Pontificia Universidad JAVERIANA Bogotá

MAESTRÍA EN EPIDEMIOLOGÍA CLÍNICA

BIOESTADÍSTICA AVANZADA

MÓDULO II

Semana 8

Regresión Logística

Material elaborado por:

Nelcy Rodriguez Malagón
Bioestadística, M.P.H
Profesora Titular
Departamento de Epidemiología Clínica y Bioestadística
Facultad de Medicina - Pontificia Universidad Javeriana

Kevin Maldonado-Cañón
Médico Epidemiólogo, MD. MSc.
Estudiante - Doctorado en Epidemiología Clínica
Departamento de Epidemiología Clínica y Bioestadística
Facultad de Medicina - Pontificia Universidad Javeriana

Evaluación de supuestos del modelo

Cuando se consideran los supuestos sobre un modelo, se tienen dos momentos: El **primer momento** tiene que ver con las condiciones necesarias para poder utilizar el tipo de modelo que se desea ajustar. El **segundo momento**, ocurre cuando ya se ha hecho el ajuste, es decir, cuando hemos utilizado los datos disponibles aplicando dicho modelo, para estimar los parámetros desconocidos α y β_i.
Una vez ajustado el modelo, la interpretación de los coeficientes no es el paso final del procesamiento y análisis de datos. Antes de pensar en conclusiones, es fundamental evaluar otros supuestos tales como si el modelo está correctamente especificado y si sus estimaciones son estables y suficientemente precisas. Este proceso de diagnóstico garantiza, entre otras cosas, que si se estiman medidas de asociación, como los riesgos relativos indirectos (ORs), reflejen asociaciones o relaciones precisas y no estimaciones derivadas de problemas estructurales del modelo.

Para ilustrar la forma de validación de los supuestos del modelo de regresión logística, consideraremos los dos momentos de validación mencionados previamente.

Evaluación de supuestos antes de ajuste del modelo

En este primer momento, usaremos la base de datos `icu` y consideraremos el problema de la potencial relación o asociación entre las variables `sta` y `ser`, ya mencionados en la semana anterior, a fin de evaluar los supuestos correspondientes.

Supuesto 1: La variable dependiente debe ser binaria

- La **variable dependiente**, en este caso, `sta`, es de tipo cualitativo y dicótomo, con dos posibles valores: vivo (*Lived*) o muerto (*Died*).

```
install.packages("arsenal")
install.packages("dplyr")

## Recuerden siempre al inicio de cada ejercicio cargar la base de datos
base_icu <- aplore3::icu


base_icu %>%
  dplyr::select(sta) %>%
  arsenal::tableby( ~ ., data = ., digits = 1, test = TRUE) %>%
  summary()
```

	Overall (N=200)
sta	
Lived	160 (80.0%)
Died	40 (20.0%)

Cuando vamos a ajustar un modelo de regresión logística en R, debemos definir una variable desenlace (en este caso `sta`) y verificar sus categorías. Por defecto, R modelará la probabilidad del segundo nivel del factor como el evento. Al aplicar la función `levels`, vemos que los niveles están ordenados así: "Lived" "Died". Por lo tanto, R tomará "Died" como el desenlace.

También podemos definir explícitamente la categoría de referencia usando la función `relevel()`.

```
levels(base_icu$sta)
```

[1] "Lived" "Died"

```
base_icu$sta <- relevel(base_icu$sta, ref = "Lived")
```

Supuesto 2: Supuesto de independencia

- Al revisar la base de datos, se observa que la información de cada individuo es independiente de la observación de cualquier otro. Por tanto, vemos que se cumple el supuesto de independencia en las observaciones.

Supuesto 3: Muestra suficiente

Con relación al supuesto de muestra suficiente y aunque este es un punto central en el caso de modelos de predicción, si se supone un **número de eventos por variable en el modelo de 10**, una frecuencia esperada del evento de 20% y 4 variables independientes en el modelo, el tamaño de muestra de 200, resultaría suficiente para explorar la asociación de interés.

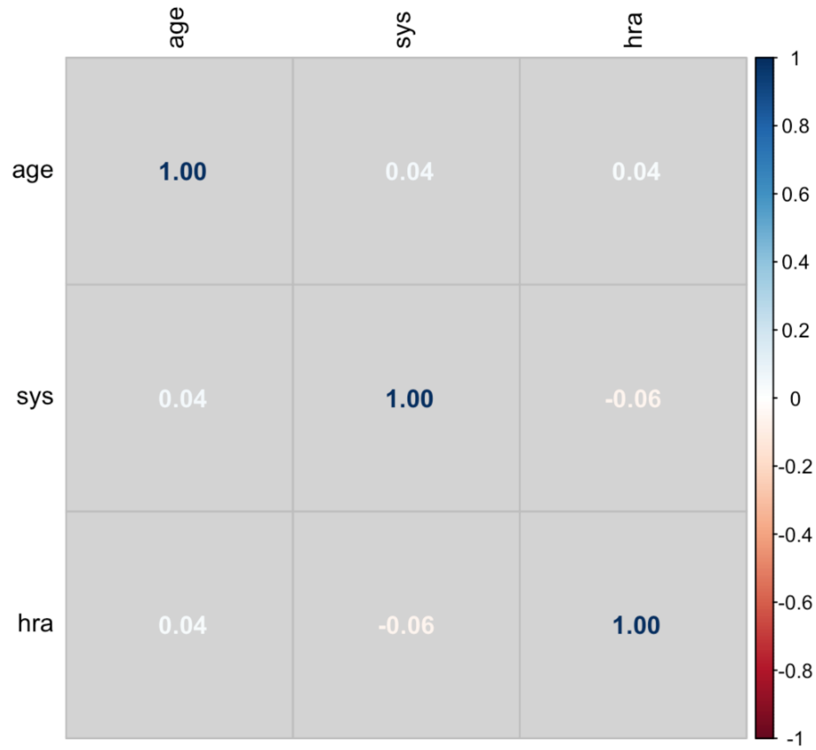
Supuesto 4: Multicolinealidad

Es importante evaluar si existen correlaciones altas entre variables independientes, ya que la colinealidad puede inflar los errores estándar, volver inestables los coeficientes y dificultar la interpretación del modelo. Una forma inicial y sencilla de explorar este problema es mediante la matriz de correlación, particularmente para variables continuas.

En nuestra base ICU, las variables continuas son:

- `age` (edad)
- `sys` (presión sistólica)
- `hra` (frecuencia cardíaca)

```
install.packages("corrplot")
# Seleccionamos las variables continuas
vars_cont <- base_icu[, c("age", "sys", "hra")]

# Calculamos la matriz de correlación de Pearson
cor_matrix <- cor(vars_cont, use = "complete.obs")

corrplot::corrplot(cor_matrix, method = "number", bg = "lightgrey", tl.col = "black")
```



La matriz de correlación muestra coeficientes entre -1 y 1:

- r ≈ 0 → no hay correlación lineal
- r = 0.7 - 0.79 → posible colinealidad relevante
- r ≥ 0.8 → colinealidad alta, potencialmente problemática

En nuestro ejemplo, todos los valores están por debajo de 0.7 por lo que podemos afirmar que no tenemos un problema de colinealidad importante entre nuestras variables independientes continuas.

Evaluación de supuestos después de ajustar el modelo

En este segundo momento, planteamos un modelo de regresión logística utilizando tanto **variables cuantitativas continuas** (edad, presión sistólica) como **variables cualitativas nominales** (servicio y sexo); variables que se pueden considerar como potenciales factores de confusión en la relación que se busca estudiar. Por esta razón, se debería controlar o ajustar en el modelo, por estas variables.

```
## Asignamos "Surgical" como nuestra categoría de referencia
base_icu$ser <- relevel(base_icu$ser, ref = "Surgical")

## Ajustamos el modelo
modelo <- glm(
  sta ~ ser + age + sys + gender,
  data = base_icu,
  family = binomial
)
summary(modelo)
##
## Call:
## glm(formula = sta ~ ser + age + sys + gender, family = binomial,
##     data = base_icu)
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.804363   1.090162  -1.655  0.09790 .
## serMedical    0.900696   0.387143   2.327  0.01999 *
## age           0.029912   0.011266   2.655  0.00793 **
## sys          -0.014791   0.005836  -2.535  0.01126 *
## genderFemale  0.030296   0.385757   0.079  0.93740
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 200.16  on 199  degrees of freedom
## Residual deviance: 177.64  on 195  degrees of freedom
## AIC: 187.64
##
## Number of Fisher Scoring iterations: 5
```

Supuesto 4: Multicolinealidad (después del ajuste)

Otra forma de evaluar el supuesto de **multicolinealidad**, esta vez después del ajuste, es a través de los VIFs (*Variance Inflation Factors*). Los VIFs cuantifican cuánto se incrementa la varianza de un coeficiente estimado debido a la colinealidad con otras variables del modelo.

En otras palabras, los VIFs responden a la pregunta:

- ¿Qué tan bien puede predecirse esta variable usando las otras variables del modelo?

Matemáticamente, para una variable X_j:

$$VIF_j = \frac{1}{1 - R_j^2}$$

donde R_j^2 es el coeficiente de determinación obtenido al regresar X_j en función del resto de variables independientes.

El R_j^2 proviene de una regresión auxiliar en la que esa variable se usa como desenlace y las demás como predictores.

Para calcular los VIFs en R utilizaremos la función `vif` del paquete `car`:

```
install.packages("car")
car::vif(modelo)
##       ser      age      sys   gender
## 1.022102 1.016713 1.014869 1.009487
```

El resultado se interpreta de la siguiente manera:

- VIF ≤ 5: ausencia de colinealidad
- VIF > 5: sugiere colinealidad
- VIF > 10: indica colinealidad severa

En nuestro caso, tenemos evidencia suficiente para decir que no tenemos colinealidad entre nuestras variables independientes.

VIFs al considerar variables categóricas

En nuestro modelo:

- `ser` tiene 2 niveles
- `gender` tiene 2 niveles

```
levels(base_icu$ser)
## [1] "Surgical" "Medical"
levels(base_icu$gender)
## [1] "Male"   "Female"
```

R las convierte internamente en una sola variable dummy. La categoría de referencia toma el valor 0 y la categoría comparada toma el valor 1. Cuando una variable categórica se codifica como factor, la categoría de referencia corresponde por defecto al primer nivel del factor. En este caso, *Surgical* es la referencia para la variable `ser`, y *Male* es la referencia para `gender`.

El valor 1 no implica mayor riesgo; simplemente indica el grupo que se compara contra la referencia:

- `serMedical` (toma los valores: `Surgical` = 0, `Medical` = 1)
- `genderFemale` (toma los valores: `Male` = 0, `Female` = 1)

Como cada factor genera una sola columna, el VIF se calcula exactamente igual que para una variable continua.

Cuando una variable tiene más de 2 categorías R crearía 3 variables dummy, entonces ya no es una sola columna, sino un conjunto de columnas que representan esa variable.

En ese caso no se puede calcular un VIF clásico para cada dummy por separado. Como solución, R calcula algo llamado GVIF (*Generalized VIF*) y luego lo ajusta para que sea comparable con el VIF tradicional.

Supuesto 5: Linealidad del logit

La regresión logística no requiere normalidad, pero sí que las **variables cuantitativas continuas** tengan relación lineal con el logaritmo (logit) de la probabilidad de ocurrencia del desenlace o variable dependiente.

El supuesto de una relación lineal entre las variables independientes continuas (edad: *age* y presión sistólica: *sys*) y el logit de la probabilidad de ocurrencia de nuestro desenlace (*sta*), muestra los siguientes resultados:

Método 1: Test de Box-Tidwell

El Test de Box–Tidwell evalúa el supuesto de linealidad en el logit de la probabilidad de ocurrencia del desenlace en modelos de regresión logística.

Bajo este modelo, se asume que las variables continuas tienen una relación lineal con el logaritmo de las odds (logit) del desenlace:

$$log\left(\frac{p}{1-p}\right)$$

El test consiste en añadir al modelo un término de interacción entre la variable continua X y su logaritmo:

$$X \times log(X)$$

Si este término es estadísticamente significativo, se concluye que la relación entre la variable y el logit no es lineal. Este test es **relevante únicamente para variables continuas**, ya que las variables dicotómicas no requieren evaluación de linealidad en el logit.

```
base_icu$age_log <- base_icu$age * log(base_icu$age)
base_icu$sys_log <- base_icu$sys * log(base_icu$sys)

modelo_bt <- glm(
  sta ~ age + age_log + sys + sys_log,
  data = base_icu,
  family = binomial
)

summary(modelo_bt)
```

```
##
## Call:
## glm(formula = sta ~ age + age_log + sys + sys_log, family = binomial,
##     data = base_icu)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  6.902513   4.399315   1.569   0.1166
## age         -0.017164   0.244198  -0.070   0.9440
## age_log      0.008685   0.049300   0.176   0.8602
## sys         -0.369884   0.150424  -2.459   0.0139 *
## sys_log      0.060597   0.025566   2.370   0.0178 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 200.16  on 199  degrees of freedom
## Residual deviance: 177.71  on 195  degrees of freedom
## AIC: 187.71
##
## Number of Fisher Scoring iterations: 5
```

El resultado se interpreta de la siguiente manera:

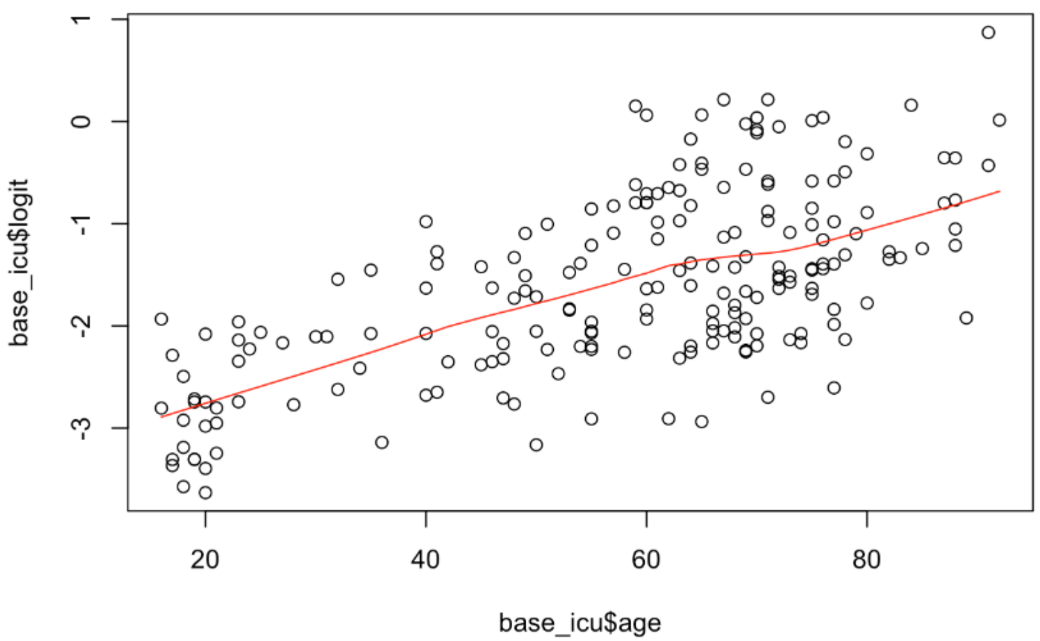
- El término age_log NO es significativo (p = 0.8602)
    - No hay evidencia de violación del supuesto de linealidad en el logit para la edad
    - La relación entre edad y logit de mortalidad puede considerarse aproximadamente lineal

- El término sys_log sí es significativo (p = 0.0178).
    - El supuesto de linealidad en el logit se viola para sys
    - El efecto de la presión sistólica sobre la probabilidad de morir no es lineal
    - Puede existir curvatura (por ejemplo efecto más fuerte en valores extremos)

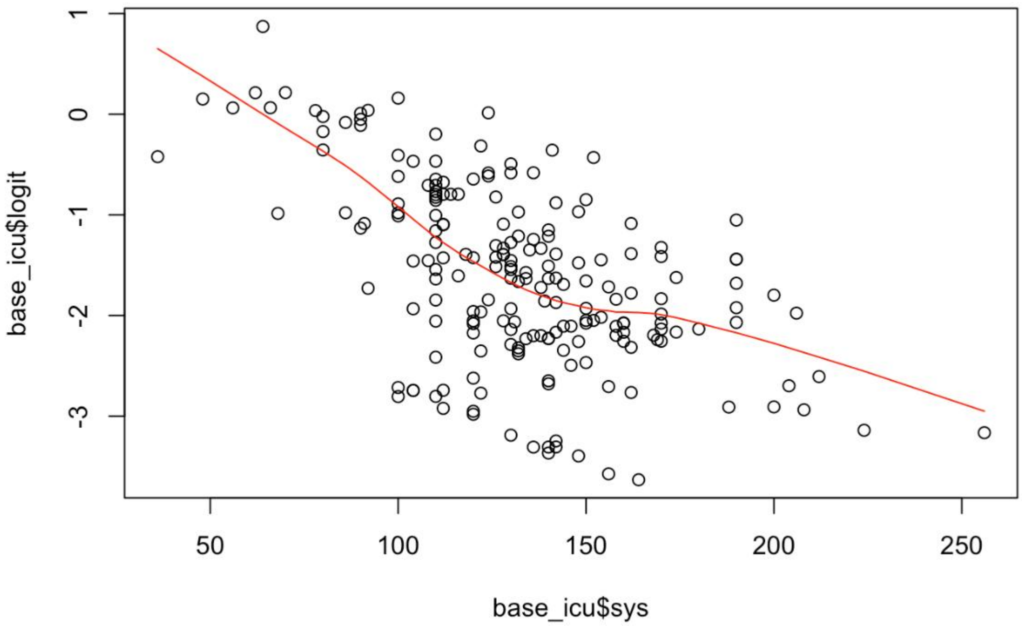
Método 2 (visual): gráfico del logit

Ahora evaluaremos el supuesto de linealidad en el logit de forma gráfica, que es complementaria al test de Box-Tidwell.

```
base_icu$logit <- log(fitted(modelo) / (1 - fitted(modelo)))


plot(base_icu$age, base_icu$logit)

lines(lowess(base_icu$age, base_icu$logit), col="red")
```




```
plot(base_icu$age, base_icu$logit)

lines(lowess(base_icu$age, base_icu$logit), col="red")
```



En ambos gráficos:

- Eje X → variable continua (`age` o `sys`)
- Eje Y → logit estimado (`base_icu$logit`)

Lo que buscamos es responder la pregunta:

- ¿La nube de puntos sigue aproximadamente una tendencia lineal?

Edad

- Se aprecia una tendencia claramente creciente
- No se observa curvatura marcada
- La relación parece aproximadamente recta
- Puede haber dispersión, pero no patrón en U o S
    - Patrón en U: ocurre cuando al aumentar una variable, la otra primero disminuye y luego vuelve a aumentar (o al revés). Es decir, los puntos en la gráfica forman una curva en forma de “U”. Esto indica una relación no lineal, donde los valores extremos se parecen entre sí y difieren de los valores intermedios.
    - Patrón en S: aparece cuando la relación cambia de manera gradual al inicio, luego más rápidamente en el centro y finalmente vuelve a estabilizarse, formando una curva tipo “S” (sigmoide). También representa una relación no lineal estructurada.

Gráficamente, la edad parece cumplir el supuesto de linealidad en el logit. Esto coincide con el resultado del Box-Tidwell (p = 0.86 → no violación).

Presión sistólica

- Se aprecia una tendencia general negativa
- Pero hay indicios de curvatura
- En valores extremos la pendiente parece cambiar
- La nube no sigue una recta tan clara

Aquí sí parece haber una ligera no linealidad. Esto es consistente con el Box-Tidwell significativo (p = 0.0178).

¿Qué podríamos hacer con la Presión Sistólica

Tenemos varias opciones:

- Transformarla (log, polinómica)
- Usar splines (usualmente recomendado)
- Categorizarla (menos ideal estadísticamente)
- Modelar con términos polinomiales (ej. sys^2)

Para efectos del ejemplo con el que estamos trabajando, no haremos ninguna modificación

Supuesto 6: Observaciones influyentes

La Distancia de Cook es una medida de influencia que evalúa cuánto cambian los coeficientes del modelo si se elimina una observación específica.

Combina dos elementos:

- Leverage (influencia estructural del punto)
- Residuales (magnitud del error)

Un valor alto de distancia de Cook indica que esa observación tiene un impacto considerable sobre el modelo.

- $Di > \frac{4}{n}$ puede considerarse potencialmente influyente.
- Valores cercanos o mayores a 1 suelen indicar observaciones altamente influyentes.

Papel en una variable dicotómica (desenlace binario)

En regresión logística con un desenlace dicótomo:

- La distancia de Cook identifica observaciones cuya combinación de variables independientes y resultado (0/1) altera significativamente los coeficientes estimados.
- Puede detectar sujetos con combinaciones poco frecuentes (por ejemplo, patrón extremo de covariables con un evento inesperado).
- Es especialmente relevante cuando existe desbalance en la variable dependiente.

Es importante aclarar que:

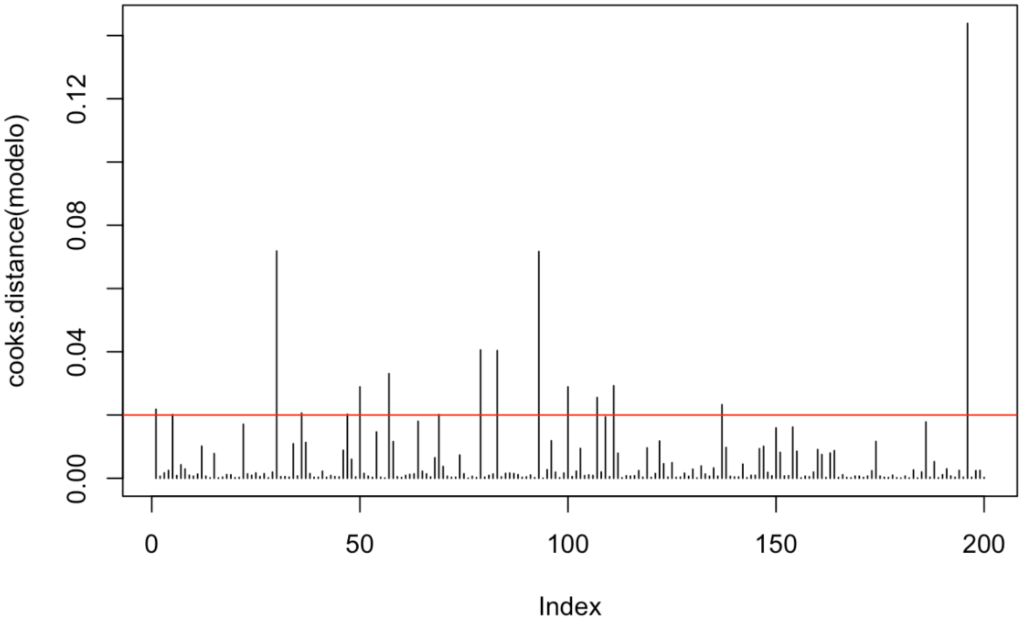
- Una observación influyente no necesariamente es errónea.
- Debe evaluarse clínicamente antes de decidir excluirla.
- Eliminar observaciones sin justificación puede introducir sesgo.

El resultado se interpreta de la siguiente manera:

- Valores por encima de la línea roja pueden ser influyentes $Di > \frac{4}{n}$.

La gráfica de la distancia de Cook muestra la influencia individual de cada observación sobre los coeficientes del modelo de regresión logística. En el eje horizontal se presenta el índice de las observaciones y en el eje vertical el valor de la distancia de Cook.

```
plot(cooks.distance(modelo), type="h")
abline(h = 4/nrow(base_icu), col="red")
```



La mayoría de los puntos se concentran cerca de cero, lo que indica que la mayor parte de los casos tienen un impacto mínimo en la estimación del modelo. Sin embargo, se observan algunas observaciones que superan la línea de referencia (aproximadamente 4/n), lo que sugiere influencia potencial moderada.

¿Qué hacer con las observaciones influyentes?

Destaca particularmente una observación con un valor considerablemente mayor que el resto, lo que indica que podría estar ejerciendo una influencia relativamente alta sobre los coeficientes estimados. No obstante, dado que ninguno de los valores alcanza magnitudes extremas (por ejemplo, cercanas a 1), no se evidencia una influencia desproporcionada severa, aunque sí se recomienda revisar de manera específica los casos que superan el umbral para evaluar su plausibilidad clínica y su impacto en la estabilidad del modelo.

1. Identificar las observaciones

```
unname(which(cooks.distance(modelo) > 4/nrow(base_icu)))
```

```
##  [1]   1   5  30  36  47  50  57  69  79  83  93 100 107 111 137 196
```

2. Revisar clínicamente esas observaciones

Preguntarse:

- ¿Son valores plausibles?
- ¿Son errores de digitación?
- ¿Son casos extremos reales (ej. paciente muy joven con PAS muy baja que murió)?

3. Hacer un análisis de sensibilidad

Podemos comparar el modelo completo vs. modelo sin esas observaciones.

Y según los resultados podemos evaluar el impacto en la estabilidad del modelo:

- Los coeficientes cambian mucho → modelo inestable
- Cambian poco → impacto limitado

Prueba de bondad y ajuste Hosmer Lemeshow

Esta evaluación está especialmente orientada a evaluar un modelo que busca predecir la probabilidad de ocurrencia de un evento. No solo queremos saber si los coeficientes son significativos, sino si el modelo está bien calibrado, es decir si las probabilidades predichas se corresponden adecuadamente con lo observado en los datos.

La prueba de Hosmer–Lemeshow es un método clásico para evaluar la bondad de ajuste en regresión logística, comparando frecuencias observadas y esperadas del evento en grupos de riesgo.

La prueba:

- Ordena las observaciones según la probabilidad predicha
- Divide la muestra en g grupos (por defecto R usa 10 grupos)
- Compara el número observado y esperado de eventos en cada grupo
- Calcula un estadístico Chi-cuadrado

La hipótesis que estamos probando con esta prueba es:

- H_0: El modelo se ajusta adecuadamente (no hay diferencias entre observado y esperado).
- H_1: El modelo no se ajusta bien.

Un p > 0.05 indica que no hay evidencia para rechazar buen ajuste.

En R, lo podemos hacer usando la función `hoslem.test` del paquete `ResourceSelection`:

```
install.packages("ResourceSelection")
```

```
modelo <- glm(
  sta ~ ser + age + sys + gender,
  data = base_icu,
  family = binomial
)

# Obtener probabilidades predichas
prob_pred <- fitted(modelo)

# Re-codificar el desenlace como 0/1 si es necesario
base_icu$sta_num <- ifelse(base_icu$sta == "Died", 1, 0)

# Prueba de Hosmer-Lemeshow (g=10 grupos por defecto)
ResourceSelection::hoslem.test(base_icu$sta_num, prob_pred, g = 10)
```

```
##
##  Hosmer and Lemeshow goodness of fit (GOF) test
##
## data:  base_icu$sta_num, prob_pred
## X-squared = 10.045, df = 8, p-value = 0.2618
```

Los grados de libertad se calculan como el número de grupos menos dos (g − 2), porque el modelo ya utiliza dos parámetros globales (intercepto y pendiente) al estimar las probabilidades. Esos parámetros "consumen" 2 grados de libertad para estimar la forma general de la relación entre variables independientes y desenlace.

Interpretación:

- Dado que el p-valor = 0.2618 es mayor que 0.05, no se rechaza la hipótesis nula. Por lo tanto: No hay evidencia estadísticamente significativa de falta de ajuste. El modelo muestra una calibración aceptable según el test de Hosmer–Lemeshow.

Desempeño global del modelo: AIC, BIC y Deviance

Además de la calibración, debemos evaluar el desempeño global del modelo y compararlo con alternativas (p.ej. un modelo más simple (menos variables), un modelo más complejo (más covariables o interacciones), un modelo con diferentes transformaciones, un modelo con distinta especificación funcional).

Para ello utilizamos medidas basadas en la verosimilitud como la deviance, el AIC y el BIC.

Deviance

En regresión logística, la *deviance* mide la discrepancia entre el modelo ajustado y el modelo saturado (el que ajusta perfectamente los datos; es decir, el modelo que no comete ningún error porque reproduce perfectamente los resultados observados al asignar una probabilidad exactamente igual a lo que ocurrió en cada observación). Indica cuánto se aleja nuestro modelo del ajuste perfecto.

```
summary(modelo)
##
## Call:
## glm(formula = sta ~ ser + age + sys + gender, family = binomial,
##     data = base_icu)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.804363   1.090162  -1.655  0.09790 .
## serMedical    0.900696   0.387143   2.327  0.01999 *
## age           0.029912   0.011266   2.655  0.00793 **
## sys          -0.014791   0.005836  -2.535  0.01126 *
## genderFemale  0.030296   0.385757   0.079  0.93740
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 200.16  on 199  degrees of freedom
## Residual deviance: 177.64  on 195  degrees of freedom
## AIC: 187.64
##
## Number of Fisher Scoring iterations: 5
```

En la salida de R usando `summary(modelo)` vemos:

- Null deviance: 200.16
  - Es la deviance del modelo más simple posible: Aquel que solo incluye el intercepto, asume que todos los individuos tienen la misma probabilidad del evento y no incluye variables independientes. Representa el punto de partida.
- Residual deviance: 177.64
  - Representa el ajuste después de introducir las variables independientes al modelo.

La diferencia entre ambas representa cuánto mejora el modelo al incluir ciertas variables independientes.

Podemos evaluarlo formalmente:

```
anova(modelo, test = "Chisq")
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: sta
##
## Terms added sequentially (first to last)
##
##
##        Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                     199     200.16
## ser     1   6.9190       198     193.24 0.008528 **
## age     1   8.5968       197     184.65 0.003368 **
## sys     1   6.9943       196     177.65 0.008177 **
## gender  1   0.0062       195     177.65 0.937432
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Si la reducción es significativa ($p < 0.05$), el modelo con esa variable independiente mejora significativamente respecto a un modelo que no la considere.

Conceptualmente, se puede interpretar como:

- Menor deviance = mejor ajuste.

AIC (Akaike Information Criterion)

Conceptualmente, el AIC estima la pérdida de información cuando usamos un modelo para aproximar la realidad. Parte de la idea de que ningún modelo es "verdadero", sino una simplificación del proceso generador de datos. El AIC intenta identificar el modelo que mejor equilibra:

- Buen ajuste (alta verosimilitud)
- Baja complejidad (menos parámetros)

Un modelo mejora al aumentar parámetros a ser estimados porque se ajusta más a los datos, pero demasiados parámetros generan sobreajuste.

En R se calcula:

```
AIC(modelo)
## [1] 187.6447
```

El AIC introduce una penalización proporcional al número de parámetros. Por tanto:

- Menor AIC (disminución de 2 unidades respecto a otro modelo) → mejor equilibrio entre ajuste y complejidad
- No tiene interpretación absoluta (como un número aislado)
- Solo sirve para comparar modelos ajustados sobre los mismos datos

En términos prácticos, el AIC es particularmente útil cuando trabajamos en modelos de predicción ya que favorece modelos que predicen mejor, incluso si son un poco más complejos.

BIC (Bayesian Information Criterion)

El BIC tiene una base bayesiana y una filosofía ligeramente distinta. También equilibra ajuste y complejidad, pero:

- Penaliza más fuertemente el número de parámetros
- La penalización aumenta con el tamaño de muestra

Conceptualmente, el BIC intenta aproximar la probabilidad de que un modelo sea el modelo "verdadero" dentro del conjunto comparado.

En R se calcula:

```
BIC(modelo)
## [1] 204.1362
```

- Menor BIC → modelo preferible
- No tiene interpretación absoluta (como un número aislado)
- Tiende a seleccionar modelos más parsimoniosos que el AIC
- Es más conservador cuando el tamaño de muestra es grande

Hasta este punto hemos evaluado el desempeño global del modelo completo, que incluye todas las variables inicialmente consideradas relevantes. Sin embargo, en modelamiento estadístico no basta con identificar qué variables son significativas de forma individual; también debemos preguntarnos si todas ellas aportan información útil al modelo en términos de ajuste y parsimonia.

Para decidir si una variable independiente debe retirarse de un modelo de regresión logística, es importante considerar tanto criterios estadísticos como criterios clínicos. Basar la decisión únicamente en la ausencia de significancia estadística de una asociación puede conducir a modelos incorrectos o a la eliminación de variables relevantes desde el punto de vista conceptual.

En primer lugar, deben evaluarse **criterios estadísticos**. Una variable puede considerarse candidata a eliminación si no muestra asociación estadísticamente significativa con el desenlace y si su exclusión no deteriora el ajuste global del modelo (esto es lo que hemos venido evaluando con indicadores como el AIC y el BIC).

En paralelo, es también fundamental evaluar si la variable podría actuar como **factor de confusión**. Si al retirar la variable los coeficientes cambian en más del 10–15%, esto sugiere que la variable estaba controlando confusión y, por tanto, debería mantenerse en el modelo, incluso si su valor p no es significativo.

Finalmente, algunas variables se consideran importantes a priori por su **plausibilidad o relevancia clínica** o por **evidencia previa en la literatura**, como la edad o el sexo. En estos casos, se puede optar por mantenerlas en el modelo independientemente de su significancia estadística.

Con fines didácticos exploraremos un **modelo reducido** sin la variable `gender` (ya que no mostró una asociación estadísticamente significativa) con el objetivo de comparar ambos modelos mediante criterios como AIC y BIC y determinar si una especificación más simple ofrece un desempeño similar o mejor.

```
modelo_reducido <- glm(sta ~ ser + age + sys,
                       family = binomial,
                       data = base_icu)


AIC(modelo, modelo_reducido)
##                 df      AIC
## modelo           5 187.6447
## modelo_reducido  4 185.6508
BIC(modelo, modelo_reducido)
##                 df      BIC
## modelo           5 204.1362
## modelo_reducido  4 198.8441
## Uso la función tbl_regression para exponenciar los coeficientes del modelo_reducid
o
gtsummary::bold_p(gtsummary::tbl_regression(modelo_reducido, exponentiate = TRUE, pva
lue_fun = ~ gtsummary::style_pvalue(.x, digits = 2)))
```

Characteristic	**OR**	**95% CI**	**p-value**
ser			
Surgical	—	—	
Medical	2.46	1.17, 5.37	**0.020**
age	1.03	1.01, 1.06	**0.008**
sys	0.99	0.97, 1.00	**0.011**

Abbreviations: CI = Confidence Interval, OR = Odds Ratio

Interpretación:

- El modelo reducido tiene menor AIC y BIC
- Al eliminar `gender`, el modelo sigue explicando bien los datos y lo hace de manera más simple. Esto justifica preferir el **modelo reducido** como modelo final, desde el punto de vista estadístico
- Esta tabla final no se usa para elegir el modelo, sino para interpretar el modelo que ya fue seleccionado (en este caso, el modelo reducido):
    - Los pacientes del servicio médico tienen **2.46 veces el odds** de morir en comparación con los pacientes quirúrgicos, **ajustando por** edad y presión sistólica.

O dicho de otra manera:

- El **chance** de morir en los pacientes del servicio médico es **2.46 veces** el **chance** de morir en los pacientes quirúrgicos, **después de ajustar por** edad y presión sistólica.

Nota sobre la traducción de "odds":

En regresión logística, el término inglés *odds* no tiene una traducción directa y universalmente aceptada en español. A veces se utiliza la palabra "*chance*" para facilitar la comprensión; sin embargo, es importante recordar que, aunque en el uso cotidiano del español, chance suele interpretarse como probabilidad, en estadística *odds* y *probabilidad* no son lo mismo.

Recordemos que:

$$odds = \frac{p}{1-p}$$

donde p es la probabilidad del evento.

Por esto, al decir "*el chance de morir en los pacientes del servicio médico es 2.46 veces…*" se debe tener cuidado y no interpretar incorrectamente el resultado como si se tratara de probabilidades o riesgos, cuando en realidad el modelo estima es una razón de odds (OR: *odds ratio*).

Tanto el AIC como el BIC son herramientas comparativas, no pruebas de hipótesis, y nunca reemplazan el juicio clínico ni la coherencia conceptual del modelo.

Lecturas complementarias:

- Kleinbaum, D. G., & Klein, M. (2010). Logistic Regression: A Self-Learning Text (3rd ed.). Springer
- Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression. John Wiley & Sons
- Kleinbaum, D. G., Kupper, L. L., & Muller, K. E. (1988). Applied regression analysis and other multivariable methods (2nd ed.). ChPTER 21. PWS-Kent Publishing Company