

1. Algunas especialidades vienen por parejas: por ejemplo, traumatología y reumatología; o bien urología y nefrología. Entre ellas hay muchas diferencias: por ejemplo, traumatología es quirúrgica atendiendo patologías más agudas ("más nuevas"); y reumatología es médica, con más crónicos (de más larga evolución). El traumatólogo **T** presume con su amigo reumatólogo **M** sobre la satisfacción de sus pacientes y ambos deciden pasar la misma encuesta de valoración ("¿Está Vd. satisfecho con la evolución de su enfermedad?") a aquellos pacientes que llevan más de 3 meses siendo atendidos por ellos. Y observan que las *odds* de satisfacción son 3 a 2 para T y 2 a 3 para R. Razone por qué un posible sesgo de selección o bien efectos confundidos podrían explicar esta situación. Y explique cómo solucionaría ambos retos (sea en la recogida de datos o el análisis). (4p)

Sesgo de selección: explicación

T="más nuevos" frente a M="duran más" implica que los pacientes que queden a los 3 meses serán diferentes. Seleccionar (bloquear) por cierta evolución (respuesta) implica riesgo por sesgo de selección aunque el motivo exacto de la selección sea desconocido. Podría ser que los pacientes vayan pasando con el tiempo de T a M en función de su evolución: Por ejemplo, los pacientes de T que van mal podría ir hacia M. [Que no estén enmascarados genera otro riesgo de sesgo, quizás de información]

Sesgo de selección: solución

La solución sería no seleccionar a los pacientes que llevan 3 meses, sino a los que acuden por primera vez a la consulta y preguntarles a TODOS a los 3 meses.

Efectos confundidos: explicación

El enunciado ya dice que son pacientes diferentes [T="más nuevos" frente M="duran más"]. En cualquier caso, nada garantiza que sean iguales, "comparables". Esta diferencia inicial y la intervención T/M tendrán sus efectos confundidos.

Efectos confundidos: solución

Convendría ajustar (bloquear) por las variables que marcan las diferencias iniciales, para poder comparar situaciones más homogéneas. [Y si fuera cierta la premisa no contrastable de conocer todas las posibles variables confusoras, sería posible afirmar la relación causal.]

2.- Los pacientes con cierta enfermedad grave pueden recibir una intervención en su domicilio o en la ambulancia, camino del hospital. Los investigadores del hospital han recogido la tabla: (6p)

	Buena evolución	Mala evolución
Tratamiento SI	100	200
Tratamiento NO	40	40

a. Calcule e interprete el OR del efecto de la intervención

Cálculo (1p):

$$OR = (100/200)/(40/40) = 1/2$$

Interpretación (1p):

Hay relación inversa: evolucionan mejor los no tratados.

Los investigadores sospechan que el patrón de envío al hospital podría variar en los 4 subgrupos. Así, preguntan al servicio de urgencias móviles, quienes les dicen que siempre van al hospital todos los tratados (primera fila) y los que “intuyen” que tendrán buena evolución (y cabe prever que terminarán en la primera columna). En cambio, de acuerdo con las “últimas voluntades” que constan en la historia informatizada, no suele ser así en los que no cumplen las condiciones previas. Así, usando su memoria, estiman *grosso modo* (“aproximado”) que en las 3 primeras celdas (Trat SI \cup Buena evol) un 80% llega al hospital; pero los de mala evolución y no tratados, sólo llega un 10%.

b. Calcule el efecto de la intervención ajustando mediante IPW esta selección de casos (1p)

Para aplicar IPW hay que multiplicar 3 celdas por $1/0.8 = 5/4$; y la última por $1/0.1 = 10$

IPW / nueva 'n'	Buena evolución		Mala evolución	
T SI	5/4	125	5/4	250
T NO	5/4	50	10	400

$OR = (125/250)/(50/400) = 4$

c. Comente la calidad de la información recogida sobre: (A) los datos observados en la tabla; y (B) las probabilidades de ser enviado al hospital. De acuerdo con esta calidad, ayude a estos autores a redactar su interpretación y sus recomendaciones

Calidad (1p):

Hay 2 calidades de información: A los datos recogidos, que permiten re-análisis y cálculos concretos de su incertidumbre ($IC_{95\%}$) son más documentables, replicables y reproducibles. B es una simple opinión.

Interpretación (1p):

Aunque los datos recogidos apuntan a un efecto negativo de la intervención, un posible sesgo de selección podría ser su explicación.

Recomendaciones (1p):

Nuestra recomendación es seguir por ahora con la intervención, pero repetir el estudio seleccionando en su propio domicilio todos los casos que presentan esta enfermedad, sea cual sea el lugar final de tratamiento.

3. Con lo aprendido en equivalencia, defienda que el consejo de TRIPOD sobre las pruebas de hipótesis del ajuste del modelo, aplica también a las pruebas de Normalidad de Kolmogoroff (o de Shapiro-Wilks). Y sugiera análisis alternativo(s) (4p)

The TRIPOD Statement: Explanation and Elaboration (Box G)

It is common to apply statistical tests for agreement between predicted and observed probabilities using the Hosmer–Lemeshow test or the counterpart tests for survival models (...). Such tests have limited statistical power to evaluate poor calibration and are sensitive to the grouping and sample size: they are often nonsignificant for small N and nearly always significant for large N. Furthermore, they convey no indication of magnitude or direction of any miscalibration, hence the preference for calibration plots.

Igual que equivalencia no requiere la estricta igualdad, el ajuste del modelo no puede ser perfecto (no son “leyes”, sino modelos): basta que sea razonable.

Las pruebas de bondad del ajuste de una distribución continua (Kolmogorov, Shapiro), igual que la pruebas de Hosmer Lemeshow (basadas en X^2 para discretas) NO ponen en H_0 la hipótesis que se desea rechazar (o tener evidencia en su contra). Lo que lleva a varias consecuencias no deseadas: (1) si no se logra rechazar, formalmente no se ha demostrado nada: “ausencia de evidencia, no es evidencia de ausencia”; (2) si aumenta mucho el tamaño muestral, se puede acabar demostrando que hay diferencias aunque sean mínimas, sin relevancia práctica; (3) los valores de P no informan ni de la magnitud de las diferencias, ni de su sentido, ni de si se concentran en algún lado u otro.

Análisis alternativo(s)

El análisis debería ser gráfico y descriptivo. En continuas, por ejemplo, el NPP o gráfico de Normalidad. Como medida descriptiva de la calidad del ajuste la distancia (máxima) de Kolmogoroff entre la distribución empírica y la teórica [O la correlación de Shapiro-Wilks entre datos y modelo.]. Y en discretas, el gráfico (de calibración) entre las probabilidades predichas por el modelo y las observadas.

Si se trata de premisas, conviene tener información sobre su razonabilidad antes de empezar el estudio.

[Si se tratara (¿algún ejemplo real?) de auténtica hipótesis, en el sentido de que se quiere demostrar que la aproximación a la normal es suficientemente buena, podría recurrirse a un planteamiento de equivalencia, poniendo en H_0 aquella distancia al modelo teórico lo suficientemente importante como para negar el modelo.]

4. Ciertos autores sostienen que el *Hazard Rate Ratio (HRR)* es una medida del efecto que sólo entienden estadísticos expertos. Por ello, los médicos que deben usarla para recomendar intervenciones a sus pacientes, basan su criterio sólo en si el valor de P es o no es significativo. Así, estos autores hacen un estudio para mostrar que, si los clínicos entendieran la magnitud del efecto, harían otras recomendaciones. Para ello, proponen seleccionar clínicos senior con buen nivel investigador y comparar sus recomendaciones según la presentación de resultados: a) un HRR con su IC_{95%}; o b) una descriptiva con una diferencia de medianas (DM) de los tiempos de vida. Los dividen en 2 grupos al azar a los que presentan los mismos datos con ambos métodos, pero con distinto orden: en unos HRR, DM; y en otros, DM, HRR —de forma parecida al *cross-over* clásico. Los clínicos puntúan entre 0 (mínimo) y 10 (máximo), el grado en que recomendarían la intervención a cierto paciente prototipo.

Grupo HR_DM	HR	DM	Suma	Difer	n
Media	8,00	6,10	14,10	1,90	10
desv tip			2,10	0,60	
Grupo DM_HR	HR	DM	Suma	Difer	n
media	4,20	4,00	8,20	0,20	10
desv tip			2,00	0,70	

En la tabla figuran las medias de sus recomendaciones, así como las medias y desviaciones típicas de las variables suma y diferencia (Atento a qué diferencia!) de las recomendaciones por el mismo clínico. Se pide:

a. Estime por intervalo el efecto de la presentación.

(3p)

Como mismas 'n, $S^2_{D, POOLED} = (0.6^2 + 0.7^2)/2 = 0.6519^2$

Como las diferencias son ya de HR-DM, se sumarán en lugar de restar:

$$IC_{95\%}(\text{efecto directo}) = \frac{1}{2} [(1.9 + 0.2) \pm 2.1 * 0.6519 \sqrt{(1/10 + 1/10)}] \approx \\ \approx \frac{1}{2} [2.1 \pm 0.6] = [0.75, 1.35]$$

Presentar en forma de HR tiende a subir la recomendación entre 0.75 y 1.35.

b. Ponga a prueba y estime por intervalo el efecto arrastrado ("*carry-over*"). Interprete.

(3p)

Como mismas 'n, $S^2_{S, POOLED} = (2.1^2 + 2^2)/2 \approx 2.05^2$

Efecto grupo = (14.1-8.2) \approx 5.9

SE(efecto grupo) = $2.05 \sqrt{(1/10 + 1/10)} \approx 0.916787871$

t-test: $5.9/0.916787871 \approx 6.435512716$

El nivel de recomendación depende del orden en que se presentan los resultados

$$IC_{95\%}(\text{efecto retardado}) = (14.1-8.2) \pm 2.1 * 2.05 \sqrt{(1/10 + 1/10)} \approx \\ \approx 5.9 \pm 1.9 = 4.0, 7.8 \quad \rightarrow 2, 3.9$$

En el grupo HR_DM, la media de las recomendaciones es entre 2 y 3.9 puntos superior.

Presentar primero HR tiende a subir la siguiente recomendación entre 4 y 7.8 puntos.

La existencia de efecto tardío invalida la estimación anterior (nótese lo complicado que es interpretar los resultados observados)