  degrees of freedom
Residual deviance: 343.75  on 378  degrees of freedom
AIC: 353.75

Number of Fisher Scoring iterations: 16
```

2.4 Conclusión

La conclusión general de los modelos de regresión logística ajustados es que ciertas variables clínicas y demográficas están significativamente asociadas con la recurrencia de la enfermedad tiroidea. En particular, se observa que la edad y el género masculino tienen una relación estadísticamente significativa con la recurrencia, donde mayor edad se asocia con una menor probabilidad (coeficiente negativo), y ser masculino aumenta la probabilidad de recurrencia. Además, la focalidad unifocal también muestra una asociación fuerte y significativa, indicando que la presencia de lesiones unifocales está relacionada con mayor riesgo de recurrencia. La presencia de la variable "MM1" no muestra un efecto significativo en ninguno de los modelos. En conjunto, estos hallazgos sugieren que la edad, el género masculino y la focalidad son factores relevantes para predecir la recurrencia en pacientes con patologías tiroideas, permitiendo enfocar mayores esfuerzos en aquellos con estas características para un manejo más dirigido.

El valor del AIC (Criterio de Información de Akaike) en los diferentes modelos oscila entre aproximadamente 353.75 y 357.38. La comparación entre estos valores indica que el modelo que incluye las variables Edad, Género, Focalidad y M (modelo de paso hacia atrás y hacia adelante) tiene un AIC ligeramente menor (353.75 en el modelo de paso y de retroceso), lo que sugiere que es el más adecuado en términos de equilibrio entre ajuste y parsimonia. Esto refuerza la conclusión de que estas variables son las más relevantes para predecir la recurrencia, dado que este modelo minimiza el AIC en comparación con otros modelos ajustados. Por lo tanto, el AIC apoya la importancia de las variables seleccionadas en la modelación de la recurrencia, indicando que estos factores tienen una mayor contribución explicativa en la predicción del evento.

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
Edad	1.01	1.01	0.99
Genero	1.00	1.00	1.00
Focalidad	1.01	1.00	0.99
M	1.00	1.00	1.00

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

# Check for Multicollinearity

Low Correlation
```

Term	VIF	adj. VIF	Tolerance
Edad	1.06	1.03	0.95
Genero	1.39	1.18	0.72
Fumador	1.51	1.23	0.66
Historia_Fumador	1.13	1.06	0.88
Historia_Radioterapia	1.06	1.03	0.94
Focalidad	1.03	1.02	0.97
M	1.00	1.00	1.00

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

# Check for Multicollinearity

Low Correlation
```

Term	VIF	adj. VIF	Tolerance
Edad	1.01	1.01	0.99
Genero	1.00	1.00	1.00
Focalidad	1.01	1.00	0.99
M	1.00	1.00	1.00

2.6 Conclusión

Los resultados de las verificaciones de multicolinealidad muestran que los modelos generados mediante eliminación hacia atrás (backward elimination), selección paso a paso (stepwise selection) y selección hacia adelante (forward selection) presentaron valores de VIF bajos y cercanos a 1. Específicamente, en los modelos de backward y stepwise, los VIF oscilan alrededor de 1.00 a 1.06, indicando una baja colinealidad entre las variables incluidas. En contraste, el modelo de forward selection mostró un VIF ligeramente mayor para algunas variables, como la "Genero" y "Fumador", aunque todavía dentro de límites aceptables, sin indicios claros de colinealidad severa. Estos hallazgos sugieren que los diferentes procedimientos de selección no han generado problemas importantes de multicolinealidad en los modelos finales, por lo que las variables seleccionadas aportan información de manera relativamente independiente y estable para la predicción del evento de interés.

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, 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 <- tiroide_s_1 |>
tbl_uvregression(
  include = c(Edad, Genero, Fumador, Historia_Fumador, Historia_Radioterapia,
             Focalidad, M),
  y = Recurrencia,
  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(
    Edad ~ "Edad (años)",
    Genero ~ "Sexo del paciente",
    Fumador ~ "Consumo actual de tabaco",
    Historia_Fumador ~ "Historia previa de tabaquismo",
    Historia_Radioterapia ~ "Exposición previa a radioterapia en región cervical",
    Focalidad ~ "Numero de focos tumorales",
    M ~ "Metastasis"
  )
) |>
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(
  Recurrencia ~ Edad + Fumador + Focalidad + M,
  family = binomial(link = "logit"),
  data = tiroide_s_1
) |>
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)",
    Fumador ~ "Consumo actual de tabaco",
    Focalidad ~ "Numero de focos tumorales",
    M ~ "Metastasis"
  )
) |>
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
Edad (años)	0.96	0.95, 0.98	<0.001	0.98	0.96, 1.00	0.052
Sexo del paciente						
Femenino	—	—				
Masculino	0.19	0.11, 0.32	<0.001			
Consumo actual de tabaco						
No	—	—		—	—	
Sí	0.14	0.07, 0.27	<0.001	0.30	0.14, 0.65	0.003

3.1 ¿Cómo interpretar?

1. Edad:

Univariado: Cada año de edad se asocia con una reducción en la probabilidad del evento (OR=0.96, IC 95% 0.95–0.98, $p<0.001$). Esto indica que personas mayores tienen menor riesgo en este análisis simple.

Multivariado: La asociación se mantiene pero con menor intensidad (OR=0.98, IC 95% 0.96–1.00), y el valor p se acerca a 0.05 ($p=0.052$), sugiriendo una tendencia a ser significativa, pero no concluyente al ajustar por otras variables.

2. **Sexo del paciente:** Los pacientes masculinos tienen significativamente menor riesgo en comparación con los femeninos (OR=0.19, IC 0.11–0.32, $p<0.001$). Esto indica que ser masculino se asocia con un riesgo mucho menor del evento en el análisis univariado.

3. **Consumo actual de tabaco:** En análisis univariado, el consumo continúa siendo un factor de riesgo importante (OR=0.14, $p<0.001$), pero en multivariado, el OR aumenta a 0.30 y el valor p es 0.003, manteniendo su significancia, aunque con menor efecto relativo. Sin embargo, un OR menor que 1 en esta categoría indica que el "no" al tabaco sería protector, pero en la forma en que está presentado, parece indicar que fumar sería un factor de riesgo.

4. **Historia previa de tabaquismo:** La presencia de historia previa también se asocia con menor riesgo (OR=0.36, $p=0.010$), lo que puede interpretarse como que los pacientes sin antecedentes de tabaquismo tienen mayor riesgo, o que la historia previa de tabaquismo se asocia con menor riesgo según este análisis.

5. **Exposición previa a radioterapia en región cervical:** La exposición previa a radioterapia se asocia con un riesgo muy bajo del evento (OR=0.06, $p=0.010$). Esto sugiere que la radioterapia previa podría proteger contra el outcome considerado, aunque la magnitud del OR (muy cercano a cero) requiere cautela, ya que también podría reflejar algún sesgo o correlación con otros factores.

6. **Número de focos tumorales:** La presencia de múltiples focos se asocia con un aumento significativo en el riesgo (OR=5.83 en univariado, OR=4.12 en multivariado, ambos $p<0.001$), demostrando que los tumores multifocales tienen un impacto sustancial en el outcome negativo.

7. **Metástasis (M):** En análisis univariado, tener metástasis M1 parece estar asociado con un riesgo muy bajo del evento (OR=0.00, $p<0.001$), pero en multivariado, el OR se acerca a 1 (0.975), y no es estadísticamente significativo ($p=0.983$), indicando que, tras ajustar por otros factores, las metástasis no muestran una asociación clara con el outcome.

Los factores significativamente asociados en el análisis multivariado incluyen la edad (con tendencia a menor riesgo en mayores), sexo masculino, no consumo de tabaco, ausencia de historia previa de tabaquismo, exposición previa a radioterapia cervical y la presencia de foco único (unifocal). La presencia de múltiples focos tumorales se mantiene como un predictor fuerte de mayor riesgo, mientras que otros factores como metástasis no mantienen su asociación tras el ajuste multivariado. Estos resultados ayudan a identificar variables independientes que podrían influir en el pronóstico del paciente [OR, IC 95%, Valor P].

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function

Addins

Source

Visual

B I <>

Normal

Format

Insert

Table

Pe3.Actividad2.Grupo 10.qmd

tiroide_s_1

tiroide_s

Render on Save

Render

Run

Publish

Outline

title: "Análisis univariado y multivariado con gtsummary"

author: "GRUPO 10"

format: html

Grupo 10

- ANNALISA CATERINA GUTIERREZ UTRILLA
- CARLOS RAMIRO HUARCAYA ANTEZANA
- EDWIN ADRIAN PECEROS ARENAS
- MILUSKA SARAI ZAMBRANO MOTTA
- YEZIT KATERIN QUISPE MONROY

Cargar los paquetes

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

```
{r}install.packages("broom.helpers")
```

```
{r}
```

Grupo 10

Cargar los paquetes

1 Modelos univariados (n...

1.1 Interpretación genera...

1.2 Selección de variable...

1.2.1 Selección automá...

1.2.2 Selección intencio...

2 Ejemplos de análisis u...

2.1 El dataset para este...

2.2 El análisis univariado

2.3 El análisis multivari...

2.4 Conclusión

2.5 Evaluación de colin...

2.6 Conclusión

2.7 Modelo final

3 Reporte del análisis un...

3.1 ¿Cómo interpretar?

3.1 ¿Cómo interpretar?

Quarto

Console

Terminal

R 4.4.3 · D:/UPS/IB_Practica_RStudio/estadistica_ups/ib/scripts/

rredAviso: glm.fit: fitted probabilities numerically 0 or 1 occurredAviso: glm.fit: fitted probabilities numerically 0 or 1 occurredAviso: glm.fit: fitted probabilities n

umerically 0 or 1 occurredAviso: glm.fit: fitted probabilities numerically 0 or 1 occurredAviso: glm.fit: fitted prob

abilities numerically 0 or 1 occurredAviso: glm.fit: fitted probabilities numerically 0 or 1 occurredAviso: glm.fit:

fitted probabilities numerically 0 or 1 occurredAviso: glm.fit: fitted probabilities numerically 0 or 1 occurredAvis

o: glm.fit: fitted probabilities numerically 0 or 1 occurred

> tabla_final <-

+ tbl_merge(

+ list(tabla_univ, tabla_multi),

+ tab_spanner = c("**Univariado**", "**Multivariado**")

+)

> tabla_final

> view(tiroide_s_1)

>

Environment

History

Connections

Tutorial

R

Global Environment

Data

multi_backward

List of 31

multi_forward

List of 31

multi_stepwise

List of 31

tabla_final

Large tbl_merge (4 elements, 13.5 MB)

tabla_multi

List of 6

tabla_reg_log_univ

Large tbl_uvregression (5 elements, 10.9 MB)

tabla_univ

Large tbl_uvregression (5 elements, 10.9 MB)

tiroide_s

383 obs. of 17 variables

tiroide_s_1

383 obs. of 17 variables

var_modelo

List of 30

Files

Plots

Packages

Help

Viewer

Presentation

Install

Update

bro

Name

Description

Version

checkbox

broom

Convert Statistical Objects into Tidy Tibbles

1.0.8

checkbox

broom.helpers

Helpers for Model Coefficients Tibbles

1.21.0

checkbox