



Universidade de Vigo

Trabajo Fin de Máster

Validación de un sistema simplificado clínico-oximétrico para identificar pacientes con apnea del sueño

Avelino Domínguez González

Máster en Técnicas Estadísticas

Curso 2019-2020

Propuesta de Trabajo Fin de Máster

<p>Título en galego: Validación dun sistema simplificado clínico-oximétrico para identificar pacientes con apnea do sono</p>
<p>Título en español: Validación de un sistema simplificado clínico-oximétrico para identificar pacientes con apnea de sueño</p>
<p>English title: Validation of a simplified clinical oximetric system to identify patients with obstructive sleep apnea</p>
<p>Modalidad: Modalidad B</p>
<p>Autor/a: Avelino Domínguez González, Universidad de Vigo</p>
<p>Director/a: Javier Roca Pardiñas, Universidad de Vigo;</p>
<p>Tutor/a: Ana Clavería, Servicio Galego de Saúde;</p>
<p>Breve resumen del trabajo: El presente trabajo se basa en medir la exactitud diagnóstica del cuestionario STOP-Bang combinado con la oximetría de pulso nocturna en la apnea de sueño moderada. Los médicos de atención primaria necesitan una herramienta rápida para que optimicen la derivación de pacientes con sospecha de Síndrome de Apnea-Hipoapnea moderado (IHA \geq 15). Se compara la utilidad del diagnóstico del cuestionario STOP-Bang, oximetría nocturna y/o su combinación en Atención Primaria</p>
<p>Recomendaciones:</p>
<p>Otras observaciones:</p>

Don Javier Roca Pardiñas, Profesor de Departamento de Estadística e Investigación Operativa de la Universidad de Vigo, doña Ana Clavería, de Servicio Galego de Saúde, informan que el Trabajo Fin de Máster titulado

Validación de un sistema simplificado clínico-oximétrico para identificar pacientes con apnea del sueño

fue realizado bajo su dirección por don/doña Avelino Domínguez González para el Máster en Técnicas Estadísticas. Estimando que el trabajo está terminado, dan su conformidad para su presentación y defensa ante un tribunal.

En Vigo, a 8 de septiembre de 2020.

El director:

Don Javier Roca Pardiñas

La tutora:

Doña Ana Clavería

El autor:

Don Avelino Domínguez González

Agradecimientos

Quisiera agradecer la dirección de este trabajo fin de máster realizada por Javier Pardiñas así como la tutorización realizada por Ana Clavería. Además agradecer a todos los profesores de las tres universidades gallegas que durante estos años en este Máster de Técnicas Estadísticas me han hecho disfrutar tanto trabajando con números. También agradecer a mi familia por estar ahí y a Lola. Por mi formación científica como biólogo siempre estuve interesado en trabajar la biología desde el punto de vista de los números. Este máster me ha ayudado a comprender que siempre detrás de una gráfica, de un punto, hay una información que nos habla por si misma y que nosotros debemos interpretar y comprender. Este trabajo fin de máster es la materialización de ese interés.

Índice general

Resumen	XI
1. Introducción	1
1.1. Diagnóstico de la apnea-hipoapnea del sueño	2
1.2. Impacto clínico y económico de la apnea del sueño	2
1.3. Cuestionarios utilizados para el SAHS	3
1.4. Objetivos	4
2. Material y métodos	7
2.1. Sujetos de estudio y ámbito	8
2.2. Muestra	9
2.3. Variables	9
2.4. Procedimiento	10
3. Pruebas diagnósticas	13
3.1. Evaluación de Pruebas Diagnósticas y curvas ROC	15
3.2. Clasificación con modelos de respuesta binaria	17
3.3. AUC	17
3.3.1. Validación cruzada	18
3.4. Índice de Youden	21
4. Modelos de regresión de respuesta binaria	27
4.1. Introducción	27
4.2. Modelos GLM	28
4.3. Selección de variables	30
4.3.1. Interpretación de resultados con Odds Ratio	32
4.4. Modelos GAM	33
5. Resultados	37
6. Conclusión	43
6.1. Líneas futuras	44
Apéndice A	47
6.2. Software utilizado	47
Índice de figuras	49
Listado de acrónimos	57

Resumen

Resumen en español

En el presente trabajo se valida un sistema simplificado para identificar pacientes con apnea de sueño. Es necesaria una herramienta rápida para que los médicos de familia optimicen la derivación de pacientes con sospecha de Síndrome de Apnea Hipoapnea de sueño moderada ($IHA \geq 15$). Se propone comparar la utilidad diagnóstica del cuestionario **STOP-Bang** así como la utilización de otras variables comparándose los resultados utilizándose modelos de regresión binaria. Las variables del estudio fueron varias, entre ellas edad, sexo, cuestionario STOP-Bang y datos de la Polisomnografía (PSG) todo llevado a cabo en un entorno de atención primaria.

Resumo en galego

No presente traballo válídase un sistema simplificado para identificar pacientes con apnea de sono. Debido a varios motivos é necesaria unha ferramenta rápida para que os médicos de familia optimicen a derivación de pacientes con sospeita de Síndrome de Apnea Hipoapnea de sono moderada ($IHA \geq 15$). Proponse comparar a utilidade diagnóstica do cuestionario STOP-Bang así como a utilización de outras variables comparándose os resultados utilizándose modelos de regresión binaria. As variables do estudio foron idade, sexo, cuestionario STOP-Bang e datos da Polisomnografía (PSG) nun entorno de atención primaria.

English abstract

In the present work a simplified system is validated to identify patients with sleep apnea. Due to various reasons, a rapid tool is necessary for family physicians to optimize the referral of patients with suspected moderate sleep Hypoapnea Apnea Syndrome ($IHA \geq 15$). We propose to compare the diagnostic utility of the STOP-Bang questionnaire as well as the use of other variables by comparing the results using binary regression models. The study variables were age, sex, STOP-Bang questionnaire and Polysomnography data (PSG) a primary care setting.

Capítulo 1

Introducción

La estadística comienza con un problema, un evento que queremos estudiar. Continúa con la recolección de datos. Procede con el análisis de los mismos y finaliza con unas conclusiones. Nuestro evento a estudiar es la búsqueda de un sistema simplificado para identificar pacientes con apnea de sueño. Se recogen datos de pacientes con varias variables lo más significativas posibles para un ajuste adecuado del modelo a crear. Se analizan los datos y por último obtenemos unas conclusiones que intentan satisfacer todas las dudas planteadas en el problema de partida.

El síndrome de apneas-hipoapneas del sueño (SAHS) es una alteración del sueño caracterizado por pausas en la respiración debido a episodios recurrentes de cierre o colapso total o parcial de la vía respiratoria superior mientras una persona duerme (Fig. 1.1). Estos episodios esporádicos son provocados por el descenso de la saturación de **oxihemoglobina** (SaO_2) y microdespertares. Estas pausas respiratorias se repiten varias veces durante la noche y lleva consigo que la persona que lo sufra no consigue tener un sueño reparador y como consecuencia de esta fragmentación del sueño el paciente padecerá cansancio diurno, alteraciones conductuales, descenso en la calidad de vida, etc.. La **oxihemoglobina** es una de las formas que puede adoptar la hemoglobina. Se trata de hemoglobina que transporta oxígeno desde los pulmones, pasando por el corazón y llegando a todas las partes con células del cuerpo. El cerebro es un órgano que necesita un aporte continuado de oxígeno. El cerebro ante la privación de ese aporte de oxígeno y como mecanismo de defensa provoca que el paciente se despierte para de esa forma reactivar de nuevo el flujo respiratorio hacia él.

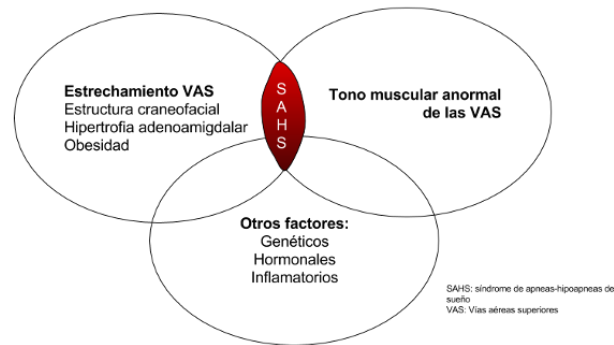


Figura 1.1: Fisiopatología SAHS

1.1. Diagnóstico de la apnea-hipoapnea del sueño

Para el diagnóstico de SAHS se utiliza el índice de apnea-hipoapnea (IAH) del sueño. Este índice se obtiene del resultado de hacer una polisomnografía (PSG) y diferencia 3 tipos de puntos de corte expuestos a continuación.

- Con $IAH \geq 5$, el SAHS es leve y se requiere del paciente hábitos higiénicos y dietéticos saludables.
- Con $IAH \geq 15$ y < 30 el SAHS es moderado y requiere un seguimiento en unidades del sueño específicas.
- Con $IAH \geq 30$ se trata de un SAHS severo y requiere de más acciones paliativas.

1.2. Impacto clínico y económico de la apnea del sueño

Recientemente se trabaja en descubrir el impacto que esta enfermedad causa en la población y su prevalencia. En España se tienen datos que indican una afectación del 4-6% en hombres y un 2-4% en mujeres (González-Mangado N, et al (2015) [11]). Estas cifras son más altas en los países industrializados, en poblaciones con altos índices de obesidad y aumentan con la edad. Son escasos los estudios de tipo comunitario, tal como expone se expone en la reciente revisión sistemática de la **Agency for Healthcare Research and Quality** (AHRQ) (Jonas D, (2017) [10]) .

1.3. Cuestionarios utilizados para el SAHS

Existen varios cuestionarios para identificar de manera adecuada a los pacientes con alta probabilidad clínica de tener SAHS.

- El más usado y validado en España es la **escala de Somnolencia de Epworth** (Johns, M. W. (1991) [1]). La escala de Epworth es una prueba subjetiva muy utilizada en diversos estudios, cuyo objetivo es evaluar la magnitud de la somnolencia diurna frente a 8 situaciones de la vida diaria, pero no tiene relación directa con el diagnóstico de SAHS.
- Cuestionario **Berlin** fue creado por un grupo de expertos (Netzer NC, Stoohs RA, Netzer CM, et al. (1996) [2]). Su gran debilidad es que usó como Gold Standard un monitor portable manejado por los pacientes en vez del PSG para su validación.
- Cuestionario **STOP** (Chung F, Yegneswaran B, Liao P, et al. (2008) [3]) fue el primer desarrollado y validado para pacientes quirúrgicos debido a la necesidad de los anestelistas de saber qué pacientes podrían padecer de SAHS no diagnosticado y que podría suponer un mayor riesgo en operaciones quirúrgicas y post operatorios.
- Cuestionario **STOP-Bang** (Chung F. (2012) [4]). Este trabajo tuvo como objetivo probar el rendimiento del cuestionario STOP-Bang para la sospecha y/o diagnóstico de apnea obstructiva del sueño.

El Cuestionario **STOP-Bang** (Snoring, Tiredness, Observed apnea, Pressure, BMI, Age, Neck circumference, Genre) consiste en un serie de 8 preguntas en las que se responde de forma afirmativa o negativa, ver Fig. 1.2. Las cuatro primeras medidas se trata de medidas subjetivas(STOP), mientras que las 4 siguientes son elementos demográficos. Cada respuesta afirmativa recibe 1 punto y 0 puntos en caso de respuesta negativa. Cuando existen 3 o más puntos el riesgo de SAHS es alto y si el valor es inferior a 3 se considera que el riesgo de SAHS es bajo.

Este cuestionario mostró una mejor sensibilidad para todos los grados de severidad del SAHS. Sin embargo, este cuestionario ha sido validado sólo en pacientes quirúrgicos y no validado en población general.

Se considera que la **polisomnografía convencional** (PSG) es la prueba de referencia (Gold Standard) para el diagnóstico del SAHS (Lloberes et al. 2011 [5]). Es una prueba a realizar en el centro hospitalario que nos aporta información de las variables neurofisiológicas del sueño y cardiorrespirato-

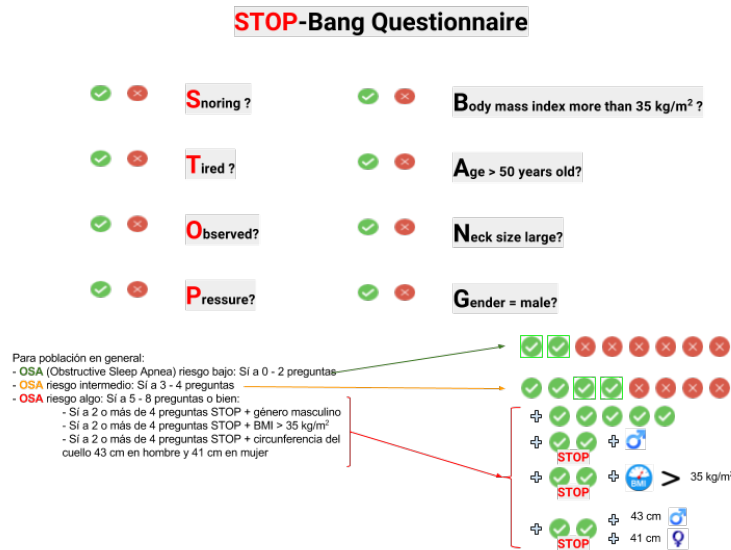


Figura 1.2: Cuestionario STOP-BANG

rias. Pero es un método de alto coste económico y complejo que requiere de personal especializado. Por ello, cada vez hay más estudios que confirman la utilidad de métodos más simplificados y domiciliarios. El más usado es la poligrafía respiratoria (PR), especialmente en aquellos pacientes en que la sospecha clínica es alta. A pesar de ello, la **Academia Americana de Medicina de Sueño (AASM)** recomienda realizar una valoración preoperatoria en una Unidad de Sueño, que incluya PSG en los pacientes antes de someterse a cirugía bariátrica (consenso)(Epstein et al. 2009 [6]).

1.4. Objetivos

Los objetivos de este trabajo van a ser analizar la utilidad diagnóstica de las pruebas seriadas **STOP-Bang** y **pulsioximetría** comparadas con la polisomnografía nocturna que se toma como referencia de diagnóstico. La prevalencia del SAHS es alta aunque es una enfermedad poco diagnosticada. Por este motivo se diseña un estudio con el objetivo de validar estos instrumentos en población española de Atención Primaria para conseguir una herramienta que puedan usar los facultativos rápida y eficazmente.

Nuestra misión va a ser conseguir que la prueba diagnóstica que utilicemos sea lo más sensible y lo más específica posible. El cuestionario STOP-Bang puede por ello convertirse en una herramienta de alta aplicabilidad y rendimiento en el cribado de pacientes con SAHS para médicos de familia.

Lo que queremos por tanto es conseguir un modelo de regresión logística que utilice varias variables explicativas que intenten predecir el comportamiento de una variable respuesta. El objetivo de este proyecto, por tanto, es validar un modelo que combine clínica y pulsioximetría nocturna, en una población comunitaria, frente al estándar oro. De obtener resultados similares, podría convertirse en una herramienta de alta aplicabilidad y rendimiento en el cribado de SAHS moderado para los facultativos de atención primaria, una herramienta que pueden utilizar para hacer un diagnóstico previo de una persona que pueda tener apnea del sueño.

Capítulo 2

Material y métodos

El diseño del experimento se basó en un estudio transversal multicéntrico de utilidad diagnóstica de la combinación de pulsioximetría y síntomas clínicos (cuestionario STOP-Bang) frente al estándar (o patrón) oro de la polisomnografía tipo 1.

El estándar de oro es el test mejor conocido (o combinación de tests) que es considerado como método de referencia para el diagnóstico de una determinada enfermedad. Cualquier nuevo método de diagnosticar una enfermedad necesita ser comparado con este estándar de oro. El estándar de oro para una enfermedad puede ser considerado obsoleto o inadecuado, pero cualquier nueva prueba diseñada para reemplazar el patrón oro debe validarse inicialmente con el patrón oro. Si la nueva prueba es realmente mejor, hay formas de demostrarlo; después de lo cual la nueva prueba puede convertirse en el estándar de oro. La investigación de una determinada enfermedad puede hacerse más precisa debido al uso de una tecnología novedosa que hace replantear todo el trabajo anterior. Por tanto, este patrón de oro no es inamovible y está sujeto a un riguroso trabajo de revisión, que va implícito cuando se trabaja en el ámbito científico.

La pulsioximetría es un método no invasivo (la colocación de un pulsioxímetro es cómodo para el paciente) que va a permitir determinar el porcentaje de saturación de oxígeno de la hemoglobina en sangre de un paciente que se estudia con la ayuda de método fotoeléctricos. Se coloca el pulsioxímetro en una parte del paciente adecuada, con buen flujo sanguíneo, como pueden ser los dedos de la mano o pie o el lóbulo de la oreja. El objetivo es diferenciar la hemoglobina oxigenada (la que transporta el oxígeno, oxihemoglobina) de la desoxigenada (la hemoglobina que vuelve a los pulmones a oxigenarse, desoxihemoglobina). Lo que va realizar el pulsioxímetro es calcular el ratio de oxihemoglobina y

desoxihemoglobina en circulación mediante la siguiente fórmula:

$$SaO_2 = HbO_2 / (HbO_2 + Hb)$$

siendo, SaO_2 la saturación de hemoglobina, HbO_2 es la hemoglobina oxigenada, y Hb es la desoxigenada.

Por otro lado el cuestionario STOP-Bang se utiliza para detectar pacientes con síndrome de apnea obstructiva del sueño (SAHS). Se evalúa en este estudio la asociación entre las puntuaciones de STOP-Bang y **la probabilidad** de SAHS.

La polisomnografía (PSG) convencional es la prueba que registra lo largo de una noche en la cual el paciente tiene que dormir en el hospital o centro de diagnóstico, de manera continuada unas variables neurofisiológicas y cardiorrespiratorias durante el sueño. Para este trabajo se utilizaron estas variables:

VARIABLE	DESCRIPCIÓN	MEDIDA
ps1	Tiempo total de registro (min)	Escala
ps2	Tiempo total de sueño (min)	Escala
ps3	Eficacia del sueño	Escala
ps4	Índice de apneas-hipoapneas	Escala
ps5	Índice de alteraciones respiratorias	Escala
ps6	% tiempo de sueño por debajo O_2 90 %	Escala
ps7	% tiempo de sueño por debajo O_2 95 %	Escala
ps8	Síndrome de Apneas-Hipoapneas	Nominal

Cuadro 2.1: Variables utilizadas en la polisomnografía

2.1. Sujetos de estudio y ámbito

Se trabajó con un población de individuos mayores de 18 años que acudían a la consulta de los facultativos que participaban en el estudio. Como criterio de exclusión se utilizaron: no aceptar par-

participar en el estudio, padecer alguna enfermedad neurológica, trastornos del sueño diagnosticados y enfermedad terminal o estar inmovilizado. En el presente estudio participaron inicialmente un total de 203 sujetos de 6 centros de salud de **Galicia** (Gondomar, Val Miñor, Matamá, Sárdoma, Porriño y A Doblada) y 35 facultativos.

2.2. Muestra

La importancia del cálculo el tamaño de muestra adecuado en las investigaciones médicas es un importante paso para que seamos capaces de detectar resultados significativos.

Para analizar la curva ROC y la correspondiente area under a curve (AUC) , se calculó el tamaño de muestra con la herramienta web **easyROC**. Para un nivel de significación α de 0.05, con AUC estándar = 0.80 y AUC clínicamente relevante del test en análisis=0.8, con relación 1:1, se precisan 91 casos y 91 controles. Asumiendo un 10 % de pérdidas, el número total de la muestra debería ser alrededor de 180 individuos.

2.3. Variables

Las variables a estudio se clasificaron en 5 categorías:

- **Variables sociodemográficas:** Edad, sexo, ocupación.
- **Variables clínicas:** Tabaco (cigarros/día), alcohol (unidades/semana), antecedentes (sí/no) de diabetes, reflujo gastroesofágico, asma, depresión, arritmias, insuficiencia cardiaca, problemas cardiovasculares, tratamientos con fármacos depresores del SNC, otras enfermedades clínicamente sugerentes de SAHS (cefalea matutina, falta de concentración, cansancio, disfunción sexual, nicturia), menopausia, accidentes).
- **Cuestionario STOP-Bang**, el cual puntúa con 1 la presencia de ronquidos (de la pregunta "¿Ronca usted fuerte?"), somnolencia diurna, paradas respiratorias durante el sueño y elevada presión arterial, IMC mayor de 30, edad superior a los 50 años, circunferencia del cuello mayor de 40 cm y sexo masculino.
- **Pulsioximetría nocturna**, se considera el tiempo de sueño por debajo de SaO_2 del 90%.

- **PSG**, se analiza el índice de apneas-hipoapneas y la evaluación clínica por facultativo especialista en Neurofisiología. El registro se evaluó manualmente en épocas de treinta segundos de acuerdo a los criterios de **Rechtschaffen** y **Kalen** (1969) [7]. Se considera la presencia de apnea en el caso de parada del flujo oronasal de al menos 10 segundos de duración e hipoapnea un descenso mínimo del flujo acompañado de una caída de un 4 % en la SaO₂ con o sin arousal acompañantes. La variable de resultado es la apnea de sueño moderada (sí/no), definida como el índice IHA igual o superior a 15 observado en la PSG.

2.4. Procedimiento

La captación de los pacientes para el estudio se llevó a cabo por los facultativos participantes. Se solicitó la inclusión de pacientes con síntomas subjetivos de SAHS y un número similar de pacientes sin sintomatología. A cada uno se le explicó el estudio, se le solicitó un consentimiento informado y se realizó una entrevista clínica. Todos los participantes recibieron las mismas instrucciones (no debían tomar café, té o cola) y se programó el día de la prueba.

El estudio se llevó a cabo en un espacio habilitado para ello, en ausencia de ruido, con personal sanitario acompañante en el Centro de Salud del Val Miñor, con atención las 24 horas del día. Se siguieron los requisitos y recomendaciones publicados por Sánchez Armengol. El equipo utilizado fue el **PSG Cadwell Easy II** de 32 canales; incluye un pulsioxímetro Nonin, que analiza la saturación de oxígeno con el EasyNet SpO₂ Oximeter Module y almacena los valores de saturación detectados cada 4 segundos. A continuación, un profesional acreditado (neurofisióloga) realizó el diagnóstico clínico de SAHS por comunicación electrónica con el equipo PSG y sin acceso a la historia clínica del paciente.

Finalmente, se remitió al médico facultativo el informe de las exploraciones para citar al paciente, informarle de los resultados y llevar a cabo las actuaciones que correspondieran de acuerdo con los protocolos.

El fichero de datos incluía las siguiente variables que fueron las que se utilizaron para crear el modelo que pueden verse en la tabla siguiente.

NOMBRE	ETIQUETA	MEDIDA	NOMBRE	ETIQUETA	MEDIDA
s1 - S	¿ronca usted fuerte?	Nominal	s1.1	Frecuencia (roncar)	Ordinal
s2 - T	¿cansancio o somnolencia durante el día?	Nominal	s2.1	Frecuencia (somnolencia)	Ordinal
s3 - O	¿lo vio alguien dejar de respirar durante el sueño?	Nominal	s3.1	Frecuencia (apnea)	Ordinal
s4 - P	¿tiene hipertensión o está siendo tratado por hipertensión?	Nominal	s5 - B	Índice de masa corporal (kg/m ²)	Nominal
s5.a	IMC kg/m ²	Escala	s6 - A	Edad	Nominal
s6.a	Edad (años)	Escala	s7 - N	Circunferencia del cuello (cm)	Nominal
s7.a	Cuello (cm)	Escala	s8 - G	¿sexo masculino?	Nominal
ps1	va a medir el tiempo total de registro (min)	Escala	ps2	tiempo total de sueño (min)	Escala
ps3	eficacia del sueño	Escala	ps4	Índice de apneas-hipoapneas	Escala
ps5	Índice de alteraciones respiratorias	Escala	ps6	Saturación media de O ₂ durante el sueño (mm de Hg)	Escala
ps7	% de tiempo de sueño por debajo de saturación de O ₂ del 90%	Escala	ps8	% de tiempo de sueño por debajo de saturación de O ₂ del 95%	Escala
ps9	Síndrome de apnea-hipoapnea del sueño	Nominal	IHA5	IHA valor 5	Ordinal
iha15	iha mayor 15	Nominal	IHA30	IHA mayor 30	Ordinal
sb	suma stop-bang	escala	iha5r	IHA < 5	nominal
iha15r	IHA valor 15	Nominal	ihar30	IHA valor 30	Nominal

Cuadro 2.2: Variables utilizadas en el estudio, remarcando las que se utilizan para el STOP-Bang

Este trabajo va a utilizar un modelo de respuesta binaria en el cual la variable que mide el **Índice Hipoapnea-Apnea** de pacientes con un valor mayor de 15, llamada *ihar15*, se va a utilizar como variable respuesta y se va a estudiar el modelo de otras variables que hacen referencia al cuestionario STOP-Bang. Se realizaron varios modelos con estas variables.

Capítulo 3

Pruebas diagnósticas

Una prueba diagnóstica es un procedimiento médico realizado para detectar, diagnosticar o monitorizar enfermedades, procesos de enfermedades o determinar el coste de un tratamiento. Suelen tratarse de pruebas clínicas y los resultados suelen utilizarse apoyados en herramientas estadísticas.

Una prueba diagnóstica puede ser mostrada utilizando una tabla 2x2 en la cual aparece en las columnas la presencia o ausencia de enfermedad (resultados reales) y en las filas los resultados de la prueba diagnóstica que queremos testear. Se puede ver en la Fig. 3.1 una serie de pruebas diagnósticas simples. En esta tabla se identifican los individuos en presencia o ausencia de enfermedad (+ para presencia y - para no presencia).

		Presencia de Enfermedad	
		+	-
Prueba diagnóstica	+	VP	FP
	-	FN	VN

$\text{Valor predictivo positivo (VPP)} = \text{VP} / (\text{VP} + \text{FP})$

$\text{Valor predictivo negativo (VPN)} = \text{VN} / (\text{FN} + \text{VN})$

$\text{SENSIBILIDAD} = \text{VP} / (\text{VP} + \text{FN})$

$\text{ESPECIFICIDAD} = \text{VN} / (\text{VN} + \text{FP})$

VP = verdadero positivo
FN = falso negativo
FP = falso positivo
VN = verdadero negativo

Figura 3.1: Pruebas diagnósticas

Es necesario definir varios términos que se utilizan comúnmente en las pruebas diagnósticas tales como:

1. **Sensibilidad.** Es el número de individuos (o proporción de ellos) con la enfermedad y que

nuestra prueba diagnóstica cataloga correctamente con enfermedad. Su fórmula sería: VP (verdaderos positivos) / (VP + FN (falsos negativos)). En la sensibilidad estudiamos a los individuos enfermos.

2. **Especificidad.** Es el número de individuos (o proporción de ellos) sanos y tienen una prueba negativa. Su fórmula sería: VN (verdadero negativo) / (VN + FP). Por tanto, lo que se estudia son los individuos sanos.
3. **Valor predictivo positivo (VPP).** Es la probabilidad de que un individuo con resultado positivo en nuestra prueba tenga realmente la enfermedad. Son los positivos de nuestra prueba entre el total de positivos. Su fórmula sería: VP / (VP + FP (falsos positivos)).
4. **Valor predictivo negativo (VPN).** Es la probabilidad de que un individuo con resultado negativo en la prueba no tenga la enfermedad. Son los sanos con resultado del test negativo dentro de todos aquellos con prueba negativa. Su fórmula sería: VN / (FN + VN).
5. **Valor global de la prueba o precisión de la prueba.** Es la probabilidad de que un individuo sea clasificado correctamente por la prueba. Es decir que un individuo sano sea diagnosticado según la prueba como sano, y un individuo enfermo sea diagnosticado correctamente como enfermo. La fórmula sería: (VP + VN) / (VP + VN + FP + FN).
6. **Razón de verosimilitud positiva (RVP).** La RVP Es la probabilidad de una prueba positiva en presencia de la enfermedad (sensibilidad), dividida por la probabilidad de una prueba positiva en ausencia de la enfermedad (1- especificidad). Este indicador nos mide cuantas veces es más probable que el test sea positivo en los éxitos.
7. **Razón de verosimilitud negativa (RVN).** La RNP Es la probabilidad de una prueba negativa cuando la enfermedad está presente (1 - sensibilidad), dividida por la probabilidad de una prueba negativa en ausencia de la enfermedad (especificidad). Este indicador nos mide cuantas veces es más probable que el test sea negativo en los enfermos que en los no enfermos.

La sensibilidad y la especificidad son inversamente proporcionales, es decir, si aumenta una disminuye la otra y viceversa. Todo va a depender del valor umbral que tomemos como referencia porque en función de ese valor vamos a tener más positivos o más negativos en nuestra prueba diagnóstica. Nosotros partimos de saber exactamente los pacientes que son positivos y la prueba diagnóstica tendrá más aciertos o fallos en función de ese valor umbral.

Tanto la RVP como la RVN dependen de la prevalencia de una enfermedad (proporción de individuos que presentan una enfermedad o característica a estudio en la población). La prevalencia del **SAHS** es alta, pero es una enfermedad infradiagnosticada. Si se pudiera hacer un estudio de sueño a todos los pacientes sintomáticos sería lo ideal, pero la falta de recursos y el insuficiente número de unidades de sueño hace necesario desarrollar cuestionarios de "despiste" que seleccionen a aquellos pacientes más graves para su pronto diagnóstico y tratamiento y así reducir la morbi-mortalidad y los costes asociados.

Un instrumento de elevada sensibilidad como es el uso del cuestionario STOP-Bang es de gran utilidad en centros de atención primaria ya que permite detectar los pacientes más graves y derivarlos rápidamente a la consulta especialista y ahorrar el coste que supondría la falta de tratamiento en estos pacientes. Una tabla resumen con términos y definiciones el cuadro 3.1:

Término	Definición
Apnea	Cese del flujo de aire por al menos 10 segundos
Hipoapnea	Reducción del flujo de aire al menor un 30 % por al menos 10 segundos con bajada en la saturación del oxígeno
IAH	Índice Apnea Hipoapnea
<i>Leve</i>	Con IAH ≥ 5 hasta < 15
<i>Moderado</i>	Con IAH ≥ 15 y < 30
<i>Grave</i>	Con IAH ≥ 30 se trata de un SAHS severo
SAHS	Síndrome Apnea Hipoapnea
<i>Leve</i>	Con IAH ≥ 5 , el SAHS es leve y se requiere del paciente hábitos higiénicos y dietéticos saludables.
<i>Moderado</i>	Con IAH ≥ 15 y < 30 el SAHS es moderado y requiere un seguimiento en unidades del sueño específicas.
<i>Grave</i>	Con IAH ≥ 30 se trata de un SAHS severo y requiere de más acciones paliativas.

Cuadro 3.1: cuadro resumen

3.1. Evaluación de Pruebas Diagnósticas y curvas ROC

Las **receiver operating characteristics (ROC)** son una popular manera de visualizar, organización y selección de clasificadores basados en su rendimiento. Es decir, son la representación gráfica de la capacidad discriminatoria que tiene una prueba a partir de todos los posibles "puntos de corte". En el eje Y se sitúan los valores de la *sensibilidad* y en el eje X los valores de *1-especificidad*. Visualizan, por tanto, los puntos de corte entre sensibilidad y especificidad en clasificadores binarios.

Para dibujar una curva ROC podemos hacer el simul de utilizar una tortuga. La tortuga considera todos los casos que tienen un valor positivo en una prueba. Dependiendo si el valor es realmente posi-

tivo o un falso positivo se comportará de forma u otra. Cuando la tortuga pasa por un valor verdadero positivo (TP, true positive) seguirá su camino en dirección hacia el eje y, es decir, hacia arriba en el gráfico y cuando pasa por un valor falso positivo tomará una dirección hacia la derecha, hacia el eje x. El tamaño de los pasos son inversamente proporcionales al número de actuales positivos (en la dirección del eje y) o negativos (en la dirección del eje x), así que el paseo siempre finaliza en la coordenada (1, 1). El resultado es un gráfico de los TPR (true positive rate o sensibilidad) y los FPR (false positive rate o 1 - especificidad). Un ejemplo de curva ROC puede verse en la Fig. 3.2.

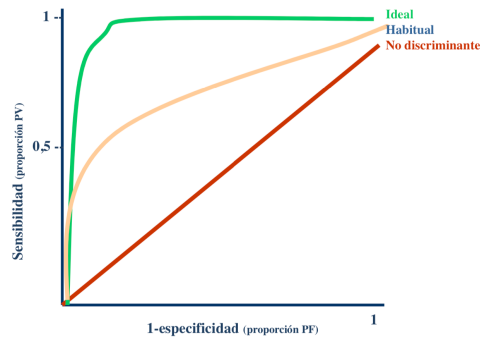


Figura 3.2: Ejemplo de curva ROC

La curva ROC es una curva interpolada hecha de puntos cuyas coordenadas son funciones puntos de corte:

punto de corte (threshold) = $\theta \in \mathbf{R}$, y donde $\theta \in [0, 1]$

$$ROC_x(\theta) = FPR(\theta) = \frac{FP(\theta)}{FP(\theta) + TN(\theta)} = \frac{FP(\theta)}{\#N}$$

$$ROC_{xy}(\theta) = TPR(\theta) = \frac{TP(\theta)}{FN(\theta) + TP(\theta)} = \frac{FP(\theta)}{\#P} = 1 - \frac{FN(\theta)}{\#P} = 1 - FNR(\theta)$$

siendo $\#N$ el número total de casos y $\#P$ el número total de positivos.

En términos de test de hipótesis donde rechazar la hipótesis nula se considera un resultado positivo, la tasa de falsos positivos (FPR) corresponde al **error de tipo I** y la tasa de falsos negativos (FNR) al error de tipo II y (1 - FNR) a la potencia. Un error de tipo I es cuando se rechaza incorrectamente una hipótesis nula verdadera (falso positivo), mientras que un error de tipo II es cuando uno acepta incorrectamente una hipótesis nula falsa (falso negativo). Los errores de tipo I solo se basan en el nivel alfa de la prueba (fácil de controlar), mientras que los errores de tipo II se basan en la potencia (muchos determinantes y por tanto no se controlan fácilmente).

3.2. Clasificación con modelos de respuesta binaria

Principalmente dos razones son responsables de por qué una curva ROC es una métrica potencialmente poderosa para la comparación de diferentes clasificadores. Uno de ellos es que el ROC resultante es invariante frente al sesgo de clase del conjunto de datos aplicado, lo que significa que un conjunto de datos con 60% de etiquetas positivas producirá el mismo ROC (esperado estadísticamente) como un conjunto de datos con 45% de etiquetas positivas (aunque esto afectará el coste asociado con un punto dado de la ROC). La otra es que la ROC es invariante frente a la puntuación evaluada, lo que significa que podríamos comparar un modelo que da puntuaciones no calibradas como una regresión lineal regular con una regresión logística o un modelo de bosque aleatorio cuyas puntuaciones pueden ser consideradas como probabilidades de clase.

3.3. AUC

El área bajo la curva ROC también se llama comúnmente Área Bajo la Curva (AUC). Cuanto más cercana esté la curva ROC a la esquina superior izquierda más exacta será la prueba diagnóstica. En cambio, si la curva ROC está sobre la diagonal la capacidad diagnóstica será nula.

El AUC tomará valores en el intervalo $[0, 1]$ y se interpreta del siguiente modo:

- El AUC de un test perfecto es el que discrimina perfectamente.
- Un test no informativo tendrá una AUC igual a 0.
- Por tanto, la AUC puede ser considerada como una medida resumen de la curva ROC y puede servir para comparar la capacidad diagnóstica de distintos marcadores.

El Área Bajo la Curva (AUC) ofrecen interesantes interpretaciones:

- El AUC tiene una propiedad estadística importante: el AUC de un clasificador es equivalente a la probabilidad de que el clasificador clasifique una instancia positiva elegida aleatoriamente más alta que una instancia negativa elegida al azar.
- El AUC también tiene varias interpretaciones intuitivas naturales, una de las cuales es que es la sensibilidad media de un clasificador bajo la suposición de que es igualmente probable que elija cualquier valor de la especificidad - bajo la suposición de una distribución uniforme sobre la especificidad.

Dado que el propio ROC es variable con respecto a un conjunto de datos dado, es necesario el promedio de múltiples ROCs derivados de diferentes conjuntos de datos para llegar a una buena estimación de la función ROC verdadera de un clasificador. Para este trabajo se han elegido dos variables que pueden utilizar los facultativos, y se toma como referencia el verdadero valor, nuestra variable Y , que va a ser la variable $iha15$ obtenida tras la polisomnografía y que clasifica sin ningún género de dudas a los individuos enfermos de los sanos.

Es decir, el AUC de la curva ROC es equivalente a la probabilidad de que un sujeto elegido al azar posea un resultado del test menor que el resultado del test de un individuo enfermo elegido al azar. Se ha utilizado la variable $iha15$ como variable independiente y las variables sb (resultado del stop-bang) y $ps7$ (% de tiempo de sueño por debajo del 90%) obteniéndose un valor de **0.8801** de AUC tal como se muestra en la Fig. 3.3.

Area under the curve: 0.8801

95% CI: 0.8296-0.9306 (DeLong)

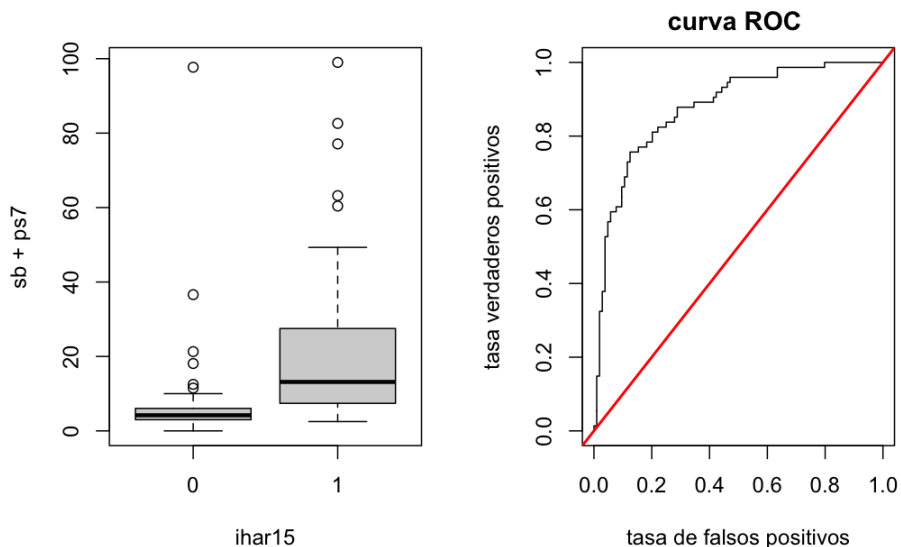


Figura 3.3: Boxplot variables utilizadas y curva ROC

3.3.1. Validación cruzada

Las medidas de diagnóstico (precisión, AUC, . . .) vistas para obtener la precisión de un clasificador pueden no ser realistas ya que se ha utilizado el mismo conjunto de datos para establecer el modelo y

[1] 0.8018519

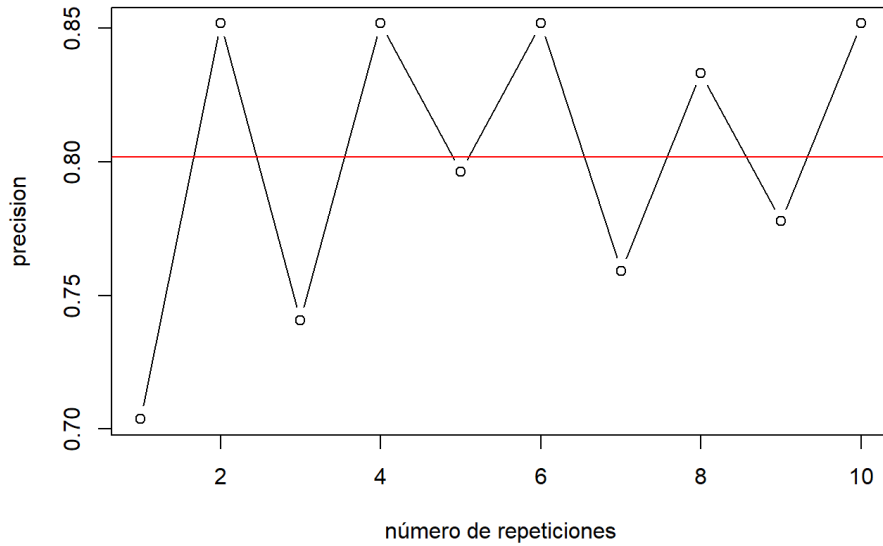


Figura 3.5: Validacion cruzada utilizando método esquina superior izquierda (closest.topleft)

La esquina superior izquierda es un criterio de predicción que se situaría en un punto en la esquina superior izquierda, o coordenada (0,1) del espacio de la curva ROC, representando un 100 % de sensibilidad (es decir, ningún falso negativo) y un 100 % también de especificidad (ningún falso positivo). Nótese que el valor de precisión que habíamos obtenido antes (utilizando los mismos datos para ajustar el modelo y para evaluar su funcionamiento) era del 85.4%. Sin embargo, al utilizar este método de validación cruzada la precisión obtenida se reduce al 80,1%. Ahora repetimos lo anterior utilizando como criterio el AUC.

k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10
0.810	0.942	0.839	0.883	0.842	0.866	0.838	0.888	0.842	0.896

[1] 0.8646636

Se obtiene un media por validación cruzada de 0.86 y el resultado puede verse en la Fig. 3.6

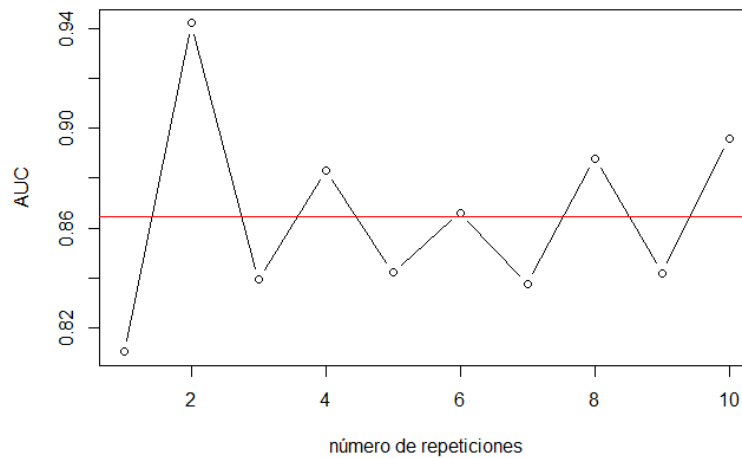


Figura 3.6: Validacion cruzada utilizando método AUC

Éste valor de **0.86** es similar al valor de AUC calculado anteriormente de **0.88**.

3.4. Índice de Youden

El **índice de Youden** (IY) nos va a permitir relacionar la sensibilidad y la especificidad de una prueba diagnóstica y medir su eficacia (Youden, W. J. (1950) [8]). El test de diagnóstico ideal debería discriminar de manera infalible entre individuos enfermos y sanos. Si existiese ese test no habría problema estadístico. La búsqueda de ese test infalible la llevamos a cabo mediante varios test de diagnóstico que dan lugar a un éxito parcial. Por tanto, existe la necesidad de encontrar un índice que permita calificar las pruebas de diagnóstico de manera objetiva y elegimos la que mejor encaja entre nuestro test diagnóstico y el valor real de estudio.

El **IY** Viene dado por:

$$IY = \text{sensibilidad} + \text{especificidad} - 1 = P(\text{verdaderos positivos}) - P(\text{falsos positivos})$$

Debido a que el **YI** maximiza las diferencias entre la fracción de verdaderos positivos (TPF) y la fracción de falsos positivos en todos los puntos de corte puede escribirse de la siguiente forma:

$$YI = \max(\text{Sensibilidad} + \text{Especificidad} - 1)$$

Cuánto más cercano esté el **IY** a 1, mejor será la prueba diagnóstica que se está evaluando. Es un índice no afectado por la probabilidad de la enfermedad. Uno de sus inconvenientes es que deja de ser una buena prueba diagnóstica si es buena en especificidad o sensibilidad.

Los puntos de corte van a definir todos los cálculos que realicemos con nuestros datos. La elección de un buen punto de corte es fundamental para la clasificación de los datos. El punto de corte θ se obtiene de la siguiente regla de clasificación:

- $\text{test} \geq \theta$, se clasifica al individuo como enfermo.
- $\text{test} < \theta$, se clasifica al individuo como sano.

El empleo de una prueba diagnóstica va a llevar consigo la elección de un valor de punto de corte. No existe una metodología cerrada para calcular el valor de corte óptimo es habitual determinar dicho corte a través de algunos de los siguientes criterios:

- Punto de corte θ que maximiza el índice de Youden.
- Elegir un punto de corte θ que corresponda con aquel valor de la curva ROC más cercano al vértice superior izquierda del gráfico.
- Otra forma es utilizar un punto de corte que iguale el valor de sensibilidad y especificidad.

Para la selección de los puntos de corte se utilizó el paquete **Optimalcutpoints**. Se utilizó como marcador *ps7*, como variable respuesta *ihar15* y teniendo en cuenta la variable *sexo* obtenemos la siguiente salida que se muestra en la Fig. 3.7 y Fig 3.8. Se obtiene un valor de AUC de 0,91 para mujeres y de un 0,87 para hombres.

Como marcador se utilizó la variable *ps7* y como variable respuesta *ihar15*. Se discriminó por sexo y se obtuvieron los siguiente resultados:

```
optimal.cutpoints.default(X = "ps7", status = "ihar15", tag.healthy = 0,
methods = "Youden", data = ap, categorical.cov = "sexo",
pop.prevalence = NULL, control = control.cutpoints(), ci.fit = TRUE,
conf.level = 0.95, trace = FALSE)
```

```
*****
```

```
MUJERES
```

```
*****
```

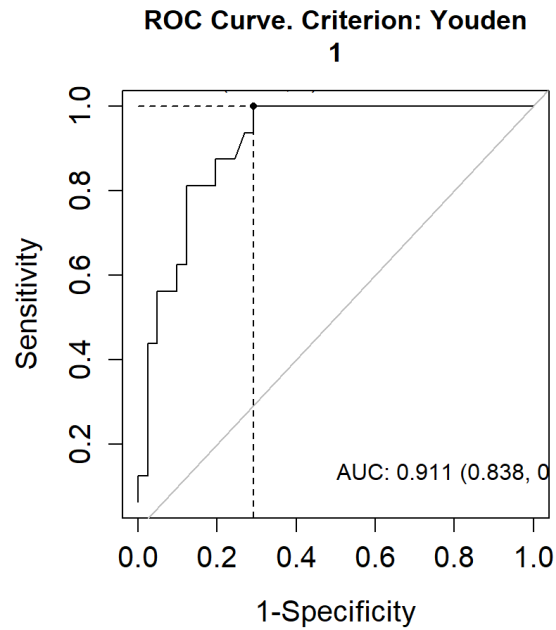


Figura 3.7: Curva ROC para Mujeres

Area under the ROC curve (AUC): 0.911 (0.838, 0.984)

CRITERION: Youden

Number of optimal cutoffs: 1

Estimate 95% CI lower limit 95% CI upper limit

cutoff	0.3000000	-	-
Se	1.0000000	0.7940928	NaN
Sp	0.7073171	0.5446260	0.8387014
PPV	0.5714286	0.3975397	NaN
NPV	1.0000000	0.8748437	1.0000000
DLR.Positive	3.4166667	2.1229815	5.4986871
DLR.Negative	0.0000000	0.0000000	NaN
FP	12.0000000	-	-
FN	0.0000000	-	-

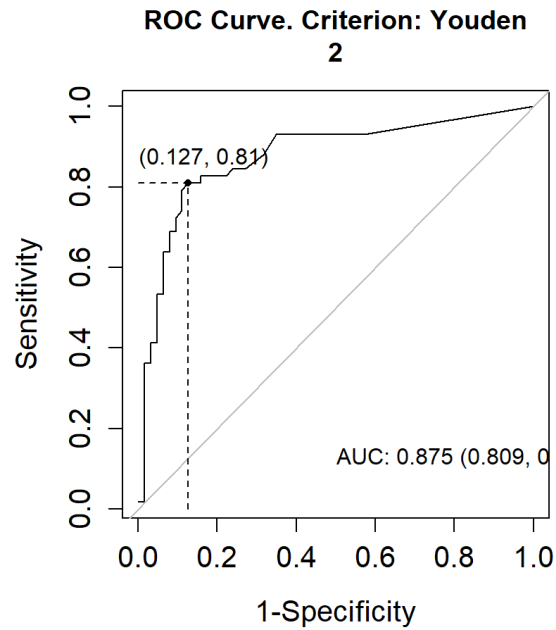


Figura 3.8: Curva ROC para Hombres

Optimal criterion 0.7073171

HOMBRES

Area under the ROC curve (AUC): 0.875 (0.809, 0.941)

CRITERION: Youden

Number of optimal cutoffs: 1

Estimate 95% CI lower limit 95% CI upper limit

cutoff	1.4000000	-	-
Se	0.8103448	0.6859491	0.9013362
Sp	0.8730159	0.7650344	0.9435487

PPV	0.8545455	0.7356142	0.9262603
NPV	0.8333333	0.7187832	0.9239886
DLR.Positive	6.3814655	3.3004999	12.3384650
DLR.Negative	0.2172414	0.1265665	0.3728777
FP	8.0000000	-	-
FN	11.0000000	-	-
Optimal criterion	0.6833607	-	-

Por tanto se obtuvo para las mujeres un área bajo la curva de 0,91, definiendo el punto (0.83, 0,98) con una sensibilidad del 100 %, una especificidad del 70 %, un valor predictivo positivo del 57 % y un valor predictivo negativo de un 100 %. En el caso de los hombres se obtuvo un área bajo la curva de 0.87, en el punto (0.80, 0,94) con una sensibilidad del 81 % y una especificidad del 87 %, un valor predictivo positivo del 85 % y un valor predictivo negativo del 83 %.

Capítulo 4

Modelos de regresión de respuesta binaria

4.1. Introducción

Para representar la dependencia que posee una variable Y (variable respuesta o variable dependiente) con respecto a otra u otras variables X (variable explicativa o variable explicativa), se utilizan los modelos de regresión. Por tanto, los análisis de modelos de regresión es una colección de procedimientos estadísticos para estimar las relaciones entre variables.

Mediante los modelos de regresión tratamos de obtener una clasificación supervisada” donde se dispone de una muestra de individuos divididos en dos posibles grupos dados por una variable respuesta binaria Y medidora de la ocurrencia (éxito, $Y=1$) o no (fracaso, $Y=0$) de un suceso de interés. Además de cada individuo se dispone de la información dada por un conjunto de variables explicativas $\mathbf{X} = (X_1, \dots, X_p)$. Con estos ingredientes interesará conocer

- Obtener la probabilidad condicionada de éxito $\pi(\mathbf{X}) = \pi(Y=1|\mathbf{X})$.
- Relacionar la probabilidad anterior con las variables explicativas, y seleccionar las variables importantes.
- Clasificar en éxito/fracaso a partir de las probabilidades anteriores, y evaluar la capacidad de clasificación del modelo.

El valor de $\pi(X)$ coincide con la respuesta media $E[Y|X]$. Sin embargo, la utilización de un modelo lineal del tipo

$$\pi(X) = a_0 + a_1 \cdot X_1 + \dots + a_p \cdot X_p$$

tendrá varios problemas como serían:

1. **Normalidad.** No se cumple debido a que el error ϵ en este caso es una variable discreta que solo puede tomar dos valores. Dado X , solo hay dos posibles valores para el error: $1 - \pi(X)$ y $0 - \pi(X)$.
2. **Homocedasticidad.** Además, la varianza condicionada $Var(Y|X) = \pi(X) \cdot (1 - \pi(X))$, que depende del vector de covariables X y no se cumple la suposición de homocedasticidad.
3. Por último, el modelo puede dar lugar a probabilidades absurdas fuera del intervalo $[0, 1]$ y es una restricción que no está presente en modelos clásicos, en los cuales $p(X)$ puede tomar cualquier valor en la recta real \mathbf{R} .

ya que no se cumplen las suposiciones de normalidad y de homocedasticidad que se suelen suponer en la realización de este tipo de modelos.

Una forma de evitar este inconveniente es utilizar una transformación de $\pi(X)$ que puede tomar cualquier valor. Una de estas transformaciones es la razón de probabilidades comúnmente conocida como *Odds*,

$$\text{Odds}(X) = \frac{\pi(X)}{1 - \pi(X)}$$

Una transformación logarítmica de Odds proporciona una medida que varía entre $-\infty$ y $+\infty$, denominada transformación **logit**:

$$\text{Logit}(X) = \log \text{Odds}(X) = \log \frac{\pi(X)}{1 - \pi(X)}$$

4.2. Modelos GLM

Los Modelos Lineales Generalizados (GLM) permiten relajar un poco el supuesto de linealidad estricta de los modelos lineales, al permitir que el valor esperado de la respuesta dependa de una función monótona suave del predictor lineal. De manera similar, la suposición de que la respuesta se distribuye normalmente se relaja al permitir que siga cualquier distribución de la familia exponencial (por ejemplo, normal, Poisson, binomial, gamma, etc.). Los modelos logísticos son el resultado de aproximar linealmente la transformación logit:

$$\text{Logit}(X) = \log \text{Odds}(X) = \log \frac{\pi(X)}{1 - \pi(X)} = a_0 + a_1 \cdot X_1 + \dots + a_p \cdot X_p$$

que da lugar a un modelo equivalente

$$\pi(X) = \log \text{Odds}(X) = \frac{\exp(a_0 + a_1 X_1 + \dots + a_p X_p)}{1 + \exp(a_0 + a_1 X_1 + \dots + a_p X_p)}$$

Para estimar el modelo anterior se dispondrá de una muestra $(X_1, Y_1), \dots, (X_n, Y_n)$ siendo $X_i = (X_{i1}, \dots, X_{ip})$.

La estimación resultante se hará por el método de máxima verosimilitud, tratando de obtener los coeficientes que maximizan el logaritmo de la función de verosimilitud.

$$\sum_{i=1}^n Y_i \cdot \log \pi_i + (1 - Y_i) \cdot \log (1 - \pi_i)$$

siendo

$$\pi_i = P(Y_i = 1 | X_i)$$

Las ecuaciones resultantes del problema de maximización son no lineales y se requiere el uso de un método iterativo. Uno de los más utilizados es el algoritmo de **Fisher scoring** (McCullagh P et al (1989) [12]), el cual cada iteración del algoritmo puede ser vista como un problema de mínimos cuadrados ponderados.

Un **modelo lineal generalizado (GLM)** es una generalización flexible de la regresión lineal que tienen modelos de distribución de errores distintos de una distribución normal. El modelo matemático que se busca en un modelo lineal generalizado es encontrar una interpretación de la relación que existe entre una variable respuesta Y , con un conjunto de p variables \mathbf{X} (X_1, \dots, X_p) conocidas también como covariables. Se busca por tanto conocer la influencia de los valores de las covariables X en la respuesta media.

Para hacer estimación e inferencia en los modelos lineales logísticos se puede utilizar la función **glm** del paquete **stats** de R. Esta función permite:

- estimar el modelo,
- construir intervalos de confianza, y
- realizar contrastes de hipótesis.

Además esta función **glm** permite definir varios tipos de modelos, cada uno con su propia "familia". Esto significa que con familia especificamos el tipo de variable respuesta que tenemos y qué tipo de

modelo nos gustaría usar. Los argumentos de esta función serían: `formula`, `family` y `data`. `Formula` y `data` son los mismos que se usan en los modelos de regresión lineal. Si se está trabajando con un `data frame`, la formula sería $y \sim x_1 + x_2 + \dots + x_p$ y luego con `data =` se escribe el nombre del `data frame`. Para regresión logística se utiliza el argumento `family = binomial`. Hay que recordar que los modelos de distribución binomial modelan la probabilidad de que los experimentos sean éxitos o fracasos (de igual modo que la variable respuesta).

4.3. Selección de variables

Un problema importante en la aplicación de modelos de regresión es seleccionar las variables importantes que entran en el modelo. De todas las variables que estamos utilizando en el estudio tenemos que quedarnos con las que nos van a resultar más útiles, nos van a dar mejores resultados explicando la variabilidad de los datos. Se realizaron varios modelos trabajando con todas las variables. Con este objetivo se puede utilizar la función `stepwise` del paquete `RcmdrMisc` que permite seleccionar el modelo por pasos según criterio **AIC** o **BIC**. En modelos de regresión binaria, si el interés se centra en la clasificación, parece razonable que la selección de variables se haga teniendo en cuenta alguna medida de diagnóstico (por ejemplo, el AUC o la exactitud). Se seleccionaron 2 modelos de regresión utilizando para la selección de variables la función `stepwise` y como criterio **BIC** (Bayesian Information Criterion):

- Modelo con variable respuesta `ihar15` y como variables explicativas `s1`, `s7` y `ps7`.
- Modelo con variable respuesta `ihar30` y como variables explicativas `s1`, `s7.a` y `ps7`.

Ambos modelos tienen variables con significación estadística. Se compararon ambos modelos utilizando

AIC

GLM.11 4 187.2057

GLM.21 4 154.5465

A la hora de comparar AIC de varios modelos siempre nos debemos quedar con el que tenga menos AIC, es decir el segundo modelo y se añadió al conjunto de datos. El AIC mide lo lejos que está el modelo de la realidad por lo que cuanto menor es el valor, mejor es el modelo.

También se trabajó con otro modelo en donde se utilizó con la variable `ihar15` como variable respuesta

y las variables `sb + ps2 + ps3 + ps6 + ps7 + ps8` como explicativas y mediante la función *stepwise* se escogieron las variables `sb` (STOP-Bang) y `ps7` (% del tiempo de sueño por debajo de saturación de O_2 del 90%) obteniéndose el siguiente modelo:

Call:

```
glm(formula = ihar15 ~ sb + ps7, family = binomial(logit), data = ap)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-4.0930	-0.7461	-0.4552	0.8365	2.1338

Coefficients:

Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.21854	0.64063	-5.024 5.06e-07 ***
sb	0.50174	0.12714	3.946 7.94e-05 ***
ps7	0.09361	0.02482	3.772 0.000162 ***

Signif. codes:

0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 241.68 on 177 degrees of freedom

Residual deviance: 180.89 on 175 degrees of freedom

AIC: 186.89

Number of Fisher Scoring iterations: 5

(Intercept)	sb	ps7
0.04001341	1.65159902	1.09813462

Las dos variables explicativas tienen significación estadística. La variable respuesta *ihar15* es una variable binaria. Si el *ihar15* es mayor de 15 entonces tenemos un valor de 1 y si es menor de 15 de 0,

por tanto, nos va a indicar de manera clara y definitiva si una persona tiene (valor 1) o no (valor 0) apnea de sueño.

4.3.1. Interpretación de resultados con Odds Ratio

Se suelen interpretar los resultados obtenidos por los modelos de regresión logística en términos de la Odds Ratio (OR) para un determinado valor de referencia de un hipotético factor de riesgo. Se considera que la variable respuesta *ihar15* indica la presencia o ausencia de síndrome de apnea hipoapnea del sueño. Se fija un valor de antemano de referencia f_0 , asociado a un valor de referencia con riesgo bajo, y se comprueba como varía el riesgo de la enfermedad a medida que el nivel de exposición F se desvía de dicho valor. Dado un valor f , se trata de obtener el valor de al OR

$$OR_{f_0}(f) = \frac{odds_X(f)}{odds_X(f_0)} = \frac{\pi(X, f)/(1 - \pi(X, f))}{\pi(X, f_0)/(1 - \pi(X, f_0))}$$

La expresión anterior se puede simplificar de la siguiente forma: $OR_{f_0}(f) = e^{b(f-f_0)}$

Las OR basadas en modelos logísticos se pueden obtener con la función `logistic.display` de la librería `epiDisplay` en **R**. Estos serían los resultados en el modelo:

```
{\tiny Logistic regression predicting ihar15 : 1 vs 0
```

```
crude OR(95%CI)   adj. OR(95%CI)
sb (cont. var.)  1.81 (1.45,2.27)  1.65 (1.29,2.12)
ps7 (cont. var.) 1.12 (1.07,1.19)  1.1 (1.05,1.15)
```

```
P(Wald's test) P(LR-test)
sb (cont. var.) < 0.001      < 0.001
ps7 (cont. var.) < 0.001      < 0.001
```

```
Log-likelihood = -90.4439
No. of observations = 178
AIC value = 186.8879}
```

4.4. Modelos GAM

Un Modelo Aditivo Generalizado (GAM) es un modelo lineal generalizado en el cual el predictor lineal depende linealmente de funciones suaves desconocidas de algunas variables predictoras, y el interés se centra en la inferencia sobre estas funciones suaves. Hastie e Tibshirani (1990) proponen un modelo en el cual no es necesario suponer linealidad, utilizando modelo aditivos generalizados (GAM). Se pueden considerar los modelos GAM como una extensión de los modelos GLM, debido a que aportan un matiz de flexibilidad que no permitían los modelos GLM.

Este tipo de modelo suponen la estructura:

$$\text{logit}(X) = a_0 + f_1(X_1) + \dots + f_p(X_p)$$

de forma que las constante a_j de los GLM (asociadas a las covariables X_j) son suplantadas por funciones f_j . Bajo este modelo, el efecto de cada covariable no sigue una forma paramétrica fija (como en los GLM), si no que depende de una función totalmente desconocida a la que únicamente se le exige cierto grado de suavidad.

En R para el cálculo de modelo aditivos (generalizados) se utiliza el paquete `mgcv` empleando regresión por splines. Permite crear varios tipos de modelos como

- Modelo lineal. En este caso el ajuste sería `gam(y ~ x1 + x2 + x3)`
- Modelo aditivo con efectos no paramétricos para x_1 y x_2 , y un efecto lineal para x_3 . En este caso el ajuste sería `gam(y ~ s(x1) + s(x2) + x3)`
- Modelo no aditivo (con interacción). En este caso el ajuste sería `gam(y ~ s(x1, x2))`
- Modelo con distintas combinaciones. El ajuste sería `gam(y ~ s(x1, x2) + s(x3) + x4)`

Primero definimos las variables a utilizar y se crearon varios modelos. Con los resultados obtenidos se eligieron dos y se creó el siguiente modelo GAM logístico:

$$\text{logit}(s1, ps7) = a_0 + a_1s1 + s(ps7, \text{según sexo})$$

Al final se utilizaron las variables con significación estadística y se distinguió la variable `ps7` según el sexo dando lugar a estos resultados:

Formula:

```
ihar15 ~ s1 + s(ps7, by = sexo)
```

Parametric coefficients:

Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.5429	0.9692	-1.592	0.1114
s11	1.8933	0.9039	2.094	0.0362 *

Approximate significance of smooth terms:

edf	Ref.df	Chi.sq	p-value	
s(ps7):sexo1	3.553	4.251	16.11	0.00376 **
s(ps7):sexo2	5.792	6.631	29.37	8.07e-05 ***

R-sq.(adj) = 0.479 Deviance explained = 44.1%

UBRE = -0.11412 Scale est. = 1 n = 178

Existe significación estadística para la variable $s1$ así como para la variable $ps7$ distinguida por sexo. Se obtuvo un resultado que se muestra en la Fig. 4.1 en donde se ve también la estimación para un intervalo de confianza del 95 %.

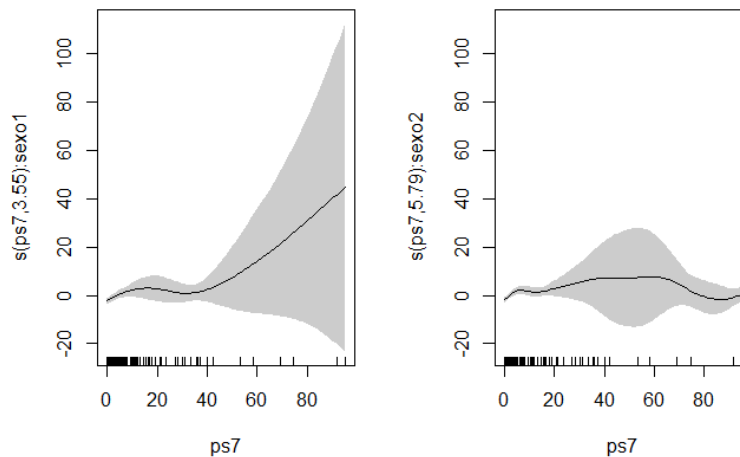


Figura 4.1: Resultado modelo GAM por sexos, mujeres izquierda, hombres derecha

Bajo el modelo GAM se deducen varios hechos interesantes.

- Existe una clara distinción entre los resultados de mujeres (sexo1) y de hombres (sexo2). Los intervalos de confianza son muy grandes en el grupo de mujeres cuando tenemos pocos datos (parte derecha) pero apenas cambia al inicio, que es donde está la mayor parte de los datos del estudio.
- El efecto en los hombres es 'aproximadamente' lineal. El intervalo de confianza es mayor en los datos centrales con apenas cambio en los datos iniciales.

El Fig. 4.1 podría cambiarse la escala de los ejes para un mejor encaja de los resultados obtenidos y así encontrar un ajuste más linealizado.

Capítulo 5

Resultados

De las 203 personas reclutadas para hacer el estudio y tras revisión detallada de los mismos, se excluyeron 25 casos debido a dos motivos, uno, el registro realizado durante la **PSG** no ofreció garantías para su análisis o bien por cumplir criterios de exclusión. Al final el estudio se redujo a **178** sujetos, de los cuales **57** eran mujeres y **121** eran hombres (Fig. 5.1), con edades comprendidas entre los 18 años y los 77 años de edad. . La media de edad fue de **50.34** años y con una desviación estándar de **13.32**.

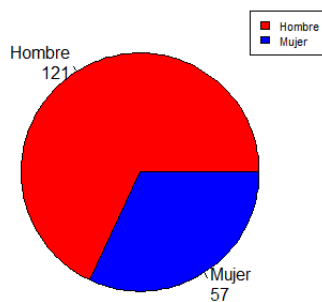


Figura 5.1: Distribución por sexos

El histograma por edad puede verse en la Fig. 5.2.

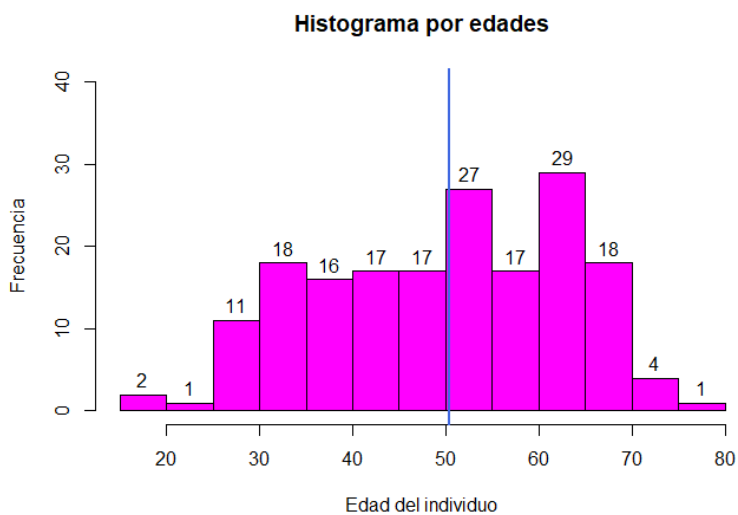


Figura 5.2: Histograma

Se incluyeron 104 (58.63%) personas con una IHA ≥ 5 , 74 (41.57%) con un IHA ≥ 15 y 54 (30.34%) con un IHA ≥ 30 . Ninguna de las variables cuantitativas estudiadas seguía una distribución normal. En el análisis bivariante, la diferencia fue significativa en todas las variables cuantitativas y en la presencia de problemas cardiovasculares, ronquido, HTA, IMC superior a 30, edad superior a 50 años y circunferencia del cuello superior a 40.

Al analizar la utilidad diagnóstica del IHA comparado con el diagnóstico clínico, su AUC es de 0.996 (0.990-1.003), presenta un punto de corte óptimo igual a 5, con sensibilidad de 0.991 (0.949-1.000), una especificidad de 0.971 (0.901-0.997), VPP de 0.982 (0.934-1.000), VPN= 0.990 (0.923-0.998), RV+ de 34.676 (8.846-135.929), RV- de 0.010 (0.001-0.067), 2 falsos positivos y 1 falso negativo.

En el análisis multivariante se utilizó el modelo que tenía un menor AIC que incluía como variables independientes responder sí a "¿ronca usted fuerte?" la SP90 por sexo obteniéndose el resultado de esta tabla con un p-valor menor de 0.036:

Familia: binomial

Link function: logit

Formula:

$\text{ihar15} \sim \text{¿ronca fuerte?} + \text{s(ps7, por = sexo)}$

Parametric coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept)	-1.5429	0.9692	-1.592	0.1114
s11	1.8933	0.9039	2.094	0.0362 *

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

Approximate significance of smooth terms:

edf Ref.df Chi.sq p-value

s(ps7):sexo1	3.553	4.251	16.11	0.00376 **
s(ps7):sexo2	5.792	6.631	29.37	8.07e-05 ***

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

R-sq.(adj) = 0.479 Deviance explained = 44.1%

UBRE = -0.11412 Scale est. = 1 n = 178

Las AUC para STOP-Bang, Sp90 y la combinación de roncar fuerte con Sp90 por sexo dio lugar a un valor de 0.737 (0.667, 0.808), 0.883 (0.831, 0.936) y 0.903 (0.858, 0.949) respectivamente y se muestran en la Fig. 5.3. Se compararon por tanto 3 curvas ROC. Para ello se utilizó la **probabilidad obtenida mediante el modelo logístico como prueba diagnóstica**. Una utilizando la variable *ps7*, otra utilizando la variable *sb* (resultado del STOP-Bang) y otra combinando ambas y se obtuvo el resultado que puede verse en la Fig. 5.3. Se estudiaron los comportamiento de varias curvas entre sí dos a dos resultando ser significativo la diferencia entre la curva *ps7* y *sb*, no habiendo diferencias entre la curva combinada y *ps7*, y encontrando diferencias significativas entre la curva combinada y *sb*. Las AUC para Sp90 y para el modelo GAM seleccionado no presentan diferencias significativas entre ellas, pero sí con la AUC del STOP-Bang. Los resultados con el punto de corte óptimo por sexo se presentan en la Tabla 3. Se utilizó el índice de Youden como criterio para calcular los puntos de corte óptimos. Las limitaciones de la pulsioximetría se fundamentan en que son estudios no vigilados y/o colocados por el paciente, lo que origina pérdida de datos o que no son técnicamente válidos, obligando a su repetición. Por ello es de suma importancia el entrenamiento de los pacientes en la colocación de los

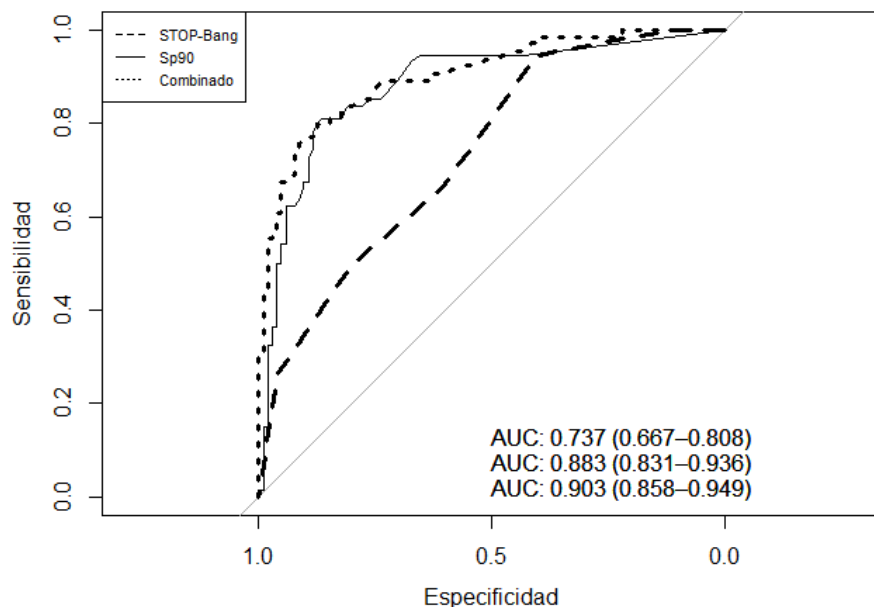


Figura 5.3: Comparación de 3 curvas ROC

dispositivos. Pero no nos permitió analizar la concordancia del pulsioxímetro en domicilio con el de la PSG. Esta información, raramente facilitada, sería de alto interés para la selección y adquisición de pulsioxímetros más allá de comparar las señales en estudios de laboratorio, para diferentes condiciones clínicas, y con pequeñas muestras (19).

Otra de las posibles limitaciones de este tipo de estudios podría deberse a una mayor tasa de participación en los sujetos con trastornos durante el sueño, que aumentaría la prevalencia pretest. En nuestra serie, el número de pacientes sanos fue inferior al de enfermos por las dificultades para encontrar voluntarios sanos (a pesar de la colaboración de los propios investigadores, sus familiares y médicos de familia en formación). También fue escaso el número de mujeres participantes, pero nuestros resultados señalan claramente la pertinencia del análisis por género en esta patología.

Entre los cuestionarios más utilizados para SAHS en pacientes comunitarios, Jonas et al destacan el cuestionario Berlin y el MVAP. El Berlin, para una IHA ≥ 15 presentó una sensibilidad de 43,0% (41,2-44,8), muy inferior a las obtenidas en este estudio. El MAVP se validó en pacientes con hipertensión y en pacientes con somnolencia diurna.

El análisis individual de la IHA comparándola con el diagnóstico clínico permite por tanto cuantificar su utilidad diagnóstica; esto, unido al elevado desempeño de la Sp90 y/o su combinación con variables clínicas, conducen a concluir que puede ser utilizado para el seguimiento de los pacientes por médicos de familia/enfermeras con técnicas sencillas. Sería un método de bajo coste para seleccionar los pacientes a derivar puesto que el precio de un pulsioxímetro con registro nocturno es de 400 mientras que un polígrafo respiratorio son 1.800 euros; y la tarifa de un estudio del sueño en el ámbito hospitalario es 1.200 euros en Galicia.

Capítulo 6

Conclusión

Aunque existen publicaciones relativas a la validación de PSG II, III y IV en domicilio del cliente, no nos consta que se hayan presentado previamente el uso de la PSG I en atención primaria. Con ello, se pretende iniciar una línea de investigación en este ámbito, que se podría orientar a patologías de importante prevalencia como la obesidad, hipertensión, fatiga en mujeres, etc..

Existe muchos trabajos de investigación que hablan de las diferencias entre mujeres y hombres con respecto a la apnea del sueño pero son pocos los estudios que cuantifiquen estas diferencias. No se han identificado puntos de corte específicos por sexo. Con los resultados obtenidos en este trabajo, el cuestionario STOP-Bang permitiría seleccionar los pacientes que se debieran referir a otro ámbito, acompañado de una entrevista clínica estructurada y en el marco de una atención integrada a estos pacientes. En los hombres se podría buscar un punto de corte que maximice sensibilidad a expensas de la especificidad, con el acuerdo de los servicios de referencia correspondientes.

Como conclusión, el uso de este tipo de procedimientos para investigar la patología en atención primaria puede resultar muy útil y ahorrar dinero al no realizar costosas pruebas diagnósticas. Y con la utilización de métodos estadísticos relativamente simples, se puede ayudar a diferenciar por sexo la actuación en la apnea de sueño, patología de prevalencia creciente y con grandes dificultades en su gestión.

Un aspecto a destacar de este trabajo son las grandes diferencias al analizar por sexo en los índices de utilidad diagnóstica y en los puntos de corte óptimos de los test, con las repercusiones que pueda tener en la práctica clínica diaria en Atención Primaria.

6.1. Líneas futuras

Como ya se ha comentado, este trabajo puede ser el punto de partida de nuevas líneas de investigación en un futuro cercano en el cual se trabajen con otras patologías como puede ser la obesidad, la hipertensión resistente, fatiga en mujeres, gestantes, etc. El poder enfrentar una variable respuesta binaria con variables de fácil manejo para un médico de familia repercutirá en una optimización del tiempo para la toma de decisiones.

Si bien existen numerosas preguntas de investigación con poca o insuficiente evidencia para el tratamiento de la **síndrome de Apnea Hipoapnea del sueño**, los pacientes con esta enfermedad, sus profesionales de atención médica y la sociedad en general se beneficiarían de re enfocar los esfuerzos de investigación en las preguntas de investigación priorizadas y de simples comparaciones de resultados a corto plazo para poder actuar de forma adecuada y realizar intervenciones específicas si son necesarias. La creciente prevalencia del Síndrome de Apnea Hipoapnea del sueño y la creciente evidencia de su asociación con enfermedades cardiometabólicas exige más investigación para abordar el verdadero impacto de este trastorno en un futuro.

Es interesante destacar que se podría plantear el hacer un seguimiento para aquellos pacientes que están en el borde de sufrir este trastorno. Por ejemplo, en el grupo de trabajo se encontraron varios pacientes que sin haber llegado a uno de los criterios clínicos de los 50 años estaban a pocos años de llegar a él. Por tanto, si se repite el cuestionario STOP-Bang en x años es probable que ese individuo entre a formar parte de los que sí van a padecer apnea.

También sería interesante plantear un modelo similar al STOP-Bang para el estudio de cuadros clínicos de pacientes que tengan síntomas de la enfermedad por coronavirus (2019) (COVID-19). La COVID 19 es una enfermedad infecciosa causada por un coronavirus (*SARS-COV-2*) recientemente descubierta. El triaje podría ser realizarles una serie de preguntas clarificadoras y según el valor de la prueba podría ser necesario derivar al paciente a hacer una prueba confirmatoria o bien enviarlo para casa con medidas higiénicas y confinamiento. El facultativo haría preguntas sobre:

- Tener fiebre.
- Tos seca.
- Cansancio.
- Molestias y dolores.

- Dolor de garganta.
- Diarrea.
- Conjuntivitis.
- Dolor de cabeza.
- Pérdida del sentido del olfato o del gusto.
- Erupciones cutáneas o pérdida del color en los dedos de las manos o de los pies.

Y a partir de aquí se podría tener un diagnóstico inicial, no confirmado de tener la COVID-19. Posteriormente si se tiene la sospecha a través del test realizado que el individuo es posible positivo se tendría que realizar una reacción en cadena de la polimerasa (PCR) para confirmar que el paciente efectivamente tiene el virus en su cuerpo.

Éste enfoque que planteo sería para individuos que presentan síntomas de la enfermedad puesto que las preguntas dan como resultado un número y según ese número podríamos derivar al individuo a un centro médico a hacer la PCR o no.

Por otra banda, podríamos también plantear un modelo para individuos asintomáticos. En este caso se debería hacer un estudio previo con cientos de variables para poder encontrar una relación entre ser asintomáticos y transmitir el virus. Posiblemente existe una razón genética para que un individuo no presente síntomas, teniendo en su cuerpo el virus. La utilización de marcadores genéticos sería muy interesante llevarla a cabo. Los marcadores genéticos son segmentos de ácido desoxirribonucleico (ADN) físicamente localizado en un locus (en una posición de un cromosoma de los 23 pares de cromosomas que tiene la especie humana) y cuya herencia genética se puede rastrear. Obtener una población asintomática y hacer un estudio de varios marcadores genéticos podría darnos una pista para saber porque esta enfermedad afecta a unos pacientes más que a otros, a unas razas más que a otras.

Apéndice A

6.2. Software utilizado

Los análisis se llevaron a cabo en R, un entorno y lenguaje de programación para estadística. Se emplearon varios paquetes entre los que destacaría:

- **pROC**. Permite mostrar y analizar curvas ROC.
- **OptimalCutpoints**. Permite calcular puntos de puntos para test diagnósticos o marcadores continuos.
- **Paquete epiDisplay**. Permite calcular los OR de modelos logísticos.
- **mgcv**. Proporciona funciones para modelos aditivos generalizados (tanto gam como bam) y modelos aditivos generalizados mezclas.
- **ROCR**. Paquete que estima medias de rendimiento y graficar estas medidas en un rango de puntos de corte.
- **bdpv**. Cálculo de intervalos de confianza asintóticos para valores predictivos negativos y positivos en pruebas de diagnóstico binarias.
- **MASS**. Permite el cálculo mediante análisis discriminante utilizando la función *lda*.
- **RcmdrMisc**. Se utilizaron varias funciones estadísticas, gráficas y de gestión de datos.

También se trabajó con **PSPP**. PSPP es una aplicación de software libre para el análisis de datos. Se presenta en modo gráfico y está escrita en el lenguaje de programación C. Es un reemplazo libre para el software propietario SPSS.

Para la creación de bibliografía se utilizó **Mendeley**. Mendeley es una aplicación web y de escritorio, propietaria y gratuita. Permite gestionar y compartir referencias bibliográficas y documentos de investigación, encontrar nuevas referencias y documentos y colaborar en línea.

Indice de figuras

Índice de figuras

1.1. Fisiopatología SAHS	2
1.2. Cuestionario STOP-BANG	4
3.1. Pruebas diagnósticas	13
3.2. Ejemplo de curva ROC	16
3.3. Boxplot variables utilizadas y curva ROC	18
3.4. Esquema k-fold cross validation, con K=10 y un solo clasificador	19
3.5. Validacion cruzada utilizando método esquina superior izquierda (closest.topleft)	20
3.6. Validacion cruzada utilizando método AUC	21
3.7. Curva ROC para Mujeres	23
3.8. Curva ROC para Hombres	24
4.1. Resultado modelo GAM por sexos, mujeres izquierda, hombres derecha	34
5.1. Distribución por sexos	37
5.2. Histograma	38
5.3. Comparación de 3 curvas ROC	40

Bibliografía

- [1] Johns, M. W. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep* 14, 540-545 (1991).
- [2] Netzer NC, Stoohs RA, Netzer CM, et al. Using the Berlin Questionnaire to identify patients at risk for the sleep apnea syndrome. *Ann Intern Med.* 1999 Oct 5;131(7):485-91. Epub: 1999/10/03. PMID: 10507956. https://en.wikibooks.org/wiki/LaTeX/Manually_Managing_References.
- [3] Chung F, Yegneswaran B, Liao P, et al. STOP questionnaire: a tool to screen patients for obstructive sleep apnea. *Anesthesiology.* 2008 May;108(5):812-21. PMID: 18431116.
- [4] Chung F. The Official STOP-Bang Questionnaire Website. Toronto Western Hospital, U. H. N., and University of Toronto. Acceso: <http://www.stopbang.ca/> (2012).
- [5] Lloberes P, Durán-Cantolla J, Martínez-García MÁ, et al. Diagnóstico y tratamiento del síndrome de apneas-hipopneas del sueño. *Arch Bronconeumol* (2011);47:143-56.
- [6] Epstein, J. L, Sanders, M. G., Simon, B. S., Salinas, K. C, Jansorn,N. R., and Van Voorhis, F. L. (2009). *School, family, andcommunity partnerships: Your handbook for action.* Thousand Oaks, CA: Corwin Press.
- [7] Rechtschaffen, A. y Kales, A. (1968) *A manual of standardised terminology, techniques and scoring system of sleep stages of human subjects.* Public Health Service, Government Printing Office (Washington).
- [8] Youden, W. J. (1950). Index for Rating Diagnostic Tests. *Cancer*, 3(1), pp. 325.
- [9] Efron Bradley, Tibshirani Robert. *An introduction to the Bootstrap.* Chapman & Hall. (1993).

- [10] Jonas D, Amick H, Feltner C, Palmieri R, Arvantis. Screening for Obstructive Sleep Apnea in Adults: An Evidence Review for the U.S. Preventive Services Task Force. AHRC. 2017;No 14-0521(121).
- [11] González-Mangado N, Egea Santaolalla C, Chiner Vives E, Saiz A, de Lucas-Ramos P. Síndrome de apnea-hipopnea del sueño. Arch Bronconeumol Monográfico. 2015;1?21.
- [12] McCullagh P, Nelder JA. Generalized Linear Models. London: Chapman and Hall, 1989.
- [13] Hastie, T.J., Tibshirani, 1990. Generalized Additive Models. Chapman and Hall, London.

Más bibliografía utilizada en el trabajo:

1. Barbé F, Pericás J, Muñoz A, et al. Automobile accidents in patients with sleep apnea syndrome. An epidemiological and mechanistic study. Am J Respir Crit Care Med 1998;158:18-22.
2. Bibbins-Domingo K, Grossman DC, Curry SJ, et al. Screening for Obstructive Sleep Apnea in Adults: US Preventive Services Task Force Recommendation Statement. JAMA 2017;317:407.
3. Cadarso-Suárez C, Roca-Pardiñas J y Figueiras A (2006). Effect Measures in Nonparametric Regression with Interactions between Continuous Exposures. Statistics in Medicine 25 : 603-621.
4. Efron Bradley, Tibshirani Robert. An introduction to the Bootstrap. Chapman & Hall. 1993.
5. Efron, B., 1979. Bootstrap methods: another look at the jackknife. The Annals of Statistics 7, 1-26.
6. Elisabeth R. DeLong, David M. DeLong and Daniel L. Clarke-Pearson (1988) - Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44, 837- 845.
7. Epton MJ, Kelly PT, Shand BI, et al. Development and outcomes of a primary care-based sleep assessment service in Canterbury, New Zealand. NPJ Prim care Respir Med 2017;27:26.
8. E. S. Venkatraman (2000) - A Permutation Test to Compare Receiver Operating Characteristic Curves. Biometrics 56, 1134-1138. DOI: 10.1111/j.0006-341X.2000.01134.x.
9. Fawcett, T. (2005). An introduction to ROC analysis. Irbm, 35(6), 299-309. <https://doi.org/10.1016/j.patrec.2005.10.010>
10. González-Mangado N., Egea Santaolalla C., Chiner Vives E., Saiz A., de Lucas-Ramos P. Síndrome de apnea-hipopnea del sueño. Arch Bronconeumol Monográfico 2015:1-21.
11. Hadley Wickham (2011) - The Split-Apply-Combine Strategy for Data Analysis. Journal of Statistical Software, 40, 1-29. URL: www.jstatsoft.org/v40/i01.

11. Hand, D. J. (2009). Measuring classifier performance: A coherent alternative to the area under the ROC curve. *Machine Learning*, 77(1), 103-123. <https://doi.org/10.1007/s10994-009-5119-5>.
12. Hastie, T.J., Tibshirani, 1990. *Generalized Additive Models*. Chapman and Hall, London
13. James A. Hanley and Barbara J. McNeil (1982) - The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143, 29-36.
14. Jonas DE, Amick HR, Feltner C, et al. Screening for Obstructive Sleep Apnea in Adults. *JAMA* 2017;317:415.
15. López-Ratón M, Rodríguez-Álvarez MX, Suárez CC, et al. OptimalCutpoints: An R Package for Selecting Optimal Cutpoints in Diagnostic Tests. *J Stat Softw* 2014;61:1-36.
16. Lloberes P, Durán-Cantolla J, Martínez-García MÁ, et al. Diagnóstico y tratamiento del síndrome de apneas-hipopneas del sueño. *Arch Bronconeumol* 2011;47:143-56.
17. McCullagh P, Nelder JA. *Generalized Linear Models*. London: Chapman and Hall, 1989.
18. Roca-Pardiñas J, Cadarso-Suárez C, Nacher V, Acuña C. Bootstrap-based methods for testing factor-by-curve interactions in Generalized Additive Models: assessing prefrontal cortex neural activity related to decision-making. *Statistics in Medicine* 2006; 25: 2483-2501. (DOI: 10.1002/sim.2415).
19. Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., and Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12(1), 77. <https://doi.org/10.1186/1471-2105-12-77>
20. Sánchez Armengol A, Carmona Bernal C, Ruiz García A. Polisomnografía, poligrafía, oximetría. Requisitos e interpretación de resultados. En: Soto Campos JG, ed. *Polisomnografía, poligrafía, oximetría. Requisitos e interpretación de resultados*. 2006:199-210.
21. Sateia MJ. International Classification of Sleep Disorders-Third Edition. *Chest* 2014;146:1387-94.
22. Terán-Santos J, Jiménez-Gómez A, Cordero-Guevara J. The association between sleep apnea and the risk of traffic accidents. Cooperative Group Burgos-Santander. *N Engl J Med* 1999;340:847-51.
23. Wood SN. Fast stable restricted maximum likelihood and marginal likelihood estimation of semi-parametric generalized linear models. *Statistical Methodol* 2011;73:3-36.
24. Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R*. Chapman & Hall/CRC.
25. Xavier Robin, Natacha Turck, Alexandre Hainard, et al. (2011) pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 7, 77. DOI: 10.1186/1471-2105-12-77.
26. Xu Sun and Weichao Xu (2014) - Fast Implementation of DeLong's Algorithm for Comparing the Areas Under Correlated Receiver Operating Characteristic Curves. *IEEE Signal Processing Letters*,

21, 1389-1393. DOI: 10.1109/LSP.2014.2337313.

27. Youden, W. J. (1950). Index for Rating Diagnostic Tests. *Cancer*, 3(1), pp. 325.

Listado de acrónimos

AUC area under a curve

COVID-19 enfermedad por coronavirus (2019)

GAM Modelo Aditivo Generalizado

GLM Modelos Lineales Generalizados

IAH índice de apnea-hipoapnea

PSG polisomnografía

ROC receiver operating characteristics

SAHS síndrome de apneas-hipoapneas del sueño