

UNIVERSIDAD MIGUEL HERNÁNDEZ
FACULTAD DE CIENCIAS SOCIALES Y JURÍDICAS
TRABAJO FIN DE GRADO

GRADO EN ESTADÍSTICA EMPRESARIAL



UNIVERSITAS
Miguel Hernández

**DOLOR MUSCULoesquelético EN CIRUJANOS ESPAÑOLES: ANÁLISIS
ESTADÍSTICO Y DESARROLLO DE UNA HERRAMIENTA PREDICTIVA DE
RIESGO**

AUTOR: Iván Mena Martínez-Santos

TUTOR: Xavier Barber

Departamento de Estadística, Matemáticas e Informática

Curso académico 2024/2025

Convocatoria Junio

Resumen

Este Trabajo Fin de Grado se centra en el estudio del dolor musculoesquelético en cirujanos españoles, partiendo de la implementación de la encuesta «ErgoSurgeon» entre miembros de la Asociación Española de Cirujanos. En la introducción se revisa la literatura sobre ergonomía quirúrgica y los mecanismos que explican la aparición de molestias en cuello, espalda y extremidades durante la actividad en el quirófano. A continuación, se detalla el diseño muestral y la construcción del cuestionario, que recoge información demográfica, características de la práctica quirúrgica (tanto frecuencia y tipo de intervenciones) y aspectos ergonómicos como la disposición de pantallas, el uso de instrumental y la realización de pausas intraoperatorias.

La metodología combina un análisis descriptivo de las variables recogidas, pruebas estadísticas inferenciales para explorar las relaciones entre factores de riesgo y presencia de dolor, y técnicas de modelización predictiva mediante regresión logística y algoritmos de Random Forest. En la parte aplicada se describe el desarrollo de una calculadora interactiva en R Shiny, basada en los resultados de los modelos, que permite estimar el riesgo individual de padecer dolor moderado o severo en distintas regiones corporales. Finalmente, se discuten las implicaciones prácticas de estos hallazgos para la formación ergonómica, el diseño de protocolos preventivos y la mejora del entorno quirúrgico con miras a preservar la salud laboral de los cirujanos.

Palabras clave: dolor musculoesquelético, regresión logística, Random Forest, modelos predictivos, aprendizaje automático.

Abstract

This Bachelor's Thesis investigates musculoskeletal pain among Spanish surgeons, grounded in the «ErgoSurgeon» survey administered to members of the Spanish Association of Surgeons. The introduction reviews the literature on surgical ergonomics and the mechanisms underlying discomfort in the neck, back, and extremities during operating room procedures. The sampling design and questionnaire development are then detailed, covering demographic characteristics, surgical practice parameters (frequency and type of procedures), and ergonomic factors such as monitor placement, instrument ergonomics, and intraoperative breaks.

The methodology integrates a descriptive analysis of the collected variables, inferential statistical tests to examine associations between risk factors and pain occurrence, and predictive modeling using logistic regression and Random Forest algorithms. In the application section, the development of an interactive R Shiny calculator based on the predictive models is described, enabling users to estimate their individual risk of moderate to severe pain in various body regions. Finally, the practical implications of these findings are discussed in terms of ergonomic training, preventive protocol design, and operating room improvements aimed at preserving surgeons' occupational health.

Keywords: musculoskeletal pain, logistic regression, Random Forest, ergonomic, predictive modeling, machine learning.

Índice

1. Introducción y contexto del estudio	7
1.1. Introducción al problema del dolor y lesiones musculoesqueléticas en cirujanos/as españoles	7
1.2. Impacto en la salud y el desempeño laboral	8
1.3. Estudios previos y relevancia del estudio	9
1.4. Objetivos del estudio	10
1.5. Metodología de la Encuesta ErgoSurgeon	11
2. Modelos estadísticos y aprendizaje automático.....	13
2.1. Introducción a los modelos de regresión logística	13
2.2. Odds Ratio e interpretación de coeficientes.....	18
2.3. Bondad de ajuste.....	20
2.4. Uso del modelo de Random Forest en la clasificación del riesgo	26
2.5. Importancia de las variables en la predicción del riesgo	28
3. Análisis descriptivo e inferencial.....	30
3.1. Descriptivos	32
3.2. Distribución de los factores de riesgo y prevalencia de dolor.	36
3.2.1. Cuello	37
3.2.2. Hombros.....	39
3.2.3. Zona dorsal.....	41
3.2.4. Lumbares	43
3.2.5. Muñeca	45
3.2.6. Manos y dedos	47
3.2.7. Rodillas	49
3.2.8. Tobillos y pies	51
4. Desarrollo de la calculadora de riesgos.....	53
4.1. Fundamentos teóricos para la estimación del riesgo	53
4.2. Implementación y visualización de la calculadora en Shiny	54
5. Conclusiones y recomendaciones	56
6. Bibliografía	58
7. Anexos.....	60

Índice Figuras

Figura 1. Esquema general del funcionamiento de un modelo de regresión logística (Sotaquirá, 2018)	13
Figura 2. Función logística	15
Figura 3. Transformación del espacio probabilístico al de odds	19
Figura 4. Ejemplo de curva ROC y valor AUC	26
Figura 5. Algoritmo Random Forest	27
Figura 6. Calculadora de prevalencia de dolores musculoesqueléticos	54
Figura 7. Resultados calculadora de prevalencia de dolores	55

Índice Tablas

Tabla 1. Tabla de clasificación.....	25
Tabla 2. Ventajas y desventajas Random Forest	28
Tabla 3. Diccionario de variables	32
Tabla 4. Características de la muestra según el Género.....	33
Tabla 5. Características según Dolor en los Últimos 12 Meses.....	34
Tabla 6. Prevalencia de Dolor Musculoesquelético por Zona y Género	35
Tabla 7. Factores asociados con dolor cervical en cirujanos	37
Tabla 8. Matriz de confusión y métricas en zona cervical.....	38
Tabla 9. Factores asociados con dolor de hombros en cirujanos	39
Tabla 10. Matriz de confusión y métricas en hombros	40
Tabla 11. Factores asociados con dolor en zona dorsal en cirujanos	41
Tabla 12. Matriz de confusión y métricas en zona dorsal.....	42
Tabla 13. Factores asociados con dolor en zona lumbar en cirujanos	43
Tabla 14. Matriz de confusión y métricas en zona lumbar	44
Tabla 15. Factores asociados con dolor en muñecas en cirujanos	45
Tabla 16. Matriz de confusión y métricas en muñecas	46
Tabla 17. Factores asociados con dolor en manos y dedos en cirujanos	47
Tabla 18. Matriz de confusión y métricas en manos y dedos	48
Tabla 19. Factores asociados con dolor en rodillas en cirujanos	49
Tabla 20. Matriz de confusión y métricas en rodillas	50
Tabla 21. Factores asociados con dolor en tobillos y pies en cirujanos.....	51
Tabla 22. Matriz de confusión y métricas en tobillos y pies.....	52

1. Introducción y contexto del estudio

1.1. Introducción al problema del dolor y lesiones musculoesqueléticas en cirujanos/as españoles

El dolor musculoesquelético es uno de los principales síntomas relacionados con el trabajo de los cirujanos, afectando negativamente a su calidad de vida y rendimiento laboral.

Una de las principales consecuencias de los riesgos ergonómicos a los que están expuestos los cirujanos es el dolor musculoesquelético (DME). La OMS hace referencia sobre las condiciones musculoesqueléticas, las cuales son caracterizadas por dolor, limitaciones en la movilidad, destreza y en las capacidades funcionales, por ello con impacto sobre las habilidades tanto físicas y mentales de las personas, afectando el proceso, la eficiencia y productividad laboral, la calidad de vida y capacidad económica.

En las enfermedades musculoesqueléticas predomina el dolor como síntoma y consecuentemente una cierta alteración funcional. Puede afectar a cualquier parte del cuerpo y su gravedad va desde la fatiga postural reversible hasta afecciones periarticulares irreversibles. En una primera fase se producen síntomas de forma ocasional, para más tarde, instaurarse de forma permanente y crónica. Normalmente, no surgen por golpes fuertes, sino por usar demasiado ciertas partes.

Cuando se opera, tener DME es un asunto más común, por el tipo de labor que se realiza siempre. Quedarse en posiciones raras por mucho rato, usar muy seguido las herramientas de operación, hace más probable que a los cirujanos les duela el cuerpo en áreas como cuello, espalda, hombros y brazos. Una gran cantidad de estudios afirman que del 80 al 90 % de los cirujanos han sentido dolor en los huesos alguna vez. (Sánchez-Guillén et al., 2024).

Tener dolores en los huesos y músculos de los cirujanos no es solo un problema personal, también afecta cómo funciona el hospital y cómo se cuida a la gente. Si los cirujanos no acuden a trabajar por estos dolores, hay menos personal para realizar las operaciones. Esto hace que los demás tengan más trabajo y que el hospital funcione peor. Además, curar estos dolores cuesta mucho dinero, ya sea con tratamientos fisiológicos, medicinas o cirugías. Esto es un problema para los hospitales y para el sistema de salud en general. En países como Estados Unidos y Canadá, estos dolores son una de las mayores causas de enfermedad en el trabajo en el sector de la salud. Esto causa pérdidas de mucho dinero cada año. En España, no se ha investigado mucho sobre esto en cirujanos, pero otros estudios sugieren que pasa algo parecido.

Este estudio busca ver cuánto sufren los cirujanos españoles por dolores en sus huesos y músculos. Se creará algo así como una calculadora para predecir quién podría tener estos problemas. La idea es que todos se cuiden más, para que los cirujanos estén bien y los pacientes reciban una mejor atención.

1.2. Impacto en la salud y el desempeño laboral

Los dolores musculoesqueléticos en el ámbito quirúrgico no tienen por qué ser únicamente una molestia puntual o un efecto secundario del trabajo físico. En realidad, es un problema de salud laboral con consecuencias directas y acumulativas, por ejemplo, la intensidad del dolor, su duración, la localización de las molestias y la frecuencia con la que aparecen afectan tanto al bienestar físico como mental de los cirujanos, produciendo así una perdida progresiva de su funcionalidad y rendimiento en el entorno hospitalario.

Desde el punto de vista de la salud física, el DME afecta generalmente a la zona del cuello, hombros, zona dorsal, lumbares, muñecas, rodillas y pies. Cuando estos dolores no se identifican pronto o no se aplican medidas correctivas pueden derivar a dolores crónicos, disminución de la movilidad y aparición de trastornos más severos como por ejemplo contracturas musculares. Estos dolores afectan tanto a la capacidad de llevar a cabo procedimientos quirúrgicos como en la vida personal dificultando actividades básicas como el descanso, las actividades físicas o el uso prolongado de dispositivos médicos.

Por el otro punto de vista a nivel mental, la carga constante de dolor y la falta de soluciones precisas pueden acabar en niveles altos de estrés y de ansiedad. La relación entre el dolor y la perdida del estado de ánimo está ampliamente documentada y en el caso de los cirujanos, puede producir una pérdida de motivación, concentración y compromiso en el ámbito laboral. La salud mental se ve afectada cuando los dolores afectan a la percepción del control, lo que puede tener un efecto bastante negativo en la toma de decisiones clínicas y sobre la relación médico-paciente.

En el ámbito profesional, los DME afectan directamente en la eficiencia y continuidad laboral. Los cirujanos que presentan este tipo de dolores muestran una mayor tendencia al absentismo, reducción de la jornada laboral o necesidad de pausas frecuentes durante las intervenciones. Esto hace provocar un aumento en la carga de trabajo del resto del equipo, afectando a la dinámica del grupo y la programación quirúrgica del centro hospitalario. En los casos más extremos se puede llegar incluso a forzar al cirujano a modificar su especialización o retirarse antes de tiempo en la práctica quirúrgica, lo que supondría una pérdida del capital humano cualificado para el sistema sanitario.

En el ámbito económico también resulta significativo. El tratamiento del dolor musculoesquelético en cirujanos requiere tanto recursos humanos como materiales como por ejemplo las consultas médicas, tratamientos farmacológicos y de fisioterapia. A esto se puede sumar también los costes indirectos como por ejemplo la perdida productividad, necesidad de sustituciones o reajustar turnos de trabajo. Por este lado el DME no solo afecta a la salud individual, sino que también es un problema estructural para el sistema hospitalario.

Por ello, estudiar los DME no es únicamente relevante desde una perspectiva clínica o estadística, si no que también permite identificar oportunidades de mejora en la organización del trabajo, optimizando los recursos y protegiendo el bienestar de las personas sanitarias que desempeñan un papel clave en la atención sanitaria especializada.

1.3. Estudios previos y relevancia del estudio

El dolor musculoesquelético (DME) se han realizado varias investigaciones ya que es bastante importante para los profesionales sanitarios. En estas investigaciones la mayoría de ellas han demostrado que los cirujanos están muy expuestos a desarrollar esta patología. Algunas de estas investigaciones abarcaban aspectos como por ejemplo la prevalencia del DME, factores de riesgo asociado y sus efectos sobre la calidad de vida y calidad de trabajo profesional, habiendo tenido como conclusiones que hay una necesidad de desarrollar intervenciones específicas para su prevención y tratamiento.

Un estudio destacado publicado en Dialnet señala que un 87% de 409 neurocirujanos han experimentado algún DME, los sitios corporales más frecuentes fueron el cuello, los hombros y la espalda las zonas del cuerpo con mayor frecuencia, debido a posturas forzadas, movimientos repetitivos, uso prolongado de instrumental quirúrgico y ausencia de pausas activas durante las intervenciones quirúrgicas prolongadas. Además, el estudio identifica que las tecnologías quirúrgicas emergentes como la cirugía robótica ha conseguido mejorar la precisión de las intervenciones, pero también ha conseguido crear nuevos patrones de estrés físico, como por ejemplo las posturas mantenidas frente a las consolas, el uso continuo de los brazos, muñecas y pulgares o faltas de descanso durante las cirugías. Como conclusiones sobre este estudio, los riesgos ergonómicos en la práctica quirúrgica pueden tener un impacto significativo en la salud musculoesquelética de los cirujanos afectando a su bienestar capacidad de trabajar de manera segura y efectiva. Mejorar el diseño de las salas de operaciones y seleccionar equipos ergonómicos son medidas importantes.

El estudio “Dolor musculoesquelético en cirujanos españoles: resultado de una encuesta nacional” (Estudio ErgoSurgeon) en donde se analizó una muestra de 651 cirujanos miembros de la Asociación Española de Cirujanos (AEC), encontrando que el 84.2% respondieron tener dolor cervical, 83.4% dolor en la zona lumbar y un 74.3% dolor en hombros durante el trabajado en el quirófano. Más de la mitad de los participantes declaró haber requerido algún tipo de tratamiento, y un 11.7% no asistieron al puesto de trabajo debido a estas molestias.

La mayoría de los cirujanos encuestados en estos estudios reconocen no haber recibido formación sobre ergonomía ni en la carrera de Medicina ni durante la residencia. Esto significa que hay un vacío importante en la formación profesional y justifica la necesidad de introducir herramientas preventivas, como por ejemplo programas formativos en ergonomía, recomendaciones posturales, pausas activas en el quirófano y rediseño del entorno quirúrgico.

Este trabajo de fin de grado, se va a envolver dentro de esta línea de investigación, tratando el estudio del DME en cirujanos españoles mediante el estudio de datos obtenidos a través de la encuesta ErgoSurgeon. La novedad de esta investigación se concentra en aplicar técnicas estadísticas y modelos de aprendizaje automático para poder predecir el riesgo de sufrir DME, así como el desarrollo de una calculadora predictiva orientada a la prevención personalizada.

1.4. Objetivos del estudio

El desarrollo de estos dolores en cirujanos es un problema que cada vez hay mas casos dentro del ámbito hospitalario y quirúrgico, con consecuencias tanto individuales como organizativas. La literatura científica recoge que las posturas mantenidas durante largas intervenciones, el uso de instrumentos pocos ergonómicos y la falta de formación sobre ergonomía son factores determinantes en la aparición de estas dolencias.

A partir del estudio ErgoSurgeon surge la necesidad de tratar con esta problemática desde una perspectiva analítica y predictiva. En este contexto, este trabajo de fin de grado tiene como objetivo el análisis del dolor musculoesquelético en cirujanos españoles a partir de los resultados de la encuesta ErgoSurgeon y desarrollar una herramienta predictiva que permita estimar el riesgo individual de padecer DME. Esta herramienta va a estar basada en modelos estadísticos como la regresión logística para ver que variables son las mas influyentes y aplicar algoritmos de aprendizaje automático como Random Forest para ver como se comportan los modelos, por lo que finalmente se crea una calculadora con las variables mas relevantes que afecten a la aparición de DME a través de una aplicación web interactiva desarrollada en R Shiny.

Los pasos que seguiremos en este trabajo serán los siguientes:

1. **Análisis descriptivo:** se realizará una caracterización de la muestra recogida en la encuesta ErgoSurgeon, analizando las variables como la edad, el género, el nivel de actividad física, los años de experiencia como cirujano, así como la frecuencia y tipo de operaciones quirúrgicas que realizan. También se analizará la prevalencia del dolor musculoesquelético según el área corporal (cuello, zona lumbar, hombros, muñecas, rodillas, etc), cuánto dura el dolor, la intensidad del dolor y si se ha requerido tratamiento o baja laboral. Todo esto se mostrará a través de tablas y gráficos para facilitar la interpretación de los resultados.
2. **Análisis diferencial:** se llevará a cabo técnicas estadísticas para explorar asociaciones significativas entre la presencia de dolor musculoesquelético y variables individuales, ergonómicas o profesionales. Para las variables categóricas como el sexo, el tipo de hospital o la mano dominante, se realizará la prueba de chi-cuadrado o Fisher. En caso de las variables numéricas como la edad, el peso, la estatura se empleará prueba U de Mann-Whitney es un método para comparar diferencias entre los grupos con y sin dolor musculoesquelético. Esta elección responde a la ausencia de normalidad en la mayoría de estas variables. Este análisis permitirá identificar cuales son los factores diferenciales entre grupos y servirá para fases posteriores de modelización estadística.
3. **Modelización:** con las variables más significativas, se construirá modelos de regresión logística que nos facilite estimar la probabilidad de sufrir dolor musculoesquelético en las distintas zonas del cuerpo. Se evaluará la significación estadística de los coeficientes, el cálculo del Odds Ratio (OR) y su intervalo de confianza.

4. **Aplicación de algoritmos de aprendizaje automático:** se aplicará un modelo de Random Forest de cada zona del cuerpo con el objetivo de clasificar la presencia o ausencia de dolor en función de las variables más significativas. Este modelo nos aporta una medida de importancia relativa para cada predictor, lo que será clave para construir la herramienta predictiva. Se evaluarán todos los modelos analizando la precisión, sensibilidad y especificidad para ver cuanto de bueno es el modelo.
5. **Desarrollo de la calculadora predictiva:** a partir de los modelos creados y validados, se implementará una aplicación interactiva con R Shiny que permita al usuario introducir sus características y que aparezca una estimación del riesgo de presentar DME en cada zona corporal en función de los factores analizados. Esta herramienta tiene como objetivo que sea utilizada para profesionales en activo, formadores y gestores de recursos humanos en salud.
6. **Elaboración de recomendaciones:** basado en los resultados del análisis se recomendará actividades/prácticas dirigidas para minimizar la aparición del DME en los cirujanos. Estas recomendaciones podrán ser como por ejemplo la formación en ergonomía, propuestas de rediseño del entorno quirúrgico, incorporación de pausas activas o acceso a recursos como la fisioterapia preventiva.

Con este enfoque, el presente trabajo de fin de grado no solo tiene como objetivo profundizar en la dimensión estadística del problema si no que también generar un impacto práctico en el entorno quirúrgico, fomentando una cultura de prevención y bienestar profesional.

1.5. Metodología de la Encuesta ErgoSurgeon

En este trabajo de fin de grado se centra en los datos obtenidos mediante la encuesta ErgoSurgeon, que consiste en un cuestionario que está diseñado especialmente para evaluar la prevalencia del dolor musculoesquelético (DME) y los factores ergonómicos que influyen en el trabajo quirúrgico en España. La encuesta fue elaborada por un equipo de investigación vinculado a la Universidad Miguel Hernández de Elche, en colaboración con la Asociación Española de Cirujanos (AEC), y aprobada por el Comité de Ética e Integridad en la investigación con el código DPC.SHS.01.21, cumpliendo con los principios éticos de la investigación.

Esta encuesta se realizó de forma telemática mediante la plataforma Google Forms, se facilitó la base de datos asociados de la AEC, lo que permitió acceder directamente a los profesionales que fueron objeto de estudio. La recogida de datos se efectuó entre los meses de diciembre de 2021 y marzo de 2022, a través de tres rondas de envío por correo electrónico. Las participaciones fueron anónimas, de forma voluntaria y sin compensaciones económicas, lo que garantiza tanto la confidencialidad como la integridad de las respuestas.

Se obtuvieron una muestra total de 651 cirujanos/as, lo que representa una tasa de respuesta del 31.6% sobre el total de miembros a los que se les envío la encuesta. Esta muestra se considera representativa en términos de sexo, especialidad, tipo de hospital y años de experiencia profesional, según los datos proporcionados por la propia AEC.

El cuestionario ErgoSurgeon está diseñado con un enfoque mixto ya que se ha combinado preguntas cerradas de tipo dicotómico, de elección múltiple, escala de tipo Likert y campos numéricos. En este cuestionario está dividido en tres grandes bloques de contenido:

1. **Bloque 1:** Variables individuales y profesionales. Se recogen variables sociodemográficas como la edad, sexo, peso, estatura, y nivel de actividad física semanal. También recoge preguntas como la experiencia laboral (años en activo como cirujano), frecuencia de guardias, número aproximado de cirugías anuales y tipo de cirugía más frecuente (abierta, laparoscópica, robótica, endoscópica, etc.).
2. **Bloque 2:** Síntomas musculoesqueléticos. En este bloque se adapta al Cuestionario Nórdico Estandarizado de síntomas musculoesqueléticos, que consiste en una herramienta validada internacionalmente para evaluar molestias físicas que están relacionadas con la actividad laboral. Se recogieron información sobre la presencia de dolores en distintas zonas corporales (cuello, zona dorsal, zona lumbar, hombros, muñecas, rodillas y pies), así como la intensidad del dolor (escala del 0 al 10), duración y recurrencia de dichos síntomas. También se quiere recoger información de si estos dolores han requerido tratamiento (fisioterapia, medicación, cirugía, etc), o si por los dolores ha resultado implicar bajas laborales o reducción del rendimiento.
3. **Bloque 3:** Aspectos ergonómicos. En el último bloque se aborda el nivel de conocimiento del profesional sobre ergonomía quirúrgica. Se ha preguntado si ha recibido alguna formación específica sobre ergonomía (tanto durante el grado universitario como en la residencia), si aplica técnicas preventivas como por ejemplo descansos, estiramientos o sillas ergonómicas y su percepción sobre las condiciones posturales en el quirófano. Finalmente, se incluye preguntas abiertas para que los profesionales puedan sugerir medidas de mejora.

Además, es importante tener en cuenta algunas de las limitaciones del diseño de este tipo de encuestas. Al ser una encuesta de autocompletado, las respuestas pueden estar influenciadas por como recuerda cada profesional sus síntomas o experiencias pasadas, lo que puede generar falta de precisión. También puede suceder que aquellos cirujanos que han sentido alguna vez molestias estén más motivados a participar que quienes no han tenido dolores, lo que podría hacer que los resultados sobre el dolor parezcan más altos de los que realmente son. Aun así, el número de respuestas que se han obtenido es suficientemente amplio y diverso.

En resumen, la encuesta ErgoSurgeon es una fuente de datos robusta, detallada y específica de los profesionales quirúrgicos en España, lo que permite abordar el problema del DME desde una perspectiva rigurosa y con vocación aplicada.

2. Modelos estadísticos y aprendizaje automático

2.1. Introducción a los modelos de regresión logística

La Regresión Logística es una técnica estadística de análisis de datos que fue desarrollada en el año 1958 por David Cox. El término “Regresión” su objetivo es estudiar el comportamiento de una variable dependiente (variable de salida) que estará influenciada o no por una serie de variables independientes (variables de entrada), por lo que estima un modelo explicativo. Por lo que dicho esto la Regresión Logística es un tipo específico de regresión que analiza cómo se comporta una variable dependiente binaria y categórica en relación con un conjunto de variables independientes, que estas variables pueden ser tanto cuantitativas como cualitativas.

Como se ha mencionado antes, esta técnica para este trabajo, la variable dependiente a analizar es binaria, es decir, que solo puede tomar dos valores que normalmente son el “0” y “1”, por lo que se podría decir que las variables son dicotómicas. En este contexto la variable dependiente es si el cirujano a presentado dolores en alguna zona corporal, el valor “1” representaría que el cirujano si ha presentado dolores en esa zona corporal y el “0” representaría que el cirujano no ha presentado dolores en esa zona corporal. Esto se conoce como variable de salida. En el caso de que la variable a estudiar fuera numérica o distinta de binaria, se realizaría otra técnica como por ejemplo la Regresión Lineal. En cuanto a las variables predictoras, su nombre lo dice, tratan de predecir el valor de la variable dependiente.



Figura 1. Esquema general del funcionamiento de un modelo de regresión logística (Sotaquirá, 2018)

Este tipo de modelos de regresión logística tienen dos finalidades:

- Cuantificar la importancia de la relación existente entre cada una de las variables independientes y la variable dependiente.
- Clasificar a los individuos dentro de las categorías (presente/ausente) de la variable dependiente, según la probabilidad que tenga de pertenecer a una de ellas dada la presencia de determinadas variables independientes.

En regresión logística simple, también conocida como regresión *logit*, es una de las herramientas que ofrece los Modelos Lineales Generalizados.

El modelo de regresión logística binaria determina la probabilidad de que la variable aleatoria Y tome el valor 1 (éxito):

$$P(Y = 1 | X) = \frac{e^{b_0 + \sum_{i=1}^n b_i x_i}}{1 + e^{b_0 + \sum_{i=1}^n b_i x_i}}$$

siendo $P(Y = 1 | X)$ probabilidad de que Y tome el valor 1, en presencia de las variables independientes X.

Los parámetros de esta ecuación son:

- b_0 : constante del modelo o término independiente.
- n : número de variables independientes o covariables.
- b_i : coeficientes de las variables independientes.

Si dividimos la expresión anterior por su complementario, es decir, si se construye su odds, se obtiene la expresión:

$$\frac{P(Y = 1 | X)}{1 - P(Y = 1 | X)} = e^{b_0 + \sum_{i=1}^n b_i x_i}$$

Si ahora realizamos su transformación logarítmica con el logaritmo neperiano, obtenemos una ecuación lineal que es más fácil de manejar y comprender.

$$\ln \left(\frac{P(Y = 1 | X)}{1 - P(Y = 1 | X)} \right) = b_0 + \sum_{i=1}^n b_i x_i$$

A la izquierda de la igualdad está el logit, es decir, el logaritmo neperiano de la odds de la variable dependiente. El término a la derecha de la igualdad es la expresión de una recta, idéntica a la del modelo general de regresión lineal.

En la regresión logística la estimación de parámetros se lleva a cabo a través del método de máxima verosimilitud, de modo que los coeficientes que estima el modelo hacen nuestros datos “más verosímiles” (Visauta, 1998).

Además, también es posible obtener una representación gráfica de las probabilidades predichas en cada modelo de regresión logística. Así, en los modelos de regresión logística simple, esto es, los que cuentan con una sola variable predictora o independiente, se grafica la curva de probabilidades o función logística del modelo, que representa la probabilidad \hat{p} estimada por el modelo de que suceda el evento (eje vertical), en función del valor de la variable independiente X_1 (eje horizontal). En la siguiente imagen se muestra un ejemplo de este tipo de curva:

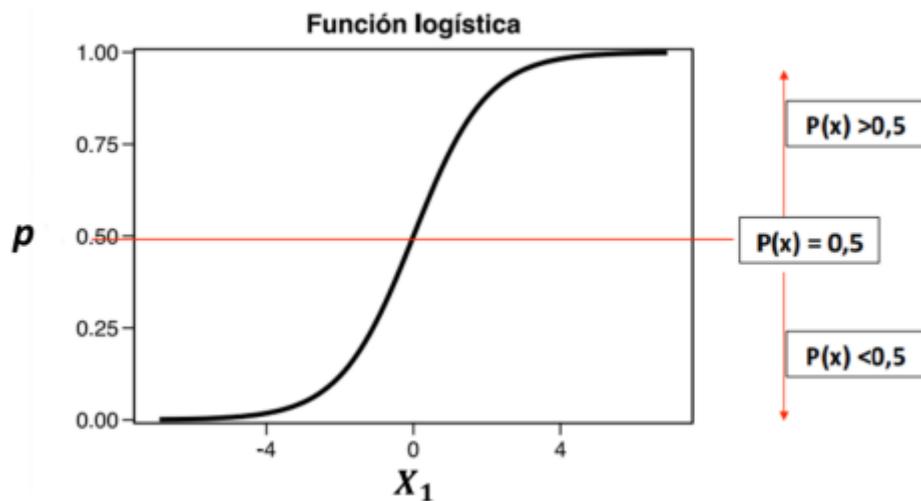


Figura 2. Función logística

Una vez se tiene el modelo inicial debe procederse a su reducción hasta obtener el modelo más reducido que siga explicando los datos. Para ello se ha de recurrir a métodos de selección paso a paso, en las cuales hay diversas técnicas para elegir:

- **Técnica de pasos hacia adelante (Forward):** consiste en ir introduciendo las variables en el modelo únicamente si cumplen una serie de condiciones hasta que no se pueda introducir ninguna más, hasta que ninguna cumpla la condición impuesta.
- **Técnica de pasos hacia atrás (Backward):** se introducen en el modelo todas las variables y se van suprimiendo si cumplen una serie de condiciones definidas a priori hasta que no se pueden eliminar más, es decir ninguna variable cumpla la condición impuesta.
- **Técnica por pasos (Stepwise):** combina los dos métodos anteriores, adelante y atrás introduciendo o eliminando variables del modelo si cumplen una serie de condiciones definidas a priori hasta que ninguna variable satisfaga ninguna de las condiciones expuestas de entrada o salida del modelo.

- **Técnica de introducir todas las variables obligatoriamente (Enter):** esta última técnica de selección de variables para construir el modelo de regresión, produce que el proceso de selección de las variables sea manual, partiendo de un modelo inicial, en el que se obliga a que entren todas las variables seleccionadas, se va evaluando qué variable es la que menos participa en él y se elimina, volviendo a construir un nuevo modelo de regresión aplicando la misma técnica, pero excluyendo la variable seleccionada y aplicando el mismo proceso de selección. Este proceso se repite reiteradamente hasta que se considere que el modelo obtenido es el que mejor se ajusta a las condiciones impuestas y que no se puede eliminar ninguna variable más de las que los componen.

AIC Y BIC

Para comparar el ajuste entre dos modelos se usará el criterio de información de Akaike (AIC) y el criterio de información de Bayes (BIC). Ambos criterios fueron propuestos por Akaike (1974) y Schwarz (1978) respectivamente. Estos criterios proporcionan una medida del ajuste de un modelo que penaliza al modelo que contiene más variables predictoras.

$$AIC = -2LL + 2k$$

$$BIC = -2LL + 2k \ln(n)$$

En donde:

- LL: logaritmo de la verosimilitud del modelo ajustado. Mide que tan bien se ajusta el modelo de los datos (cuanto mayor, mejor).
- K: número de parámetros del modelo.
- n: tamaño de la muestra.

Al ajustar un modelo es posible aumentar su precisión mediante la adición de parámetros, pero si esto se lleva a cabo, se puede derivar en un sobreajuste. Tanto el BIC y AIC resuelven este problema mediante la introducción de un término de penalización para el número de parámetros en el modelo. Este término de penalización es mayor en el BIC que en el AIC. Dado un conjunto de modelos candidatos para los datos, el modelo preferido es el que tiene el valor mínimo en el AIC y BIC.

Las ventajas que supone el empleo de la regresión logística como técnica de análisis de datos pueden ser:

- Es capaz de manejar correctamente el empleo de variables independientes de diverso tipo, tanto categóricas como numéricas.
- Su uso es muy adecuado cuando se cuenta con una variable dependiente binaria o categórica.

- Realiza una estimación de probabilidades, por lo que es posible analizar las predicciones realizadas en términos de fiabilidad.
- Presenta una sencilla implementación e interpretación, proporcionando coeficientes que reflejan la influencia o importancia de cada variable independiente en la variable dependiente o respuesta. Además, según el signo de estos coeficientes, se concluye el sentido en el que influye cada variable. También, es posible analizar si una variable es o no significativa mediante un contraste sobre cada coeficiente obtenido.
- Presenta cierta robustez ante la presencia de datos atípicos en la base de datos usada.
- Suele conseguir una alta eficacia y precisión sin requerir grandes recursos computacionales, y sus resultados pueden interpretarse de manera simple.
- Cuenta con una amplia disponibilidad en cuanto a su uso, y sus posibles campos de aplicación también son muy diversos.

Por otro lado, la regresión logística también presenta ciertas limitaciones en su empleo, algunas de estas limitaciones son las siguientes:

- Sirve para predecir variables respuesta que sean discretas o categóricas, pero no continuas, lo que limita en parte su uso. Además, si existen más de dos clases de la variable dependiente, los resultados se vuelven más difíciles de interpretar.
- Requiere que las observaciones sean independientes unas de otras para obtener resultados fiables.
- Una multicolinealidad alta entre las variables independientes, como en el caso de la regresión lineal, dificulta la interpretación del efecto de las variables independientes sobre la variable respuesta, pero no afecta a la predicción.
- Existen algoritmos más potentes que pueden crear predicciones más completas.

Además, este tipo de modelos pueden ser mejorados mediante procedimientos de selección automática de variables como por ejemplo los métodos de paso a paso. Estos tipos de métodos nos permitirán mejorar la eficiencia del modelo al eliminar las variables independientes que no aportan valor estadístico significativo. Por lo que se obtendrá un modelo más compacto, más fácil de interpretar y con mejor capacidad predictiva, lo cual es un pilar fundamental para diseñar herramientas prácticas como una calculadora de riesgos desarrollada en este trabajo.

2.2. Odds Ratio e interpretación de coeficientes

La regresión logística tiene como objetivo indicar la probabilidad de ocurrencia de una variable dependiente Y dicotómica. En este ámbito, van a ser muy importantes los términos de Odds y Odds Ratio.

Los odds asociados a un cierto suceso se definen como la razón de probabilidad de que dicho suceso ocurra y la probabilidad de que no ocurra, por lo que es un número que expresa una probabilidad de que se produzca ese suceso. Si llamamos E a ese suceso, $P(E)$ a la probabilidad de que ocurra y $O(E)$ a los odds que le corresponden, entonces se tiene:

$$O(E) = \frac{P(E)}{1 - P(E)}$$

En la literatura médica se utiliza frecuentemente el término “riesgo” para referirse a los “odds”, de forma que, con esta nomenclatura, riesgo de una determinada patología en una población es igual a la prevalencia/(1-prevalencia) en donde la prevalencia viene expresada en tanto por uno (y no en porcentaje) como se hace habitualmente.

Vamos a poner un ejemplo para los odds, si se estima que el 75% de los cirujanos presenta dolores musculoesqueléticos, se dice que los odds de que un cirujano padezca de dolores musculoesqueléticos es de 3 (ya que $0.75/0.25 = 3$). Por lo que la ocurrencia de este suceso es 3 veces más probable de que no ocurra este suceso.

La interpretación de los odds la podríamos resumir de la siguiente manera:

- Un Odds igual a 1 significa que es igual de probable que un cirujano tenga dolores musculoesqueléticos a que no.
- Un Odds mayor que 1 significa que para un cirujano es más probable tener dolores musculoesqueléticos que no tenerlos, puesto que el numerador que es la probabilidad de que ocurra el suceso tendrá más peso.
- Un Odds menor que 1, significa que para un cirujano es menos probable de tener dolores musculoesqueléticos que no, puesto que el denominador presenta la ocurrencia del suceso por lo que tendrá más peso.

Por otra parte, los odds se puede deducir la probabilidad. En general si los odds de un suceso E asciende a $O(E)$, entonces su probabilidad es:

$$P = \frac{O(E)}{O(E) + 1}$$

Si ponemos otro ejemplo para aquí, los odds de que un cirujano padezca de dolor musculoesquelético es de 0.4 esto significa que la probabilidad de ese hecho ocurra es $0.4/1.4 = 0.285$.

De modo que ambas informaciones son equivalentes y expresan la misma idea: calcular cuán probable es de que ocurra el suceso. Entre la probabilidad del suceso y los odds existe una relación directa: si aumenta la probabilidad del suceso también aumenta los odds. Si $P(E)=1$, entonces $O(E)$ tiende a infinito, pero en cambio si $P(E)=0$ entonces $O(E)$ será nulo también.

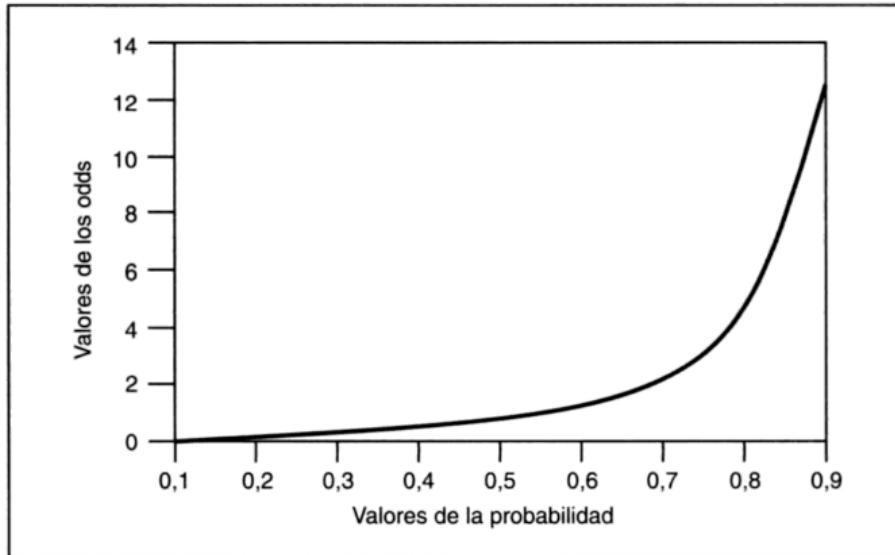


Figura 3. Transformación del espacio probabilístico al de odds

Por otro lado, existe el **riesgo relativo** de que se produzca el suceso E cuando esta presente la condición A, que lo denominaremos $P(E|A)=PA(E)$, y de que se produzca E cuando ocurre la condición B, de denotaremos como $P(E|B)=PB(E)$, es la razón:

$$RR = \frac{PA(E)}{PB(E)}$$

Con esta formulación el riesgo relativo expresa cuánto más probable es que ocurra el suceso E cuando esta en la condición A respecto cuando esta en la condición B.

Una vez definidos los odds que es una manera de expresar la probabilidad de que ocurra un suceso y el riesgo relativo expresa la relación entre dos probabilidades, por lo que tiene sentido realizar el riesgo relativo entre dos Odds, es decir, considerar la razón entre dos probabilidades dada por dos Odds, lo cual se considera como Odds Ratio.

Los odds-ratio (OR) se calculan como la razón entre los odds, donde la variable respuesta Y está presente entre los individuos, es decir, toma el valor $Y=1$, y la variable independiente X puede estar presente o no, es decir tomar los valores $X=1$ y $X=0$.

$$OR = \frac{\frac{P(Y=1 | X=1)}{1 - P(Y=1 | X=1)}}{\frac{P(Y=1 | X=0)}{1 - P(Y=1 | X=0)}}$$

La interpretación de los odds ratios la podemos ver de estas tres maneras:

- Un OR= 1 implica que no existe asociación entre la variable respuesta y la variable independiente.
- Un OR inferior a la unidad se interpreta como un factor de protección, es decir, el suceso es menos probable en presencia de la variable independiente.
- Un OR mayor que 1 se interpreta como un factor de riesgo, ya que significa que el suceso es más probable en presencia de dicha variable independiente.

En resumen, si se quiere conocer como afecta el valor de cada variable a la probabilidad de que la variable Y tome un valor u otro, hay que recurrir a los OR. Si se tiene por ejemplo una variable binaria Y que indica si un cirujano tiene dolor musculoesquelético (1) o si no tiene (0), y una variable binaria que indica no realiza pausas (1) o si realiza pausas (0), si el OR > 1, el hecho de no realizar pausas aumenta la probabilidad de que el cirujano pueda padecer de dolores, si el OR es <1, realizar pausas disminuye la probabilidad de que el cirujano padezca de dolores. Si un OR es mayor que 1 no nos puede garantizar la relación entre tener la realización de pausas y la de padecer dolores, ya que puede estar influenciada por una tercera variable que no se recoge en el modelo.

A la hora de interpretar los coeficientes del modelo debemos tener en cuenta dos cosas: la relación funcional entre la variable dependiente y las independientes, y la unidad de cambio para las variables independientes. Además, la interpretación de los coeficientes dependerá también de que tipo de variables tengamos: dicotómicas, polítómica o continuas.

2.3. Bondad de ajuste

En esta sección se van a mostrar diversas maneras de medir como de bueno es el ajuste de los valores predichos por el modelo, a los valores observados. Se dice que un modelo presenta un buen ajuste de los datos si los valores que predice reflejan de manera correcta los valores observados. Si el modelo presenta un mal ajuste, este no puede ser utilizado para extraer conclusiones ni efectuar predicciones.

Existen diversas medidas de ajuste global para comparar la diferencia entre valores predichos y valores observados. Los mas utilizados son el test basado en la devianza D, el estadístico X^2 de Pearson y el test de Hosmer-Lemeshow. Los dos primeros consisten en patrones de las variables independientes y puede ser usados en los modelos lineales generalizados en general. El tercero se basa en probabilidades estimadas y se aplica en el caso de un MLG con distribución binomial, es decir, un modelo de regresión logística.

Test de Hosmer-Lemeshow

La evaluación del ajuste del modelo se realiza con los valores de Z_i y p_i las probabilidades calculadas a partir del modelo ajustado de la muestra de tamaño n . En primer lugar, se ordenan los valores de p de menor a mayor, y se forman k grupos. Si $Z = 1$ se suman los valores de p dentro de cada uno de los grupos formados, estas sumas son los valores esperados E_i . Por último, se cuentan los valores observados O_i en los que $Z = 1$. Por lo tanto, se realiza el siguiente contraste:

$$\begin{cases} H_0 : \text{El modelo ajusta bien los datos} \\ H_a : \text{Mal ajuste del modelo} \end{cases}$$

Y el estadístico de Hosmer-Lemeshow se calcula de la siguiente manera (Silvia Aycaguer y Barroso, 2004).

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} + \sum_{i=1}^k \frac{(O_i^* - E_i^*)^2}{E_i^*}$$

En donde O_i y E_i :

$$O_i = \sum_{j \in \text{grupo } i} Y_j, \quad E_i = \sum_{j \in \text{grupo } i} \hat{p}_j$$

- O_i es el número de casos reales con $Y=1$ dentro del grupo.
- E_i es la suma de las probabilidades \hat{p}_j de esos mismos sujetos; equivale al número de eventos que el modelo “espera” en el grupo

También tenemos O_i^* y E_i^* ;

$$O_i^* = n_i - O_i, \quad E_i^* = n_i - E_i$$

- O_i^* es, pues, el número de casos que sí tienen $Y = 0$ en el grupo. Dado que en el grupo hay n_i observaciones en total y O_i de ellas corresponden a $Y = 1$, el resto ($n_i - O_i$) son los casos con $Y = 0$.
- E_i^* es el número de no-eventos esperados, que resulta de restar del total n_i el número de eventos esperados E_i . En otras palabras, si el modelo “espera” E_i eventos, el resto ($n_i - E_i$) son los no-eventos que, en promedio, deberían observarse.

donde O_i^* y E_i^* son los datos observados y esperados, respectivamente para los que $Z = 0$ y n_i es el grupo correspondiente, n_i es el tamaño del grupo y k es el grupo en cuestión. Sigue una distribución χ^2 con $k - 2$ grados de libertad siendo k el número de grupos formados (Agresti, A.(2007)). La hipótesis nula establece que no hay diferencias entre los valores observados y los pronosticados, es decir, el modelo ajusta bien los datos. Por tanto, si el ajuste es bueno, se espera un valor alto de p -valor, superior o igual a 0,05.

Esta prueba consiste en dividir el recorrido de valores de la variable dependiente dicotómica Y en una serie de intervalos. Estos intervalos deben contener un número de observaciones suficientemente grande (cinco o más). Se trata de contar intervalo por intervalo el valor esperado y el observado para cada uno de los dos resultados posibles de la variable dependiente dicotómica.

Cada vez que se incluye o elimina una variable hemos de comparar el nuevo modelo con el anterior hasta que todas las variables importantes estén en el modelo final. Un modelo de regresión logística puede resultar inadecuado por diferentes razones.

Una de ellas es la invalidez del componente lineal del modelo. Esta situación es frecuente cuando ciertas variables explicativas o términos de interacción no son incluidos en el modelo cuando deberían serlo, o también cuando no se realizan transformaciones en ellas que permitan mejorar el ajuste a los datos. La existencia de outliers puede ser también determinante en un mal ajuste.

Recordad que el modelo logístico no asume linealidad entre la variable dependiente y una covariable, pero sí respecto al logit de la odds de la variable dependiente. Si no se cumpliera esta asunción, la relación entre dicha covariable y la variable dependiente estaría infraestimada. Una manera de comprobar si alguna de las covariables no cumple esta asunción de linealidad respecto al logit es dibujar los gráficos de dispersión.

R y R²

En regresión lineal se utiliza el coeficiente de determinación, llamado R², que se define como la proporción de la varianza total de la variable explicada por la regresión y refleja la bondad de ajuste de un modelo. Este coeficiente oscila entre 0 y 1, cuanto más próximo a 1 sea su valor, mayor será el ajuste del modelo, y cuanto más próximo a 0 menos ajustado estará el modelo.

$$R^2 = \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}$$

En el caso de la regresión logística se podría considerar esta medida substituyendo \hat{y}_i por $n_i \hat{p}_i$, siendo \hat{p}_i la probabilidad estimada. Los estimadores verosímiles no son los que maximizan esta medida ni tiene en cuenta la dependencia de la varianza de Y respecto de p. Por todo esto se utilizan unas medidas alternativas, denominadas pseudo R², las principales son:

- R² de Cox-Snell: este coeficiente se obtiene en función de la verosimilitud, aunque se puede expresar también en función de las distancias. El valor de R²_{CS} ∈ [0; 1 – (V₀/V)²] y queda definido como:

$$R_{CS}^2 = 1 - \left(\frac{V_0}{V} \right)^{\frac{2}{N}}$$

donde V es la verosimilitud y N es el tamaño de la muestra.

- R^2 de McFadden: compara la log-verosimilitud del modelo ajustado, L , con la log-verosimilitud del modelo nulo, L_0 . Proporciona una idea de cuanto se reduce la distancia (es decir, la verosimilitud) al incluir variables explicativas. Su valor teóricamente oscila entre 0 y 1, pero rara vez se aproxima a 1. Se considera un buen ajuste cuando $R^2_{MF} \in [0.2;0.4]$ y excelente para valores más grandes.

$$R^2_{MF} = 1 - \frac{L}{L_0}$$

- R^2 de Nagelkerke: es una modificación del índice de Cox–Snell, usando la misma notación de log-verosimilitud (L_0 y L). Primero, definimos el Cox–Snell “pseudo- R^2 ” como

$$R^2_{CS} = 1 - \left(\frac{L_0}{L} \right)^{\frac{2}{n}}$$

donde n es el tamaño de la muestra. A partir de ahí, el coeficiente de Nagelkerke se obtiene corrigiendo el máximo teórico para que quede en el rango [0,1]:

$$R^2_N = \frac{R^2_{CS}}{1 - (L_0/L)^{\frac{2}{n}}}$$

De este modo, $R_N^2 \in [0,1]$ y puede interpretarse de manera análoga al R^2 de una regresión lineal, aunque en la práctica rara vez toma valores muy cercanos a 1.

En resumen, se trata de la correlación parcial entre la variable dependiente y cada una de las variables independientes. Este valor puede variar entre el 0 y el 1, si por ejemplo esta en un valor cercano a 0 implica que, si la variable predictora decrece, la probabilidad de que el resultado ocurra disminuye y si esta cercano al valor 1 significa que, al crecer la variable predictora, lo hace la probabilidad de que el evento ocurra.

Chi-Cuadrado de Pearson

El estadístico Chi-Cuadrado de Pearson compara frecuencias observadas y esperadas en una distribución binomial. Dicho estadístico se define como:

$$\chi^2 = \sum_{q=1}^m r_q^2$$

$$r_q = \frac{y_q - n_q \hat{p}_q}{\sqrt{n_q \hat{p}_q (1 - \hat{p}_q)}}$$

En donde p_q es la probabilidad estimada de que ocurra el evento en la celda q (por ejemplo, obtenida a partir del modelo logístico o del ajuste de la tabla de contingencia), y n_q es el número total de observaciones en dicha celda.

Donde r_q denominados por Hosmer como residuos de Pearson. El contraste para realizar para verificar su significación estadística es:

$$\begin{cases} H_0 : r_q = 0 \\ H_1 : r_q \neq 0 \end{cases}$$

Bajo la hipótesis nula, r_q tiene distribución asintótica normal con media cero y varianza estimada menor que uno. A pesar de esto, los residuos de Pearson suelen ser tratados como normales estándar. Se considerarán significativos cuando sus valores absolutos sean mayores que dos.

Curva ROC

Las curvas ROC, acrónimo de Relative Operating Characteristic, o en español COR (Característica Operativa del Receptor), se comenzaron a utilizar durante la II Guerra Mundial para el análisis de señales de radar. En ese momento fueron parte de lo que se conoce como Teoría de detección de señales (Green y Swets, 1966). La utilización de las curvas ROC se ha extendido en diferentes ámbitos, y en particular, en el campo de la medicina (Del Valle Benavides, 2017). En esta última, el análisis ROC se utiliza de manera muy extensa en problemas relacionados con la epidemiología. también en radiología se utiliza para evaluar nuevas técnicas de diagnóstico por imagen. Hace poco se han mostrado muy útiles para la evaluación de técnicas de aprendizaje automático. La primera aplicación en este ámbito fue hecha por Spackman que demostró su valor para la comparación de diferentes algoritmos de clasificación.

Supóngase que se tiene una variable aleatoria Z que se distribuye como una Bernoulli de parámetro p , tal que $p = P(Z = 1)$, que representa la probabilidad de que el individuo esté enfermo. Sea X una variable aleatoria que mide cierta característica en cada individuo, cuyo resultado puede ser continuo o discreto. Sea c el valor umbral o punto de corte que se utiliza para construir la variable Y , la dicotomización de X , como se muestra en la ecuación correspondiente.

$$Y = \begin{cases} 1 & \text{si } x \geq c \text{ el cirujano presenta dolor} \\ 0 & \text{si } x < c \text{ el cirujano no presenta dolor} \end{cases}$$

La tabla de clasificación sería esta:

	Z = 1 (Enfermo)	Z = 0 (Sano)
Y = 1 (Prueba +)	Verdadero positivo (VP)	Falsos positivos (FP)
Y = 0 (Prueba -)	Falsos negativos (FN)	Verdadero negativo (VN)

Tabla 1. Tabla de clasificación

Un parámetro o magnitud asociado a dicha curva es el AUC (Area under the ROC Curve) o área bajo la curva, que refleja el área existente bajo la curva ROC, y expresa la fiabilidad que posee el modelo analizado al diferenciar entre categorías positivas y negativas. De este modo, un mayor valor de AUC (más cercano a uno) será indicativo de que el modelo es más preciso, mientras que un valor cercano a 0,5 indicará que no mejora apenas a una clasificación que se realizase de manera totalmente aleatoria. Además, las curvas ROC pueden completarse con intervalos de confianza en torno a su “estimación puntual” del valor de AUC.

Según Hosmer & Lemeshow (2000) el valor de AUC se interpreta de la siguiente manera:

- Si el AUC es igual a 0.5, no hay discriminación.
- Si el AUC está entre 0.5 y 0.7, hay poca discriminación.
- Si el AUC está entre 0.7 y 0.8, existe una discriminación aceptable.
- Si el AUC se encuentra entre 0.8 y 0.9 es una discriminación excelente.
- Si el AUC es mayor a 0.9 es que existe una discriminación casi perfecta.

La capacidad predictiva de un modelo de regresión logística se puede resumir mediante el concepto de sensibilidad y mediante el concepto de especificidad:

$$\text{sensibilidad} = \frac{VP}{VP + FN} \quad \text{especificidad} = \frac{VN}{VN + FP}$$

Es decir, la probabilidad de predecir correctamente el “éxito” de la variable respuesta, se denomina sensibilidad y la probabilidad de predecir correctamente el “fracaso” de la variable respuesta se denomina especificidad.

Una curva de tipo *receiver operating characteristic* (ROC) es un gráfico en el que se representa la sensibilidad en función de (1 – especificidad). Si se modifican los valores del valor de corte c seleccionado para clasificar cada registro en Y=1 o Y=0, y representamos la sensibilidad (en ordenadas) frente a (1 – especificidad) (en abscisas), tenemos la curva ROC. Es una curva cóncava que conecta los puntos (0, 0) y (1, 1). Cuanto mayor sea el área bajo la curva, mejores serán las predicciones.

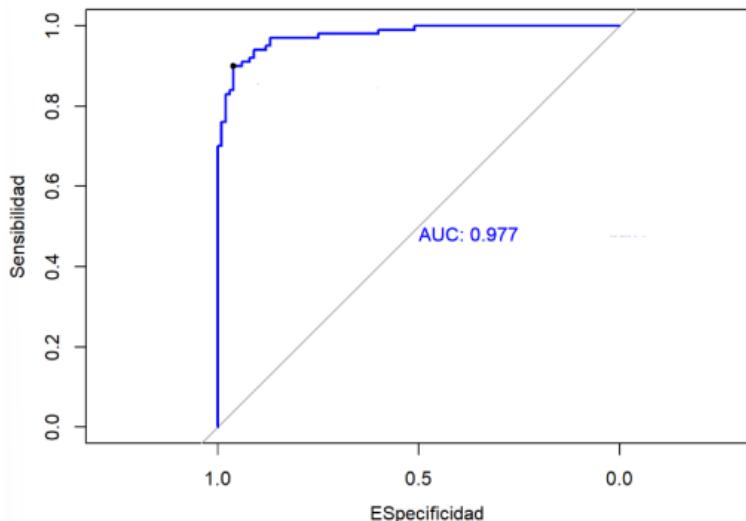


Figura 4. Ejemplo de curva ROC y valor AUC

Además, también se compararán en una misma gráfica varias curvas ROC asociadas a diferentes modelos, de manera que puedan observarse las diferencias entre ellas, y concluir qué modelo es mejor.

2.4. Uso del modelo de Random Forest en la clasificación del riesgo

Random Forest es una versión más compleja de la técnica Bagging (Bootstrap aggregating). Ambas técnicas fueron introducidas por Breiman en 2001 y 1996, respectivamente.

La técnica Bagging se basa en la generación de arboles basados en muestreos aleatorios, a fin de reducir la varianza de un estimador (función de predicción). La idea es tomar una parte del conjunto total de datos y predecir los valores fuera de la parte inicial. Aplicando estos pasos de forma iterativa y calculando el promedio de todas las predicciones, se obtiene modelos de predicción con muy buen rendimiento.

El algoritmo Random Forest es una técnica de aprendizaje supervisado que genera multitud de arboles de decisión sobre un conjunto de datos de entrenamiento: los resultados obtenidos se combinan a fin de obtener un modelo único más robusto en comparación con los resultados de cada árbol por separado (Lizares, 2017). Cada árbol se obtiene mediante un proceso de dos etapas:

1. Se genera un número considerable de árboles de decisión con el conjunto de datos. Cada árbol contiene un subconjunto aleatorio de variables m (predictores) de forma que $m < M$ (donde M = total de predictores).
2. Cada árbol crece hasta su máxima extensión.

Cada árbol que se ha generado por el algoritmo Random Forest contiene un grupo de observaciones aleatorias (elegidas mediante bootstrap, que es una técnica estadística para obtener muestras de una población donde una observación se puede considerar en mas de

una muestra). Las observaciones no estimadas en los árboles se utilizan para validar el modelo. Las salidas de todos los árboles se combinan en una salida final Y que se obtiene mediante alguna regla.

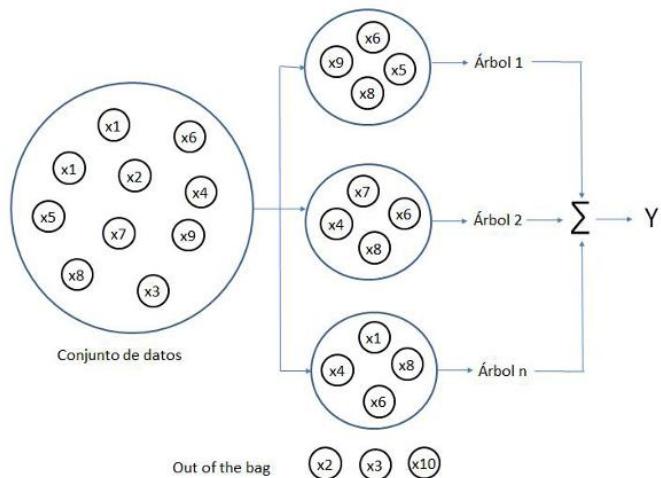


Figura 5. Algoritmo Random Forest

Las principales ventajas y desventajas del algoritmo Random Forest (Cánovas et al., 2017) son:

Ventajas	Desventajas
a) Puede utilizarse tanto para clasificación como para predicción. En clasificación, se asigna la clase con más votos; en predicción, se promedia la salida de los árboles.	a) La visualización gráfica de los resultados puede ser compleja y difícil de interpretar.
b) Es un modelo fácil de entrenar en comparación con otros algoritmos complejos, pero con rendimiento similar.	b) Puede sobreajustarse en presencia de ruido o datos inconsistentes.
c) Tiene un desempeño eficiente y es altamente certero en bases de datos grandes.	c) Las predicciones no son continuas, lo que impide extrapolar más allá del rango de entrenamiento. Puede haber sesgos con predictores categóricos de muchos niveles.
d) Puede manejar centenares de predictores, estimando automáticamente cuáles son más relevantes. Ideal para reducción de dimensionalidad.	d) Tiene poco control interno, lo que convierte el modelo en una 'caja negra' difícil de explicar en detalle.

e) Mantiene una buena precisión incluso con datos perdidos, lo que lo hace robusto ante información incompleta.	
---	--

Tabla 2. Ventajas y desventajas Random Forest

Las ventajas de Random Forest hacen que se convierta en una técnica ampliamente utilizada en muchos campos, por ejemplo, teledetección (para clasificación de imágenes), bancos (para detección de fraudes y clasificación de clientes para otorgamiento de crédito), medicina (para analizar historiales clínicos a fin de identificar enfermedades potenciales en los pacientes), finanzas (para pronosticar comportamientos futuros de los mercados financieros) y comercio electrónico (para pronosticar si un cliente comprará, o no, cierto producto), entre otros .

2.5. Importancia de las variables en la predicción del riesgo

El concepto de la importancia de las variables en la predicción del riesgo es muy usado en modelos tanto de regresión como de clasificación y otros modelos de machine learning. No obstante, es un concepto difícil de explicar debido a su difícil medición y su complejidad.

Este concepto se resumir en ver cómo afecta a la salida del modelo cuando se realizan modificaciones en las variables de entrada. Las variables de entrada que más variabilidad produzcan en la salida serán aquellas que más influencia tengan y, por tanto, serán esas variables que mejor explicarán el modelo y serán las más importantes. Para medir la importancia de dichas variables, se va a poder realizar de dos maneras:

- Reducción de la impuridad nodal media: En cada árbol a la hora de realizar el *División/Corte*, se mide la reducción en la impuridad que contribuye la variable elegida en la partición. Haciendo el promedio sobre todos los arboles, de todas las reducciones de todas las variables, se saca el promedio de este valor de impuridad, se utiliza la medida del índice de Gini. La variable que más reduzca la impuridad del bosque será la más importante.
- Incremento del Mean Squared Error (MSE): El MSE, traducido al castellano es como el error cuadrático medio, es una medida muy usada en estadística como medida de calidad de un estimador. Para todas las respuestas, se calcula el error de la siguiente manera, denotando y como la respuesta real e \hat{y} como la variable predicha.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y - \hat{y})^2$$

Entre los dos métodos de medida de la importancia de las variables, el método más utilizado, es la medida del incremento del MSE, ya que supone un tiempo computacional menor que el cálculo de la reducción de la impuridad nodal.

En el caso de Random Forest, para medir la importancia de las variables, para cada árbol que forman el bosque, se calcula el MSE con las variables. A continuación, se realiza una permutación aleatoria en una de las variables de entrada. Esto produce un cambio en las y de salida que a su vez produce un cambio en el MSE de las variables que han quedado fuera:

$$MSE(X_j) = \frac{1}{n} \sum_{i=1}^n (y - \hat{y}(X_j))^2$$

Se realizará este proceso con todas las variables de entrada y en todos los árboles del bosque, permutando aleatoriamente. Para cada variable permutada, se compara el valor del MSE antes y después de permutar. Esta diferencia, para cada árbol se suma, normaliza y se hace el promedio. Cuanto mayor sea este valor, mas importante será la variable permutada.

3. Análisis descriptivo e inferencial

Una vez definido el contexto de estudio, así como las bases metodológicas y estadísticas empleadas, en este capítulo se realizará el análisis descriptivo e inferencial sobre los datos que se han obtenido mediante la encuesta ErgoSurgeon que se ha mencionado en capítulos anteriores. El objetivo central de este capítulo es describir detalladamente la muestra obtenida y realizar un primer análisis exploratorio que permita detectar asociaciones significativas entre las variables analizadas y la presencia de dolor musculoesquelético (DME).

Para ello, el análisis se ha estructurado en dos etapas principales. La primera etapa, correspondiente al análisis descriptivo, estará compuesto por un resumen exhaustivo de las variables sociodemográficas (como la edad, género, altura, etc), variables laborales (tipo de hospital, especialidad, años de experiencia profesional, etc) y variables relacionadas con hábitos ergonómicos y clínicos (frecuencia de intervenciones quirúrgicas, tipo de cirugía realizada, presencia y duración del dolor musculoesquelético). La segunda etapa, correspondiente al análisis inferencial, tiene como objetivo evaluar las posibles relaciones estadísticas entre estas variables y la ocurrencia del dolor musculoesquelético mediante pruebas estadísticas como la prueba del chi-cuadrado para variables categóricas y el test t de Student para variables numéricas, dependiendo de las características de cada variable analizada.

Antes de presentar estos resultados, y con el fin de proporcionar mayor claridad al análisis, se expone a continuación un diccionario de variables. Esta herramienta incluye una breve descripción de cada una de las variables empleadas en la encuesta ErgoSurgeon, facilitando así la interpretación precisa de los resultados obtenidos en este capítulo en los posteriores análisis de modelos predictivos.

Para el estudio requiere de la creación y la agrupación de varias variables que no se encuentran en las bases de datos mencionadas como por ejemplo el total de cirugías en un año, índice de masa corporal, experiencia, etc. Todas estas variables son creadas en función de lo que se busca, y son añadidas al resto de variables disponibles.

La siguiente tabla se muestran las principales variables que se van a utilizar en el estudio, junto a su descripción y tipología.

Nombre	Definición	Continua	Categórica
EDAD	Edad del cirujano/a participante en años.	x	
GENERO	Género del participante (masculino, femenino o no binario).		x
BMI	Índice de masa corporal	x	
Total_Cirugias_Año	Número total de cirugías en un año.	x	
mano_dominante	Mano dominante utilizada por el participante (derecha, izquierda o ambas)		x
Ambito	Ámbito laboral del cirujano.		x
experiencia	Años de experiencia laboral como cirujano especialista.		x

dias_30_min_Af	Días por semana con al menos 30 minutos de actividad física.		x
dias_quirofano	Días semanales promedio de trabajo en quirófano.		x
laparoscopica	Frecuencia anual de cirugías laparoscópicas realizadas.		x
menor	Frecuencia anual de cirugías menores realizadas.		x
abierta	Frecuencia anual de cirugías abiertas realizadas.		x
endoscopica	Frecuencia anual de cirugías endoscópicas realizadas.		x
robotica	Frecuencia anual de cirugías robóticas realizadas.		x
guardias	Realización habitual de guardias médicas (sí/no).		x
dolor12m	Dolor musculoesquelético atribuido al trabajo en quirófano en los últimos 12 meses.		x
postura	Adopción habitual de posturas incomodas en quirófano.		x
Cuello	Presencia de dolor musculoesquelético en el cuello.		x
Hombros	Presencia de dolor musculoesquelético en los hombros.		x
Zona_dorsal	Presencia de dolor musculoesquelético en la zona dorsal.		x
Lumbares	Presencia de dolor musculoesquelético en la zona lumbar.		x
Muñeca	Presencia de dolor musculoesquelético en la muñeca.		x
Manos_y_dedos	Presencia de dolor musculoesquelético en manos y dedos.		x
Rodillas	Presencia de dolor musculoesquelético en las rodillas.		x
Tobillo_pie	Presencia de dolor musculoesquelético en tobillos y pies.		x
necesitado_asistencia	Necesidad de asistencia médica por dolor musculoesquelético (sí/no).		x
molestias_baja	Si las molestias han causado baja laboral o pérdidas de días (sí/no).		x
formación_ergo	Recepción de formación específica en ergonomía quirúrgica (sí/no).		x
mala_postura	Identificación propia de malas posturas en quirófano (sí/no).		x
factores_causa_molestias	En qué medida considera que han podido influir en la aparición de sus molestias o síntomas musculoesqueléticos por las horas seguidas en cirugía.		x

Tipos_de_cirugia_realizo	En qué medida considera que han podido influir en la aparición de sus molestias o síntomas musculoesqueléticos por el tipo de cirugía.		x
Duración_de_las_intervenciones	En qué medida considera que han podido influir en la aparición de sus molestias o síntomas musculoesqueléticos por la duración de las intervenciones.		x
Instrumental_quirurgico	En qué medida considera que han podido influir en la aparición de sus molestias o síntomas musculoesqueléticos por el instrumental quirúrgico adecuado.		x
Ubicación_de_pantallas_laparoscopia	En qué medida considera que han podido influir en la aparición de sus molestias o síntomas musculoesqueléticos por la ubicación de las pantallas laparoscopia.		x
uso_equipo_provoca_sintomas	Si el uso habitual del equipo quirúrgico genera síntomas (si/no).		x
Conoce_recomendaciones	Conocimiento de recomendaciones sobre ergonomía (si/no)		x
formación_suficiente	Si la formación en ergonomía es suficiente (si/no).		x

Tabla 3. Diccionario de variables

3.1. Descriptivos

La siguiente tabla presenta las características descriptivas más relevantes de la muestra según el género del cirujano, mostrando medianas y cuartiles (Q1,Q3) para variables numéricas y frecuencias absolutas con porcentajes para variables categóricas. Se comparan mediante distribuciones mediante la prueba de Wilcoxon para las variables cuantitativas y el test de Chi-cuadrado de Pearson para variables categóricas determinando así si existen diferencias significativas entre géneros.

Variable	N	Mujer N = 381 ¹	Hombre N = 268 ¹	p-value ²
EDAD	649	39 (31, 49)	49 (37, 59)	<0.001
TALLA_m	649	1.65 (1.60, 1.69)	1.77 (1.73, 1.82)	<0.001
mano_dominante	649			>0.9
Derecha		349 (92%)	243 (91%)	
Izquierda		18 (4.7%)	14 (5.2%)	
Ambas		14 (3.7%)	11 (4.1%)	
BMI	649	22.5 (20.3, 24.5)	25.2 (23.9, 27.8)	<0.001
ambito	649			<0.001
Público		324 (85%)	169 (63%)	
Privado		12 (3.1%)	18 (6.7%)	
Ambas		45 (12%)	81 (30%)	

guardias	649	<0.001
No	342 (90%)	204 (76%)
Sí	39 (10%)	64 (24%)
experiencia	649	<0.001
MIR 1-3er año	53 (14%)	19 (7.1%)
MIR 4º y 5º año	26 (6.8%)	14 (5.2%)
<10	113 (30%)	54 (20%)
10-20	83 (22%)	46 (17%)
>20	106 (28%)	135 (50%)
actividad_física	649	0.11
No realiza	41 (11%)	33 (12%)
Moderada	172 (45%)	99 (37%)
Intensiva	168 (44%)	136 (51%)
Total_Cirugias_Año	649	0.017
<25	15 (3.9%)	5 (1.9%)
25-50	45 (12%)	17 (6.3%)
>50	321 (84%)	246 (92%)

¹ Median (Q1, Q3); n (%)
² Wilcoxon rank sum test; Pearson's Chi-squared test

Tabla 4. Características de la muestra según el Género

En cuanto a las variables **cuantitativas**, se observan diferencias significativas por género en la edad, la talla y el índice de masa corporal (BMI). Los cirujanos hombres presentan una mayor edad mediana (49 años frente a 39 en mujeres), mayor altura (1.77 m frente a 1.65 m), y mayor BMI (25.2 frente a 22.5). Estas diferencias indican perfiles antropométricos distintos, lo que podría implicar diferentes tipos y niveles de estrés físico y ergonómico durante las actividades quirúrgicas.

En relación con las variables **cualitativas**, destacan las diferencias significativas en el ámbito profesional y experiencia laboral. Las mujeres trabajan predominantemente en el sector público (85%), mientras que los hombres tienen una mayor representación en el ámbito mixto y privado. Además, los hombres realizan más guardias hospitalarias (24% frente al 10% en mujeres) y poseen mayor experiencia profesional acumulada (50% con más de 20 años frente al 28% en mujeres), lo que podría influir en la exposición acumulativa a factores de riesgo ergonómicos.

La siguiente tabla se mostrará un análisis descriptivo de las características analizadas en la tabla anterior en función de la presencia o ausencia de dolor musculoesquelético en los últimos 12 meses.

Variable	N	No N = 59 ¹	Sí N = 590 ¹	p-value ²
GENERO	649			0.034
Mujer		27 (46%)	354 (60%)	
Hombre		32 (54%)	236 (40%)	
EDAD	649	47 (36, 57)	42 (33, 53)	0.037
TALLA_m	649	1.71 (1.63, 1.78)	1.69 (1.63, 1.75)	0.5
mano_dominante	649			0.037
Derecha		49 (83%)	543 (92%)	
Izquierda		5 (8.5%)	27 (4.6%)	
Ambas		5 (8.5%)	20 (3.4%)	
BMI	649	24.4 (22.4, 26.0)	23.9 (21.5, 26.1)	0.15
ambito	649			0.040
Público		42 (71%)	451 (76%)	
Privado		7 (12%)	23 (3.9%)	
Ambas		10 (17%)	116 (20%)	
guardias	649			0.3
No		47 (80%)	499 (85%)	
Sí		12 (20%)	91 (15%)	
experiencia	649			0.2
MIR 1-3er año		6 (10%)	66 (11%)	
MIR 4º y 5º año		0 (0%)	40 (6.8%)	
<10		16 (27%)	151 (26%)	
10-20		10 (17%)	119 (20%)	
>20		27 (46%)	214 (36%)	
actividad_fisica	649			0.081
No realiza		3 (5.1%)	71 (12%)	
Moderada		21 (36%)	250 (42%)	
Intensiva		35 (59%)	269 (46%)	
Total_Cirugias_Año	649			0.079
<25		0 (0%)	20 (3.4%)	
25-50		10 (17%)	52 (8.8%)	
>50		49 (83%)	518 (88%)	

¹ n (%); Median (Q1, Q3)

² Pearson's Chi-squared test; Wilcoxon rank sum test; Fisher's Exact Test for Count Data with simulated p-value (based on 2000 replicates)

Tabla 5. Características según Dolor en los Últimos 12 Meses

Al observar esta tabla se puede apreciar que existen diferencias significativas en cuanto a la edad entre los cirujanos que reportaron dolor musculoesquelético en los últimos 12 meses (mediana de 42 años) y aquellos que no (mediana de 47 años), mostrando una tendencia a que los cirujanos con dolor son más jóvenes. En las variables cuantitativas como la talla y el índice de masa corporal, no se aprecian diferencias estadísticamente significativas entre ambos grupos, lo cual indica que estos factores físicos no presentan una asociación directa con la presencia de dolor musculoesquelético.

En cuanto a las variables categóricas, se puede identificar diferencias significativas como el género, con una mayor prevalencia de dolor musculoesquelético entre las mujeres (60%) en comparación con los hombres (40%), y la mano dominante, en donde los cirujanos diestros presentan una prevalencia mayor de dolor (92%) en comparación con los zurdos y ambidiestros. Además, el ámbito laboral presenta diferencias significativas, mostrando que los profesionales públicos o mixtos tienen una prevalencia mayor que los que trabajan en entornos privados. El resto de las variables cualitativas analizadas no mostraron diferencias estadísticamente relevantes.

A continuación, se mostrará la tabla de prevalencia de dolor musculoesquelético en diferentes zonas del cuerpo según el género, destacando los porcentajes de afectación en mujeres y hombres. Además, se emplea una escala de colores para reflejar la intensidad del dolor en función de su frecuencia.

Prevalencia de Dolor Musculoesquelético por Zona y Género

Zona del Cuerpo	Mujeres (%)	Hombres (%)
Cuello	87.7	79.5
Hombros	78.7	68.3
Zona_Dorsal	68.0	59.7
Lumbares	83.2	83.6
Muñeca	51.7	43.3
Manos_Dedos	65.4	53.7
Rodillas	38.8	33.2
Tobillo_Pie	20.5	19.4

Tabla 6. Prevalencia de Dolor Musculoesquelético por Zona y Género

Los datos revelan que las zonas más afectadas son el cuello y el área lumbar, con porcentajes superiores al 75% en ambos géneros. En particular el 87.7% de las mujeres y el 79.5% de los hombres reportan dolor en el cuello, mientras que en la zona lumbar presentan cifras similares en ambos géneros (83.2% de mujeres y 83.6% en hombres), por lo que esto nos permite identificar que áreas son especialmente vulnerables.

En un segundo nivel de prevalencia, con porcentajes que oscilan entre el 50% y el 75%, se encuentran los hombros, la zona dorsal y las manos y dedos. En estos casos, también se observa una mayor incidencia en mujeres en comparación con los hombres. Las muñecas y las rodillas presentan una prevalencia intermedia, con valores entre el 25% y el 50%. El 51.7% de las mujeres reporta dolor en la muñeca, frente al 43.3% de los hombres, mientras que en las rodillas las cifras son del 38.8% y 33.2%, respectivamente.

Finalmente, la menor prevalencia de dolor se observa en los tobillos y pies, donde menos del 25% de los encuestados reporta molestias. Con un 20.5% en mujeres y un 19.4% en hombres, esta es la única zona en la que la diferencia de género es mínima, lo que sugiere que las extremidades inferiores no están tan expuestas a la sobrecarga o a posturas perjudiciales en el ejercicio de la cirugía.

En conclusión, los resultados obtenidos en este análisis descriptivo permiten identificar patrones claros de distribución y prevalencia del dolor musculoesquelético entre cirujanos españoles, resaltando zonas anatómicas especialmente vulnerables y destacando diferencias significativas en función del género y otros factores demográficos y laborales. Estos hallazgos subrayan la necesidad de implementar estrategias ergonómicas y preventivas específicas para reducir la incidencia de estas dolencias en el colectivo quirúrgico. Además, esta información servirá como base para futuras investigaciones y para el desarrollo de modelos predictivos que faciliten la detección temprana de riesgos laborales en cirugía.

3.2. Distribución de los factores de riesgo y prevalencia de dolor.

En este apartado, se va a realizar un análisis detallado sobre la distribución de los principales factores de riesgo identificados en cirujanos españoles y su relación de prevalencia del dolor musculoesquelético en las distintas zonas corporales estudiadas. En este análisis tiene como objetivo identificar qué factores específicos (demográficos, ergonómicos y profesionales) presentan una asociación significativa con la aparición y prevalencia del dolor musculoesquelético. Para ello, se utilizarán técnicas estadísticas de regresión logística multivariante, aplicando un proceso sistemático de selección de variables basado en criterios estadísticos, lo que permitirá obtener modelos robustos y representativos de la realidad laboral para los cirujanos/as españoles. Los resultados derivados de este análisis servirán para fundamentar la creación de medidas preventivas y estrategias específicas que reduzcan el riesgo ergonómico en la práctica diaria en los cirujanos.

Para la comparación de modelos que tengan la misma variable dependiente, se empleará el Criterio de Información de Akaike (AIC), el cual es una medida de calidad relativa y a la hora de comparar 2 o más modelos, a menor AIC, mejor es el modelo.

Lo primero de todo, el objetivo es encontrar alguna relación entre ciertas variables y que el cirujano sufra o no algún dolor musculoesquelético en algunas de las áreas. Las variables que sean significativas en los modelos de cada área corporal explicarán las posibles causas de que un cirujano sufra algún dolor musculoesquelético.

Para facilitar la interpretación de los parámetros de los modelos (β_i), estos serán transformados para obtener los odds ratios y poder observar cómo afecta cada variable a las probabilidades del modelo.

3.2.1. Cuello

Characteristic	Cuello (OR) ¹	95 % CI	p-value
EDAD	0.95	(0.93, 0.97)	<0.001
GENERO			
Mujer	—		
Hombre	0.76	(0.47, 1.25)	0.3
Duracion_de_las_intervenciones			
Sin Dolor-Leve	—		
Moderada-Importante	3.7	(2.09, 6.42)	<0.001
Ubicación_de_pantallas_laparoscopia			
Sin Dolor-Leve	—		
Moderada-Importante	2.5	(1.55, 4.24)	<0.001
AIC	490		
BIC	512		
Log-likelihood	-240		
Deviance	480		
No. Obs.	649		

¹ OR = Odds Ratio

Tabla 7. Factores asociados con dolor cervical en cirujanos

En el modelo final para el dolor musculoesquelético en la zona del cuello. Entre las variables incluidas, la edad se comporta como un factor protector, siendo estadísticamente significativa (OR = 0.95; IC95%: 0.93–0.97; p < 0.001). Esto significa que, por cada año adicional, disminuye en promedio un 5% la probabilidad de sufrir dolor cervical.

Como se puede observar en la tabla, la variable género no resulta estadísticamente significativa (OR = 0.76; p = 0.3), se ha mantenido en el modelo como variable de ajuste para controlar por posibles diferencias entre hombres y mujeres. Esta práctica es común en estudios epidemiológicos y permite estimar el efecto de otras variables de forma más precisa, reduciendo potenciales sesgos por confusión.

Por otro lado, se observa que una duración moderada o importante de las intervenciones quirúrgicas incrementa notablemente el riesgo de presentar dolor cervical ((OR = 3.7; IC95%: 2.1–6.4; p < 0.001), lo cual resalta la influencia de la exposición prolongada en el quirófano ya que tienen aproximadamente un 270% más de probabilidades de desarrollar dolor en el cuello en comparación con aquellas intervenciones que son cortas. Una ubicación inadecuada de las pantallas laparoscópicas se asocia significativamente con mayor probabilidad de dolor (OR = 2.5; IC95%: 1.6–4.2; p < 0.001), lo que subraya el papel de la ergonomía quirúrgica como determinante de salud musculoesquelética, los cirujanos expuestos a una ubicación inadecuada de las pantallas supone un incremento del 150% en la probabilidad de presentar dolor en el cuello respecto a quienes operan con

una ubicación adecuada de las pantallas, lo que subraya el papel de la ergonomía quirúrgica como determinante clave de salud musculoesquelética.

Matriz de confusión			Métricas del modelo	
	Referencia No	Referencia Sí	Valor	
Predicho No	7	10	Accuracy	0.8299
Predicho Sí	23	154	Kappa	0.2094
			Sensitivity	0.2333
			Specificity	0.9390

Tabla 8. Matriz de confusión y métricas en zona cervical

Con base en la matriz de confusión y las métricas obtenidas a partir del modelo Random Forest para predecir la presencia de dolor musculoesquelético en la zona del cuello, se observa lo siguiente:

El modelo alcanza una precisión global (accuracy) del 82.99%, lo que indica que más del 80% de las predicciones totales fueron correctas. Sin embargo, este valor debe interpretarse con cautela, ya que el conjunto de datos presenta un desequilibrio importante entre clases, siendo mucho más frecuente la clase “Sí” (presencia de dolor). Esta situación se refleja en otras métricas clave: la sensibilidad del modelo es baja (23.33%), lo que significa que el modelo identifica correctamente solo al 23% de los casos que efectivamente no presentan dolor (ya que la clase positiva está definida como “No” por defecto). En cambio, la especificidad es muy alta (93.90%), lo que indica que el modelo es muy eficaz para detectar correctamente los casos con dolor (clase “Sí”).

Esto implica que el modelo tiende a clasificar la mayoría de los casos como si tuvieran dolor (clase “Sí”), y falla en identificar adecuadamente a quienes no lo tienen. Este desequilibrio en la clasificación puede deberse al desbalance en la variable dependiente, donde la clase “Sí” predomina fuertemente. Finalmente, el valor del índice Kappa (0.2094) refleja un acuerdo entre predicción y realidad moderadamente bajo, sugiriendo que, aunque el modelo acierta muchas veces, parte de ese acierto puede deberse al sesgo de clase mayoritaria.

3.2.2. Hombros

Characteristic	Cuello (OR) [†]	95 % CI	p-value
EDAD	0.98	(0.97, 1.00)	0.058
GENERO			
Mujer	—		
Hombre	0.81	(0.49, 1.33)	0.4
PESO_kg	0.99	(0.97, 1.00)	0.086
Duracion_de_las_intervenciones			
Sin Dolor-Leve	—		
Moderada-Importante	2.9	(1.77, 4.81)	<0.001
Total_Cirugias_Año			
<25	—		
25-50	1.4	(0.48, 4.19)	0.5
>50	2.7	(0.99, 6.85)	0.045
Instrumental_quirurgico			
Sin Dolor-Leve	—		
Moderada-Importante	1.5	(1.03, 2.20)	0.037
AIC	702		
BIC	738		
Log-likelihood	-343		
Deviance	686		
No. Obs.	649		
[†] OR = Odds Ratio			

Tabla 9. Factores asociados con dolor de hombros en cirujanos

En el caso de los dolores musculoesqueléticos en la zona de los hombros, se ha identificado varios factores que influyen significativamente a la aparición de dichos dolores. En primer lugar, la duración de las intervenciones quirúrgicas se mostro como un factor de riesgo clave ya que aquellos cirujanos que están expuestos a intervenciones de larga duración presentan un incremento del 190% en la probabilidad de sufrir estos dolores con quienes reportan intervenciones de corta duración. Asimismo, el número total de cirugías también puede se asocia con un mayor riesgo, pero únicamente resulta significativamente a aquellos cirujanos que realicen más de 50 cirugías anualmente presentan un incremento del 170% en la aparición de dolor en los hombres frente a aquellos que realizan menos de 50 cirugías, lo cual sugiere un efecto acumulativo de la actividad quirúrgica sobre esa zona corporal. Por otro lado, el incorrecto uso del instrumental quirúrgico se relaciona con un aumento del 50% en la aparición del dolor frente a aquellas personas que si utilizan el instrumental quirúrgico adecuado.

En contraste, variables como el peso corporal ($p = 0.086$), la edad ($p = 0.058$) y el género ($p = 0.4$) no alcanzan significación estadística, aunque la edad y el peso muestran una tendencia que podría indicar un leve efecto protector. No obstante, se ha mantenido la

variable género como ajuste en el modelo, siguiendo criterios epidemiológicos como se ha mencionado anteriormente.

Estos resultados permiten concluir que tanto la carga de trabajo quirúrgica como las condiciones ergonómicas del entorno son factores clave a tener en cuenta en el desarrollo de estrategias de prevención del dolor musculoesquelético en la zona de los hombros en profesionales de la cirugía.

Matriz de confusión		Métricas del modelo	
	Referencia No	Referencia Si	Valor
Predicho No	4	16	Accuracy 0.6839
Predicho Sí	45	128	Kappa -0.0366
			Sensitivity 0.0816
			Specificity 0.8889

Tabla 10. Matriz de confusión y métricas en hombros

En cuanto a la capacidad predictiva del modelo de Random Forest aplicado para la clasificación de la presencia de dolor musculoesquelético en la zona de los hombros, la clase positiva en este análisis ha sido establecida como "No" (es decir, ausencia de dolor), lo que influye directamente en la interpretación de las métricas. La exactitud del modelo es del 68.39%, lo cual nos indica que aproximadamente 7 de cada 10 casos fueron correctamente, pero el índice Kappa es -0.0366, lo cual evidencia una baja concordancia entre las predicciones del modelo y los valores reales, más allá de lo que se esperaría por azar.

El modelo muestra una sensibilidad muy baja (8.16 %), lo que significa que identifica correctamente solo al 8.16 % de los individuos que no presentan dolor. En cambio, la especificidad es alta (88.89 %), lo cual implica que el modelo identifica correctamente al 88.89 % de los individuos que sí presentan dolor musculoesquelético en la zona de los hombros. Este comportamiento indica que el modelo tiende a sobre predecir la clase "No" (ausencia de dolor), probablemente como resultado de un desbalance de clases en los datos o de la falta de variables predictivas relevantes.

A pesar de que el modelo tiene una precisión aceptable para la clase "No", su utilidad clínica es limitada si el objetivo principal es identificar a quienes están en riesgo de sufrir dolor.

3.2.3. Zona dorsal

Characteristic	Cuello (OR) ¹	95 % CI	p-value
EDAD	0.97	(0.95, 0.98)	<0.001
GENERO			
Mujer	—		
Hombre	0.95	(0.65, 1.39)	0.8
BMI_categoria			
healthy	—		
overweight	0.72	(0.48, 1.08)	0.11
obese	1.5	(0.75, 3.07)	0.3
Duracion_de_las_intervenciones			
Sin Dolor-Leve	—		
Moderada-Importante	1.5	(0.88, 2.61)	0.13
Ubicación_de_pantallas_laparoscopia			
Sin Dolor-Leve	—		
Moderada-Importante	1.6	(1.07, 2.28)	0.022
Instrumental_quirurgico			
Sin Dolor-Leve	—		
Moderada-Importante	1.4	(0.99, 2.10)	0.058
factores_causa_molestias			
Sin Dolor-Leve	—		
Moderada-Importante	1.8	(1.10, 3.00)	0.019
AIC	797		
BIC	838		
Log-likelihood	-390		
Deviance	779		
No. Obs.	649		
¹ OR = Odds Ratio			

Tabla 11. Factores asociados con dolor en zona dorsal en cirujanos

En el modelo final de regresión logística multivariante para el dolor musculoesquelético en la zona dorsal, se identifican varios factores relevantes. La edad se comporta nuevamente como un factor protector, siendo estadísticamente significativa (OR = 0.97; IC95%: 0.95–0.98; p < 0.001). Esto implica que, por cada año adicional de edad, la probabilidad de experimentar dolor dorsal disminuye en un 3% aproximadamente. La variable genero siempre se va a mantener como se ha comentado en anteriores modelos.

Dentro de las variables ergonómicas, se observa que una ubicación moderada o inadecuada de las pantallas laparoscópicas se asocia con un incremento del 60% en la probabilidad de desarrollar dolor en la zona dorsal (OR = 1.6; IC95%: 1.1–2.3; p = 0.022). Este resultado refuerza la importancia de una correcta disposición del equipo en quirófano. Además, la percepción de que los factores relacionados con molestias físicas son moderada o importante también muestra una asociación significativa con el dolor dorsal, incrementando la probabilidad en un 80%.

Otras variables como el instrumental quirúrgico y la duración de las intervenciones muestran una tendencia a ser significativas, pero sus valores p se encuentran ligeramente por encima del umbral convencional de 0.05. No obstante, se sugiere prestar atención a estas variables en futuras investigaciones, ya que podrían tener un papel relevante en contextos más específicos o con mayor tamaño muestral.

Matriz de confusión		Métricas del modelo	
	Referencia No	Referencia Sí	Valor
Predicho No	18	25	Accuracy 0.6082
Predicho Sí	51	100	Kappa 0.0665
			Sensitivity 0.2609
			Specificity 0.8000

Tabla 12. Matriz de confusión y métricas en zona dorsal

En este análisis, la clase positiva para el cálculo de Sensibilidad y Especificidad es 'No Dolor'. Por lo tanto, Sensibilidad se refiere a la proporción de casos sin dolor correctamente identificados, y Especificidad a la proporción de casos con dolor correctamente identificados.

Tras aplicar el modelo de Random Forest para predecir la presencia de dolor musculoesquelético en la zona dorsal, se obtuvo una precisión global (accuracy) del 60.82 %, lo que indica que aproximadamente 6 de cada 10 predicciones fueron correctas. Sin embargo, esta cifra debe interpretarse con cautela, ya que el índice Kappa es de 0.0665, lo cual evidencia un muy bajo grado de concordancia entre las predicciones y los valores reales, apenas por encima de lo que se esperaría por azar.

Dado que en este análisis la clase positiva se ha definido como "No Dolor" (ausencia de dolor dorsal), la sensibilidad del modelo fue del 26.09 %, lo que implica que identifica correctamente solo al 26.09 % de los cirujanos que no presentan dolor dorsal. Por otro lado, la especificidad es del 80 %, lo que significa que el modelo acierta a clasificar correctamente al 80 % de los cirujanos que sí presentan dolor dorsal.

Este comportamiento refleja un claro sesgo hacia la clase "Sí Dolor": el modelo tiende a clasificar erróneamente muchos casos sin dolor como si lo tuvieran, priorizando la detección de la clase mayoritaria. Como en el conjunto de datos la mayoría de los cirujanos presentaban dolor dorsal, el modelo ha aprendido a sobre predecir "Sí Dolor", sacrificando la detección de quienes realmente no tienen dolor.

3.2.4. Lumbares

Characteristic	Cuello (OR) [†]	95 % CI	p-value
EDAD	0.98	(0.96, 1.00)	0.013
PESO_kg	1.0	(1.00, 1.03)	0.070
Duracion_de_las_intervenciones			
Sin Dolor-Leve	—		
Moderada-Importante	2.4	(1.30, 4.26)	0.004
Ubicación_de_pantallas_laparoscopia			
Sin Dolor-Leve	—		
Moderada-Importante	1.5	(0.95, 2.35)	0.087
factores_causa_molestias			
Sin Dolor-Leve	—		
Moderada-Importante	1.7	(0.92, 2.92)	0.086
AIC	560		
BIC	587		
Log-likelihood	-274		
Deviance	548		
No. Obs.	649		

[†] OR = Odds Ratio

Tabla 13. Factores asociados con dolor en zona lumbar en cirujanos

En el modelo final para el dolor musculoesquelético en la zona lumbar, nuevamente se observa que la edad tiene un efecto protector, por cada año adicional que tenga el cirujano, se reduce en promedio un 2% la probabilidad de experimentar dolor en esa zona.

Entre los factores ergonómicos, la duración moderada o importante de las intervenciones quirúrgicas destaca como un factor de riesgo significativo (OR = 2.4; IC95%: 1.3–4.3; p = 0.004). Esto implica que los cirujanos expuestos a intervenciones prolongadas presentan una probabilidad un 140% mayor de sufrir dolor lumbar en comparación con aquellos que realizan intervenciones de corta duración.

Por otro lado, variables como el peso corporal, la ubicación de las pantallas laparoscópicas y los factores que generan molestias durante la actividad quirúrgica mostraron asociaciones positivas con el dolor lumbar, aunque no alcanzaron la significación estadística (p > 0.05). Aun así, sus odds ratios superiores a 1 podrían señalar tendencias que merecen atención en futuros estudios.

Matriz de confusión			Métricas del modelo	
	Referencia No	Referencia Sí	Valor	
Predicho No	4	6	Accuracy	0.8247
Predicho Sí	28	156	Kappa	0.1215
			Sensitivity	0.1250
			Specificity	0.9630

Tabla 14. Matriz de confusión y métricas en zona lumbar

En este análisis, la clase positiva para el cálculo de Sensibilidad y Especificidad es 'No Dolor'. Por lo tanto, Sensibilidad se refiere a la proporción de casos sin dolor correctamente identificados, y Especificidad a la proporción de casos con dolor correctamente identificados.

La matriz de confusión y las métricas obtenidas permiten evaluar el rendimiento del modelo Random Forest para predecir la presencia de dolor en la zona lumbar. El modelo alcanza una precisión general (accuracy) del 82.47%, lo que indica un buen desempeño global.

La sensibilidad (0.125) revela que el modelo identifica correctamente sólo el 12.5% de los casos negativos (es decir, aquellos sin dolor lumbar), mientras que la especificidad (0.9630) indica una excelente capacidad para identificar correctamente a quienes sí presentan dolor (96.3%).

Por otro lado, el índice Kappa (0.1215) sugiere una concordancia leve entre las predicciones del modelo y los valores reales, lo que indica que, aunque el modelo tiene una alta exactitud general, su utilidad para diferenciar adecuadamente entre clases puede ser limitada debido al desequilibrio de clases.

Estos resultados reflejan un buen rendimiento del modelo para identificar casos con dolor lumbar, pero una menor eficacia para detectar correctamente los casos negativos, lo que debe tenerse en cuenta al aplicar este modelo en contextos clínicos o preventivos.

3.2.5. Muñeca

Characteristic	Cuello (OR) [†]	95 % CI	p-value
EDAD	0.97	(0.95, 0.98)	<0.001
GENERO			
Mujer	—		
Hombre	0.89	(0.63, 1.27)	0.5
Duracion_de_las_intervenciones			
Sin Dolor-Leve	—		
Moderada-Importante	1.5	(0.92, 2.62)	0.11
Ubicación_de_pantallas_laparoscopia			
Sin Dolor-Leve	—		
Moderada-Importante	1.3	(0.94, 1.94)	0.10
cirugiaEndo			
No	—		
Sí	1.9	(1.31, 2.77)	<0.001
lInstrumental_quirurgico			
Sin Dolor-Leve	—		
Moderada-Importante	1.9	(1.29, 2.65)	<0.001
AIC	838		
BIC	869		
Log-likelihood	-412		
Deviance	824		
No. Obs.	649		

[†] OR = Odds Ratio

Tabla 15. Factores asociados con dolor en muñecas en cirujanos

En el modelo para la detección de dolores musculoesqueléticos en la muñeca la variable edad continua con un efecto protector con una odds ratio de 0.97 (IC95%: 0.95–0.98; p < 0.001). La variable genero no es significativa, pero sí que se estará en el modelo como variable para la práctica habitual en estudios epidemiológicos.

La realización de cirugía endoscópica se asocia significativamente con mayor probabilidad de dolor en la muñeca (OR = 1.9; IC95%: 1.3–2.8; p < 0.001), lo que representa un aumento del 90% en la probabilidad de presentar dolor en comparación con quienes no realizan este tipo de intervenciones. De igual forma, un uso moderado o importante del instrumental quirúrgico también incrementa significativamente el riesgo (OR = 1.9; IC95%: 1.3–2.7; p < 0.001), sugiriendo que el manejo repetido o exigente del instrumental podría estar vinculado a sobrecarga articular en esta zona.

Aunque la duración de las intervenciones y la ubicación de pantallas laparoscópicas muestran tendencias hacia una asociación positiva (OR = 1.5 y OR = 1.3,

respectivamente), sus valores p (> 0.05) indican que no alcanzan significación estadística, por lo que no se puede concluir una relación clara en este modelo.

Matriz de confusión			Métricas del modelo	
	Referencia No	Referencia Sí	Valor	
Predicho No	55	34	Accuracy	0.5907
Predicho Sí	45	59	Kappa	0.1836
			Sensitivity	0.5500
			Specificity	0.6344

Tabla 16. Matriz de confusión y métricas en muñecas

En este análisis, la clase positiva para el cálculo de Sensibilidad y Especificidad es 'No Dolor'. Por lo tanto, Sensibilidad se refiere a la proporción de casos sin dolor correctamente identificados, y Especificidad a la proporción de casos con dolor correctamente identificados.

Tras aplicar el modelo de Random Forest para predecir la presencia de dolor musculoesquelético en la muñeca, se obtuvo una precisión global (accuracy) del 59.07%, lo que indica que aproximadamente 6 de cada 10 predicciones fueron correctas. El índice Kappa es 0.1836, lo cual evidencia un acuerdo bajo, pero mejor que el azar.

Dado que en este análisis la clase positiva se ha definido como "No Dolor" (ausencia de dolor en la muñeca), la sensibilidad del modelo es del 55 %, lo que implica que identifica correctamente al 55 % de los cirujanos que no presentan dolor en la muñeca. Por otro lado, la especificidad es del 63.44%, lo que significa que el modelo acierta a clasificar correctamente al 63.44 % de los cirujanos que sí presentan dolor en la muñeca.

En conjunto, el rendimiento del modelo es moderado: muestra una capacidad razonable para detectar a quienes no tienen dolor (sensibilidad del 55 %), pero es limitado para identificar a quienes sí lo padecen (especificidad del 63.44 %). Aunque la sensibilidad supera a la consistencia con "No Dolor" como clase positiva, aún queda espacio para mejorar la capacidad predictiva, especialmente en la detección de los casos con dolor.

3.2.6. Manos y dedos

Characteristic	Cuello (OR) [†]	95 % CI	p-value
EDAD	0.98	(0.97, 1.00)	0.021
GENERO			
Mujer	—		
Hombre	0.71	(0.50, 1.01)	0.058
Duracion_de_las_intervenciones			
Sin Dolor-Leve	—		
Moderada-Importante	1.8	(1.11, 3.00)	0.019
Ubicación_de_pantallas_laparoscopia			
Sin Dolor-Leve	—		
Moderada-Importante	1.4	(1.00, 2.07)	0.051
Instrumental_quirurgico			
Sin Dolor-Leve	—		
Moderada-Importante	2.1	(1.46, 3.02)	<0.001
AIC	824		
BIC	851		
Log-likelihood	-406		
Deviance	812		
No. Obs.	649		

[†] OR = Odds Ratio

Tabla 17. Factores asociados con dolor en manos y dedos en cirujanos

En el análisis multivariante para el dolor musculoesquelético en manos y dedos, se identificaron varios factores que presentan asociaciones estadísticamente significativas. En primer lugar, la edad muestra un efecto protector moderado: por cada año adicional, la probabilidad de reportar dolor en esta zona disminuye aproximadamente un 2% (OR = 0.98; p = 0.021), lo que puede estar vinculado a la experiencia y adaptación ergonómica con el tiempo.

Aunque el género masculino no alcanza significación estadística (OR = 0.71; p = 0.058), se observa una tendencia hacia un menor riesgo comparado con las mujeres. Esto sugiere la conveniencia de seguir considerando el género como una posible variable de ajuste.

Por otro lado, se evidencia que una duración prolongada de las intervenciones quirúrgicas incrementa en un 80% la probabilidad de experimentar dolor en manos y dedos (OR = 1.8; p = 0.019). Además, una ubicación moderadamente inadecuada de las pantallas laparoscópicas se asocia con un riesgo aumentado (OR = 1.4; p = 0.051), aunque en este caso el valor p roza el umbral de significación.

La variable con mayor impacto es el uso de instrumental quirúrgico percibido como inadecuado o incómodo, que multiplica por 2.1 el riesgo de padecer dolor en esta zona (p < 0.001), lo que resalta el papel clave de la ergonomía instrumental en la prevención del dolor.

Métricas del modelo			
Matriz de confusión		Valor	
	Referencia No	Referencia Sí	
Predicho No	27	18	Accuracy 0.6528
Predicho Sí	49	99	Kappa 0.2169
			Sensitivity 0.3553
			Specificity 0.8462

Tabla 18. Matriz de confusión y métricas en manos y dedos

En este análisis, la clase positiva para el cálculo de Sensibilidad y Especificidad es 'No Dolor'. Por lo tanto, Sensibilidad se refiere a la proporción de casos sin dolor correctamente identificados, y Especificidad a la proporción de casos con dolor correctamente identificados.

En cuanto al rendimiento del modelo Random Forest aplicado a la zona de manos y dedos, se observa una precisión global (accuracy) del 65,3%, lo que indica que aproximadamente dos de cada tres casos son correctamente clasificados. El índice Kappa (0.2169) sugiere un acuerdo discreto entre las predicciones del modelo y los valores reales, una mejora moderada respecto al azar.

La sensibilidad del modelo, que mide la capacidad para detectar correctamente a quienes no presentan dolor (clase positiva definida como "No"), es baja (35,5%), lo que indica que muchos individuos sin dolor fueron clasificados incorrectamente como con dolor. En cambio, la especificidad es elevada (84,6%), reflejando que el modelo es bastante efectivo al identificar correctamente a quienes sí experimentan dolor en esta zona.

Este comportamiento sugiere que el modelo es más fiable para detectar la presencia de dolor en manos y dedos, pero presenta limitaciones para detectar su ausencia.

3.2.7. Rodillas

Characteristic	Cuello (OR) [†]	95 % CI	p-value
GENERO			
Mujer	—		
Hombre	0.77	(0.55, 1.07)	0.12
Duracion_de_las_intervenciones			
Sin Dolor-Leve	—		
Moderada-Importante	2.2	(1.28, 4.04)	0.006
Ubicación_de_pantallas_laparoscopia			
Sin Dolor-Leve	—		
Moderada-Importante	1.4	(1.03, 1.99)	0.034
AIC	842		
BIC	860		
Log-likelihood	-417		
Deviance	834		
No. Obs.	649		

[†] OR = Odds Ratio

Tabla 19. Factores asociados con dolor en rodillas en cirujanos

En el modelo final estimado para el dolor musculoesquelético en la zona de las rodillas, se identificaron dos variables estadísticamente significativas. En primer lugar, la duración de las intervenciones quirúrgicas mostró una asociación clara con la aparición del dolor. Aquellos cirujanos que reportaron intervenciones de duración moderada o importante presentan un 120% más de probabilidad de experimentar dolor en las rodillas en comparación con quienes realizan intervenciones más cortas (OR = 2.2; IC95%: 1.3–4.0; p = 0.006).

Asimismo, una ubicación inadecuada o moderadamente importante de las pantallas laparoscópicas se asocia con un aumento del 40% en la probabilidad de padecer dolor en esta zona (OR = 1.4; IC95%: 1.0–2.0; p = 0.034), lo que refuerza el papel de la ergonomía como factor de riesgo relevante.

Aunque la variable género no alcanza significación estadística (p = 0.12), se mantuvo en el modelo como variable de ajuste (OR = 0.77), en línea con los criterios epidemiológicos que buscan controlar posibles factores de confusión entre hombres y mujeres. Esto permite una mejor estimación del efecto real de las variables ergonómicas y profesionales analizadas.

Matriz de confusión			Métricas del modelo	
	Referencia No	Referencia Sí	Valor	
Predicho No	123	71	Accuracy	0.634
Predicho Sí	0	0	Kappa	0.000
			Sensitivity	1.000
			Specificity	0.000

Tabla 20. Matriz de confusión y métricas en rodillas

En este análisis, la clase positiva para el cálculo de Sensibilidad y Especificidad es 'No Dolor'. Por lo tanto, Sensibilidad se refiere a la proporción de casos sin dolor correctamente identificados, y Especificidad a la proporción de casos con dolor correctamente identificados.

El modelo clasificó todos los casos de prueba como "No Dolor". La sensibilidad (para la clase "No Dolor") es del 100 %, lo que significa que todos los individuos que realmente no tenían dolor fueron correctamente identificados como "No Dolor". Sin embargo, la especificidad (para la clase "Sí Dolor") es del 0 %, lo que significa que ninguno de los individuos que realmente sí tenían dolor fue correctamente identificado como tal; todos ellos fueron incorrectamente clasificados como "No Dolor".

Aunque la precisión global (accuracy) es del 63,4 %, este valor resulta engañoso debido al fuerte desequilibrio en la clasificación. El índice Kappa es 0, lo que sugiere que el modelo no aporta mejora alguna respecto a una clasificación aleatoria.

3.2.8. Tobillos y pies

Characteristic	Cuello (OR) ¹	95 % CI	p-value
EDAD	0.97	(0.95, 0.98)	<0.001
GENERO			
Mujer	—		
Hombre	0.59	(0.34, 0.99)	0.050
PESO_kg	1.0	(1.02, 1.06)	<0.001
Duracion_de_las_intervenciones			
Sin Dolor-Leve	—		
Moderada-Importante	5.1	(1.79, 21.71)	0.008
Ubicación_de_pantallas_laparoscopia			
Sin Dolor-Leve	—		
Moderada-Importante	1.9	(1.24, 2.85)	0.003
factores_causa_molestias			
Sin Dolor-Leve	—		
Moderada-Importante	2.1	(0.99, 5.12)	0.068
AIC	605		
BIC	636		
Log-likelihood	-295		
Deviance	591		
No. Obs.	649		

¹ OR = Odds Ratio

Tabla 21. Factores asociados con dolor en tobillos y pies en cirujanos

El modelo final para el dolor musculoesquelético en la zona del tobillo y pie revela varias asociaciones significativas con factores individuales y laborales. En primer lugar, la edad actúa como un factor protector (OR = 0.97; IC95%: 0.95–0.98; p < 0.001), lo que implica que, por cada año adicional, la probabilidad de sufrir dolor en esta zona disminuye aproximadamente un 3%. Además, el género masculino se asocia con una menor probabilidad de experimentar este tipo de dolor en comparación con el femenino (OR = 0.59; IC95%: 0.34–0.99; p = 0.050), lo que indica una reducción del 41% en el riesgo para los hombres.

El peso corporal también muestra una asociación significativa, aunque sutil, con un OR de 1.0 (IC95%: 1.0–1.1; p < 0.001), lo que sugiere que un aumento en el peso podría tener un efecto leve pero estadísticamente significativo sobre el riesgo de dolor en tobillo y pie.

Entre los factores ergonómicos, la duración moderada o importante de las intervenciones quirúrgicas destaca como el principal predictor del dolor en esta zona (OR = 5.1; IC95%: 1.8–22; p = 0.008), indicando que estos profesionales tienen un 410% más de probabilidad de sufrir dolor en comparación con aquellos con intervenciones cortas. Asimismo, una ubicación inadecuada de las pantallas laparoscópicas se relaciona con un incremento del 90% en la probabilidad de dolor (OR = 1.9; IC95%: 1.2–2.8; p = 0.003), reforzando la importancia del entorno de trabajo quirúrgico. Por último, los factores que

causan molestias, cuando se consideran moderada o importante, tienden a aumentar el riesgo de dolor en un 110%, aunque este resultado no alcanza significación estadística convencional (OR = 2.1; IC95%: 1.0–5.1; p = 0.068), por lo que debe interpretarse con cautela.

Métricas del modelo			
		Valor	
Matriz de confusión			
	Referencia No	Referencia Sí	
Predicho No	141	33	
Predicho Sí	14	6	
		Accuracy	0.7577
		Kappa	0.0777
		Sensitivity	0.9097
		Specificity	0.1538

Tabla 22. Matriz de confusión y métricas en tobillos y pies

En este análisis, la clase positiva para el cálculo de Sensibilidad y Especificidad es 'No Dolor'. Por lo tanto, Sensibilidad se refiere a la proporción de casos sin dolor correctamente identificados, y Especificidad a la proporción de casos con dolor correctamente identificados.

La matriz de confusión y las métricas del modelo Random Forest para predecir el dolor musculoesquelético en la zona de tobillo y pie muestran un rendimiento desigual en la clasificación de los casos. El modelo presenta una precisión general (accuracy) del 75.77%, lo que indica que tres de cada cuatro observaciones fueron correctamente clasificadas. No obstante, el valor Kappa es bajo (0.0777), lo que evidencia un escaso acuerdo entre las predicciones del modelo y los datos reales más allá del azar.

La sensibilidad, que mide la capacidad del modelo para identificar correctamente los casos positivos (cirujanos sin dolor, ya que la clase positiva está definida como "No"), es del 90.97%, lo que implica que el modelo acierta en la gran mayoría de los casos en los que no se reporta dolor. Por el contrario, la especificidad, que refleja la capacidad del modelo para detectar correctamente a quienes sí presentan dolor, es muy baja (15.38%), lo que indica que el modelo falla al reconocer adecuadamente los casos con dolor en esta zona.

En conjunto, aunque el modelo predice bien la ausencia de dolor, no es fiable para identificar correctamente a quienes lo sufren, lo que limita su aplicabilidad clínica para la detección de casos positivos.

4. Desarrollo de la calculadora de riesgos

4.1. Fundamentos teóricos para la estimación del riesgo

La calculadora desarrollada en este trabajo tiene como base teórica fundamental el modelo de regresión logística, una herramienta estadística utilizada ampliamente para la predicción y estimación de probabilidades en estudios. En este contexto en concreto, la regresión logística va a permitir predecir la probabilidad de que un cirujano experimente dolor musculoesquelético en distintas zonas corporales, basándose en múltiples variables demográficas, laborales y ergonómicas.

Para desarrollar la calculadora se han empleado los coeficientes estimados en cada uno de los modelos de regresión logística específicos para cada zona corporal estudiada (cuello, hombros, dorsal, lumbares, muñeca, manos y dedos, rodillas, tobillo y pie). Estos coeficientes se aplican en la siguiente fórmula logística para obtener la probabilidad de experimentar dolor musculoesquelético en cada zona:

$$\eta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

Representa la combinación lineal de las variables predictoras por sus respectivos coeficientes. Y para el cálculo de la probabilidad:

$$P(Y = 1|X) = \frac{e^\eta}{1 + e^\eta} = \frac{1}{1 + e^{-\eta}}$$

Para facilitar la comprensión, se presenta a continuación un ejemplo práctico de este cálculo. Consideramos por ejemplo un cirujano hombre de 40 años, con peso de 70 kg y talla 1,70 m, que realiza más de 50 cirugías al año, siempre realiza medidas ergonómicas como la reducción del tiempo en quirófano, la colocación del monitor y la realización de micro descansos. Utilizando los coeficientes obtenidos del modelo (son distintos a los OR que hemos explicado anteriormente) específico para la zona del cuello: la calculadora generaría la siguiente combinación lineal:

$$\eta_{cuello} = 2.8618 + (-0.2679) \times 1 + (-0.0536) \times 40 + (1.3006) \times 0 + (0.93109) \times 0$$

Posteriormente, esta combinación lineal se convierte en una probabilidad mediante la fórmula logística, proporcionando una estimación del riesgo para el cirujano.

En definitiva, la aplicación del modelo de regresión logística como base teórica para esta calculadora permite obtener estimaciones objetivas y prácticas del riesgo musculoesquelético en cirujanos. De este modo, se facilita la identificación temprana de situaciones de riesgo, posibilitando una intervención preventiva efectiva, contribuyendo así a la mejora en la salud laboral y la calidad de vida del personal quirúrgico.

4.2. Implementación y visualización de la calculadora en Shiny

La calculadora diseñada para estimar el riesgo de dolor musculoesquelético se presenta como una aplicación web creada con Shiny, un paquete de R que permite diseñar interfaces reactivas directamente desde el entorno estadístico. Con esta herramienta, el cirujano introduce sus datos personales (edad, género, peso y talla) y sus hábitos ergonómicos (por ejemplo, breves pausas intraoperatorias o disposición de pantallas). Una vez llenado todos los campos pulsando el botón de cálculo, la aplicación transforma esos valores en un vector numérico, aplica los coeficientes logísticos obtenidos en el análisis estadístico y calcula la probabilidad de dolor en cada zona corporal.

La interfaz se ha estructurado de forma fluida: un primer bloque recoge los datos demográficos, seguido de secciones para las prácticas ergonómicas y la actividad quirúrgica. Cada campo incorpora validaciones de rango y texto guía para garantizar la fiabilidad de la entrada. El resultado se muestra en pantalla como una tabla de probabilidades (expresadas en porcentaje).

The screenshot shows a Shiny application window titled 'Calculadora'. The interface is divided into three main sections:

- Información Básica:** Contains fields for 'Edad' (40), 'Género' (radio buttons for 'Hombre' and 'Mujer' with 'Hombre' selected), 'Peso (kg)' (85), and 'Altura (m)' (1.75).
- Medidas para reducir el dolor:** A section with two groups of questions and radio buttons:
 - ¿Intentas reducir tiempo en quirófano?
• Nunca (selected), A veces, A menudo, Siempre
 - ¿Ajustas el monitor a altura adecuada?
• Nunca, A veces, A menudo, Siempre
- Información sobre Cirugías:** Contains fields for 'Total de Cirugías al Año' (radio buttons for '<25' (selected), '25-50', and '>50') and '¿Realizas cirugías endoscópicas?' (radio buttons for 'No' (selected) and 'Sí').

At the bottom is a blue 'Calcular' button with a gear icon.

Figura 6. Calculadora de prevalencia de dolores musculoesqueléticos

Tras pulsar el botón “Calcular”, la aplicación procesa los datos de entrada y muestra inmediatamente, debajo del formulario, una tabla de probabilidades estimadas de presentar

dolor moderado-severo en cada zona corporal analizada. La tabla se presenta en dos columnas.

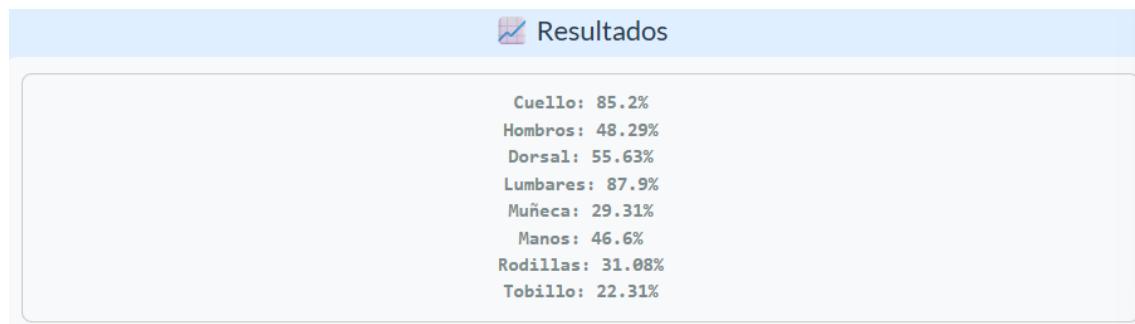


Figura 7. Resultados calculadora de prevalencia de dolores

En las zonas de Cuello (85,2%) y zona lumbar (87,9%) se sitúan muy por encima del umbral de alerta. Deberían priorizarse estrategias como descansos intraoperatorios más frecuentes, ajustes ergonómicos de la mesa de operaciones y de las pantallas, así como entrenamiento específico en posturas cervicales y lumbares. En la zona dorsal (63,6%) también requiere atención, implementando estiramientos dirigidos y revisión de la altura de los apoyabrazos. En los hombros (48,3%) y manos (46,6%) se encuentran en el límite superior de riesgo moderado. Se recomienda alternar labores de precisión con acciones de menor demanda muscular, y valorar el uso de instrumental con mango ergonómico. En las rodillas (31,1%) y muñeca (29,3%) están en el tramo bajo-moderado: seguir monitorizando su evolución y fomentar ejercicios de estiramiento y fortalecimiento fuera del quirófano. Y por último el tobillo (22,3%) ocupa la categoría de bajo riesgo, aunque conviene mantener la higiene postural y supervisar posibles incrementos en cirugías de pie prolongadas.

La aplicación ha sido desplegada en la plataforma Shiny (shinyapps.io) bajo la URL <https://ivanmena.shinyapps.io/calculadora-dolor> y puede utilizarse directamente desde cualquier navegador, sin necesidad de instalar nada adicional. De este modo, cualquier cirujano puede acceder a la herramienta, introducir sus datos y obtener al instante su perfil de riesgo de dolor musculoesquelético.

5. Conclusiones y recomendaciones

El presente trabajo ha permitido identificar y evaluar rigurosamente los factores ergonómicos y laborales puede revelar patrones de riesgo asociados al desarrollo de dolor musculoesquelético en cirujanos, un colectivo que a menudo permanece fuera del foco de la investigación en salud ocupacional. Mediante el análisis de datos recogidos a través de la encuesta ErgoSurgeon, se han implantado técnicas clásicas de regresión logística con algoritmos de Random Forest, lo que ha permitido capturar tanto el peso individual de cada variable.

Los resultados han puesto de manifiesto una alta prevalencia de dolor musculoesquelético entre los cirujanos españoles, con más del 80% de los participantes que presentan molestias en las cervicales y lumbares, y tasas cercanas al 75% en la zona de hombros, lo que confirma la existencia de un problema de salud laboral. Además de todos los cirujanos que han reportado que han sufrido molestias, un gran porcentaje de estos cirujanos afirma que han tenido que recurrir a algún tipo de asistencia para reducir su dolor e incluso perdidas de días de días laborales, lo que evidencia no solo el impacto inmediato sobre la productividad y la calidad de vida de los cirujanos. En el análisis de regresión logística ha identificado la duración prolongada de las intervenciones, la ubicación inadecuada de las pantallas laparoscópicas y el uso instrumental no adaptado como los principales factores de riesgo asociados al dolor musculoesquelético. Estos hallazgos ponen de manifiesto la necesidad de introducir de manera sistemática formación en ergonomía desde la residencia, la adaptación personalizada del entorno de quirófano (altura de mesas y pantallas, selección de instrumental) y la instauración de microdescansos intraoperatorios. Asimismo, la calculadora de riesgos desarrollada en Shiny se presenta como una herramienta práctica que, al integrar los factores más influyentes, permite diseñar estrategias de prevención individualizadas para preservar la salud musculoesquelética de los cirujanos a lo largo de toda su carrera.

Además, nuestro estudio ha puesto de manifiesto que las diferencias de género y de años de experiencia modulan de forma significativa la percepción y la gravedad de las molestias musculoesqueléticas: las cirujanas y los profesionales con menos años de residencia muestran una mayor susceptibilidad al dolor cervical y lumbar, probablemente vinculada tanto a diferencias fisiológicas en la percepción del dolor como a una adaptación insuficiente del mobiliario y del instrumental a sus características antropométricas. Por lo que es de total importancia extender los programas de formación ergonómica, y de certificar salas de cirugía que garanticen estándares mínimos de adaptabilidad para todo el personal. Por otro lado, se exige protocolos de microdescansos y ejercicios de movilidad activa intraoperatoria resultara clava para reducir la cronificación de dolor y prolongar en condiciones saludables de los cirujanos.

Antes de iniciar la intervención, resulta aconsejable efectuar ejercicios de activación muscular y articular, incidiendo especialmente en cervicales, zona lumbar, hombros, muñecas y dedos. Asimismo, una adecuada ingesta de alimentos e hidratación previa contribuye no solo al correcto funcionamiento de la musculatura, sino también a mitigar el “síndrome de fatiga quirúrgica” que aparece en procedimientos de larga duración. Dado que las posturas forzadas y la ausencia de movilidad ergonómica en el quirófano son factores desencadenantes de dolor musculoesquelético, la preparación del entorno de trabajo adquiere gran relevancia. Entre las medidas intraoperatorias más eficaces se encuentran la regulación óptima de la altura de la mesa, el ajuste preciso de lupas y

lámparas frontales, la adaptación del instrumental y la colocación de monitores a una distancia ergonómica. Varias revisiones especializadas aportan pautas según cada zona anatómica y equipo utilizado. Cambiar periódicamente la posición de bipedestación y redistribuir el peso favorece la contracción muscular equilibrada y el retorno venoso; además, mantener una leve flexión de rodillas y activar la musculatura profunda del abdomen estabiliza la región lumbar. La realización de movimientos activos de cuello, hombros y escápulas previene patrones motores inadecuados y fatiga ocular. Numerosos estudios confirman que estos ejercicios específicos o “microdescansos” no entorpecen la práctica quirúrgica ni comprometen la seguridad del paciente. Tras la operación, es beneficioso realizar estiramientos de los grupos musculares más cargados. Programas de estiramiento en otros colectivos (por ejemplo, trabajadores de oficina) han demostrado resultados positivos en la reducción del dolor musculoesquelético.

En definitiva, la adopción de un protocolo ergonómico integral que abarque desde el calentamiento previo y la organización del entorno de quirófano hasta la realización de microdescansos intraoperatorios y estiramientos posteriores, constituye una estrategia clave para minimizar el riesgo de dolor musculoesquelético en los cirujanos. La implementación de estas prácticas, unida a la formación continuada y al uso de herramientas predictivas como la calculadora de riesgos desarrollada, no solo mejorará el bienestar físico de los profesionales, sino que también incrementará la eficiencia y seguridad de los procedimientos quirúrgicos. Fomentar una cultura de prevención activa y basar las decisiones en la evidencia científica garantizará que la carrera quirúrgica se desarrolle en condiciones saludables y sostenibles a lo largo del tiempo.

6. Bibliografía

- Amat Rodrigo, J. (2016, agosto). *Regresión logística simple y múltiple*. Obtenido de RPubs: https://rpubs.com/Joaquin_AR/229736
- Barrios Arce, J. (2019, julio 26). *La matriz de confusión y sus métricas*. Obtenido de Health Big Data: <https://www.juanbarrios.com/la-matriz-de-confusion-y-sus-metricas/>
- Cerda, J., & Cifuentes, L. (2011, julio 25). *Uso de curvas ROC en investigación clínica. Aspectos teórico-prácticos*. Obtenido de SciELO: https://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0716-10182012000200003
- De Ingenieria, F. (s. f.). *Revista Ingeniería Investigación y Tecnología | Facultad de Ingeniería-UNAM*. <https://www.revistaingenieria.unam.mx/numeros/v21n3-02.php>
- Parra, F. (s. f.). *5 Modelos con Variables Cualitativas | Estadística y Machine Learning con R*. <https://bookdown.org/content/2274/modelos-con-variables-cualitativas.html>
- Gil Martínez, C. (2018, mayo). *Regresión logística (simple y múltiple)*. Obtenido de RPubs: https://rpubs.com/Cristina_Gil/Regresion_Logistica
- Greyrat, R. (2022, julio 5). *Ventajas y desventajas de la regresión logística*. Obtenido de Barcelona Geeks: <https://barcelonageeks.com/ventajas-y-desventajas-de-la-regresionlogistica/>
- Hemmati, P., Nguyen, T. C., & Dearani, J. A. (2022). Ergonomics for Surgeons by Surgeons—Posture, Loupes, and Exercise. *JAMA Surgery*, 157(9), 751. <https://doi.org/10.1001/jamasurg.2022.0676>
- Recuero de los Santos, P. (2021, diciembre 13). *Cómo interpretar la matriz de confusión: ejemplo práctico*. Obtenido de Empresas.blogthinkbig: <https://empresas.blogthinkbig.com/como-interpretar-la-matriz-de-confusion-ejemplopractico/>
- Zihni, A. M., Cavallo, J. A., Ray, S., Ohu, I., Cho, S., & Awad, M. M. (2016). Ergonomic analysis of primary and assistant surgical roles. *Journal Of Surgical Research*, 203(2), 301-305. <https://doi.org/10.1016/j.jss.2016.03.058>
- Steyerberg, E. W. (2009). *Clinical prediction models: A practical approach to development, validation, and updating*. Springer
- Silva, L. C., & Barroso, I. M. (2004). *Regresión logística*. Ed. La Muralla y Hespérides.
- Hocking, R. R. (1976). *The analysis and selection of variables in linear regression*. *Biometrics*, 32(1), 1-49.
- KeepCoding. (2023, abril 26). *¿Qué es el criterio de información Akaike AIC?*. Obtenido de <https://keepcoding.io/blog/que-es-el-criterio-de-informacion-akaike/>

SimpliRoute. (2023, enero 11). *Regresión logística: Qué es y cómo funciona*. Obtenido de <https://simpliroute.com/es/blog/regresion-logistica>

Giagio, S., Volpe, G., Pillastrini, P., Gasparre, G., Frizziero, A., & Squizzato, F. (2019). A Preventive Program for Work-related Musculoskeletal Disorders Among Surgeons. *Annals Of Surgery*, 270(6), 969-975. <https://doi.org/10.1097/SLA.0000000000003199>

Rodríguez, D. (2018, julio 23). *La regresión logística*. Obtenido de Analytics Lane: <https://www.analyticslane.com/2018/07/23/la-regresion-logistica/>

Schlussel, A. T., & Maykel, J. A. (2019). Ergonomics and Musculoskeletal Health of the Surgeon. *Clinics In Colon And Rectal Surgery*, 32(06), 424-434. <https://doi.org/10.1055/s-0039-1693026>

Zúñiga, J. J. E. (2020). Aplicación de algoritmos Random Forest y XGBoost en una base de solicitudes de tarjetas de crédito. *Ingeniería Investigación y Tecnología*, 21(3), 1-16. <https://doi.org/10.22201/fi.25940732e.2020.21.3.022>

7. Anexos

Código

Librerías

```
library(readxl)
library(tidyverse)
library(kableExtra)
library(dplyr)
library(ggplot2)
library(MASS)
library(gtsummary)
library(rlang)
library(rstatix)
library(corrplot)
library(pROC)
library(gt)
library(caret)
library(randomForest)
library(broom)
```

Limpieza de Datos

```
# Leer La base de datos
datos <- read_excel("Base_de_datos_simplificada.xlsx")

datos <- datos %>%
  rename(TALLA_m = `TALLA _m`)

datos <- datos %>%
  rename(duracion_Zona_dorsal = `duracion_Zona dorsal`)

datos <- datos %>%
  rename(uso_equipo_provoca_sintomas = `uso _equipo_ provoca_sintomas`)
resultados <- list()

# Eliminar la columna "...55" que está vacía
datos <- datos %>% select(-"...55")

# Definir Listas de variables categóricas
categorical_vars <- c("GENERO", "mano_dominante", "ambito", "Tipos_de_cirugia_realizo", "guardias")

binary_vars_01 <- c("Cuello", "Hombros", "Zona_dorsal", "Lumbares", "Munyeca",
  "Rodillas", "Manos_y_dedos", "Tobillo_pie")

binary_vars_12 <- c("dolor12m", "postura", "necesitado_asistencia", "moles_tias_baja",
  "formacion_ergo", "mala_postura", "conocce_recomendaciones",
  "aplicado_informacion", "formacion_suficiente")

# Convertir variables categóricas a factores con etiquetas
datos <- datos %>%
```

```

  mutate(GENERO = factor(GENERO, levels = c(2,1), labels = c("Mujer", "Hombre
")),
  mano_dominante = factor(mano_dominante, levels = c(1, 2, 3), labels =
c("Derecha", "Izquierda", "Ambas")),
  ambito = factor(ambito, levels = c(1, 2, 3), labels = c("Público", "Privado", "Ambas")),
#hay que cambiar esta variable no es eso
  Tipos_de_cirugia_realizo = factor(Tipos_de_cirugia_realizo,
  levels = c(0, 1, 2, 3, 4, 5),
  labels = c("Sin Dolor-Leve","Sin Dolor-Leve","Sin
Dolor-Leve","Moderada-Importante","Moderada-Importante","Sin Dolor-Lev
e")),
  guardias = factor(guardias, levels = c(1, 2), labels = c("No", "Sí")
),
  duracion_Cuello=factor(duracion_Cuello,levels=c(0,1,2,3,4),labels=c(
"Ningun dolor","Horas","Varios dias","Unas semanas","Meses")),
  duracion_Hombros=factor(duracion_Hombros,levels=c(0,1,2,3,4),labels=
c("Ningun dolor","Horas","Varios dias","Unas semanas","Meses")),
  duracion_Zona_dorsal=factor(duracion_Zona_dorsal,levels=c(0,1,2,3,4),
labels=c("Ningun dolor","Horas","Varios dias","Unas semanas","Meses")),
  duracion_Lumbares=factor(duracion_Lumbares,levels=c(0,1,2,3,4),label
s=c("Ningun dolor","Horas","Varios dias","Unas semanas","Meses")),
  duracion_Muñeca=factor(duracion_Muñeca,levels=c(0,1,2,3,4),labels=c("Ni
ngun dolor","Horas","Varios dias","Unas semanas","Meses")),
  duracion_Manos_y_dedos=factor(duracion_Manos_y_dedos,levels=c(0,1,2,3
,4),labels=c("Ningun dolor","Horas","Varios dias","Unas semanas","Meses
")),
  duracion_Rodillas=factor(duracion_Rodillas,levels=c(0,1,2,3,4),label
s=c("Ningun dolor","Horas","Varios dias","Unas semanas","Meses")),
  duracion_Piernas_gemeulos=factor(duracion_Piernas_gemeulos,levels=c(0,1
,2,3,4),labels=c("Ningun dolor","Horas","Varios dias","Unas semanas","Me
ses")),
  duracion_Tobillo_pie=factor(duracion_Tobillo_pie,levels=c(0,1,2,3,4),
labels=c("Ningun dolor","Horas","Varios dias","Unas semanas","Meses")),
  laparoscopica=factor(laparoscopica,levels=c(0,1,2,3,4,5,6),labels=c(
"Ninguna","<10","10-25","25-50","50-75","75-100",">100")),
  menor=factor(menor,levels=c(0,1,2,3,4,5,6),labels=c("Ninguna","<10
","10-25","25-50","50-75","75-100",">100")),
  abierta=factor(abierta,levels=c(0,1,2,3,4,5,6),labels=c("Ninguna","<
10","10-25","25-50","50-75","75-100",">100")),
  endoscopica =factor(endoscopica,levels=c(0,1,2,3,4,5,6),labels=c("Ni
nguna","<10","10-25","25-50","50-75","75-100",">100")),
  robotica=factor(robotica,levels=c(0,1,2,3,4,5,6),labels=c("Ninguna",
"<10","10-25","25-50","50-75","75-100",">100")),
  factores_causa_molestias=factor(factores_causa_molestias,levels=c(0,1
,2,3,4,5),labels=c("Sin Dolor-Leve","Sin Dolor-Leve","Sin Dolor-Leve",
"Moderada-Importante","Moderada-Importante","Sin Dolor-Leve")),
  Duracion_de_las_intervenciones=factor(Duracion_de_las_intervenciones,
levels=c(0,1,2,3,4,5),labels=c("Sin Dolor-Leve","Sin Dolor-Leve","Sin Do
lor-Leve","Moderada-Importante","Moderada-Importante","Sin Dolor-Leve
")),
  Iinstrumental_quirurgico=factor(Iinstrumental_quirurgico,levels=c(0,1
,2,3,4,5),labels=c("Sin Dolor-Leve","Sin Dolor-Leve","Sin Dolor-Leve",
"Moderada-Importante","Moderada-Importante","Sin Dolor-Leve")),

```

```

    Ubicación_de_pantallas_laparoscopia=factor(Ubicación_de_pantallas_lap
    aroscopia,levels=c(0,1,2,3,4,5),labels=c("Sin Dolor-Leve","Sin Dolor-Leve",
    "Sin Dolor-Leve","Moderada-Importante","Moderada-Importante","Sin Dolo
    r-Leve")),
    Otras=factor(Otras,levels=c(0,1,2,3,4,5),labels=c("Sin Dolor-Leve",
    "Sin Dolor-Leve","Sin Dolor-Leve","Moderada-Importante","Moderada-Impo
    rtante","Sin Dolor-Leve")),
    anyos_cirujano=factor(anyos_cirujano,levels=c(1,2,3),labels=c(">20",
    "10-20","<10")),

    )
datos <- datos %>%
  mutate(residente = factor(ifelse(is.na(residente), "Residente", as.character(re
  sidente)))))

# Convertir variables binarias con valores 0/1 en factores con etiquetas "No/Sí"
datos <- datos %>%
  mutate(across(all_of(binary_vars_01), ~factor(., levels = c(0, 1), labels = c("
  No", "Sí"))))

# Convertir variables binarias con valores 1/2 en factores con etiquetas "No/Sí"
datos <- datos %>%
  mutate(across(all_of(binary_vars_12), ~factor(., levels = c(2, 1), labels = c("
  No", "Sí"))))

# Variables a imputar con la media
vars_media <- c("EDAD", "PESO_kg", "TALLA_m", "BMI")

# Verificar que las variables existen en el dataframe y son numéricas
vars_media <- vars_media[vars_media %in% names(datos) & sapply(datos[vars_media], is
.numeric)] 

# Reemplazar NA con la media en cada variable
for (var in vars_media) {
  media_var <- mean(datos[[var]], na.rm = TRUE) # Calcular la media sin NA
  if (!is.na(media_var)) { # Solo imputar si la media no es NA
    datos[[var]][is.na(datos[[var]])] <- media_var
  }
}

# Eliminar la variable con demasiados NA
datos <- datos %>% select(-OTRAS_DOLENCIAS)
datos <- datos %>% select(-aplicado_informacion)
datos <- datos %>% select(-tipo_asistencia)
datos <- datos %>% select(-como_obtuvo_informacion)

# Imputar variables categóricas con la moda
moda_guardias <- names(sort(table(datos$guardias), decreasing = TRUE))[1]
datos$guardias[is.na(datos$guardias)] <- moda_guardias

```

```

moda_molestias <- names(sort(table(datos$molestias_baja), decreasing = TRUE))[1]
datos$molestias_baja[is.na(datos$molestias_baja)] <- moda_molestias

moda_conocce <- names(sort(table(datos$conocce_recomendaciones), decreasing = TRUE))[1]
datos$conocce_recomendaciones[is.na(datos$conocce_recomendaciones)] <- moda_conocce

# Reemplazar Los valores NA en La variable ambito con "Público"
datos$ambito[is.na(datos$ambito)] <- "Público"

datos <- datos %>% filter(TALLA_m > 0, BMI > 0)

datos <- datos %>%
  mutate(anyos_cirujano = case_when(
    is.na(anyos_cirujano) & residente %in% c("MIR 1", "MIR 2", "MIR 3", "MIR 4",
    ") ~ "<10",
    is.na(anyos_cirujano) ~ names(sort(table(anyos_cirujano), decreasing = TRUE))[1], # Usa La moda
    TRUE ~ as.character(anyos_cirujano)
  ))

datos$anyos_cirujano <- factor(datos$anyos_cirujano)

datos <- datos %>%
  mutate(BMI_categoria = case_when(
    BMI < 25 ~ "healthy",
    BMI >= 25 & BMI < 30 ~ "overweight",
    BMI >= 30 ~ "obese"
  ))
datos$BMI_categoria <- factor(datos$BMI_categoria, levels = c("healthy", "overweight", "obese"))

# Crear La variable "experiencia" agrupando residentes y especialistas
datos <- datos %>%
  mutate(experiencia = case_when(
    residente %in% c("MIR 1", "MIR 2", "MIR 3") ~ "MIR 1-3er año",
    residente %in% c("MIR 4", "MIR 5") ~ "MIR 4º y 5º año",
    TRUE ~ as.character(anyos_cirujano) # Usa años de experiencia para Los especialistas
  ))

# Convertir La variable experiencia en factor y ordenar Los niveles
datos$experiencia <- factor(datos$experiencia, levels = c("MIR 1-3er año", "MIR 4º y 5º año", "<10", "10-20", ">20"))

# Definir el mapa de conversión de categorías a valores numéricos

```

```

mapa_conversion <- c(
  "Ninguna" = 0, "<10" = 5, "10-25" = 17.5, "25-50" = 37.5,
  "50-75" = 62.5, "75-100" = 87.5, ">100" = 110
)

# Crear variables numéricas temporales para calcular la suma
datos <- datos %>%
  mutate(
    menor_num = as.numeric(recode(menor, !!!mapa_conversion)),
    abierta_num = as.numeric(recode(abierta, !!!mapa_conversion)),
    endoscopica_num = as.numeric(recode(endoscopica, !!!mapa_conversion)),
    robotica_num = as.numeric(recode(robotica, !!!mapa_conversion))
  )

# Crear la variable "Total de cirugías al año" sumando las categorías
# numéricas
datos <- datos %>%
  mutate(Total_Cirugias_Año_Num = rowSums(select(., menor_num, abierta_num, endoscopica_num, robotica_num), na.rm = TRUE))

# Agrupar la variable "Total_Cirugias_Año" en tres categorías
datos <- datos %>%
  mutate(
    Total_Cirugias_Año = case_when(
      Total_Cirugias_Año_Num < 25 ~ "<25",
      Total_Cirugias_Año_Num >= 25 & Total_Cirugias_Año_Num <= 50 ~ "25-50",
      Total_Cirugias_Año_Num > 50 ~ ">50"
    )
  )

# Convertir "Total_Cirugias_Año" en factor con el orden correcto
datos$Total_Cirugias_Año <- factor(datos$Total_Cirugias_Año, levels = c("<25", "25-50", ">50"))

# Eliminar las variables numéricas temporales (ya no son necesarias)
datos <- datos %>% select(-menor_num, -abierta_num, -endoscopica_num, -robotica_num)

# **Mantener las variables originales "menor", "abierta", "endoscopica" y "robotica" como factores**
datos <- datos %>%
  mutate(
    menor = factor(menor, levels = names(mapa_conversion)),
    abierta = factor(abierta, levels = names(mapa_conversion)),
    endoscopica = factor(endoscopica, levels = names(mapa_conversion)),
    robotica = factor(robotica, levels = names(mapa_conversion))
  )

datos <- datos %>%
  mutate(dias_quirofano = factor(ifelse(dias_quirofano >= 3 & dias_quirofano <= 5, "Intensiva", "No"),
                                         levels = c("No", "Intensiva")))

datos <- datos %>%

```

```

  mutate(cirugiaEndo = ifelse(endoscopica == "Ninguna", "No", "Sí"))

# Convertir a factor
datos$cirugiaEndo <- factor(datos$cirugiaEndo, levels = c("No", "Sí"))

datos <- datos %>%
  mutate(actividad_fisica = case_when(
    dias_30_min_Af == 0 ~ "No realiza",
    dias_30_min_Af %in% c(1, 2) ~ "Moderada",
    dias_30_min_Af > 2 ~ "Intensiva"
  ))

# Convertir en factor con niveles ordenados
datos$actividad_fisica <- factor(datos$actividad_fisica, levels = c("No realiza",
  "Moderada", "Intensiva"))

summary(datos)

```

Tabla descriptiva de la muestra según género

```

#CARACTERISTICAS DE LA MUESTRA SEGUN EL GENERO
tabla_genero <- datos %>%
  select(GENERO, EDAD, TALLA_m, mano_dominante, BMI, ambito,
         guardias, experiencia, actividad_fisica, Total_Cirugias_Año) %>%
 tbl_summary(by = GENERO,
             statistic = list(all_continuous() ~ "{median} ({p25}, {p75})"),
             missing = "no") %>%
  add_n() %>% # Agrega número de observaciones
  add_p() %>% # Agrega test estadísticos
  modify_header(label = "***Variable***") %>%
  modify_caption("**Tabla 1: Características de la muestra según el Género**") %>%
  bold_labels()
tabla_genero

```

Tabla descriptiva de la muestra según el dolor en los últimos 12 meses

```

#CARACTERISTICAS DE LA MUESTRA SEGUN EL DOLOR DE LOS ULTIMOS 12 MESES
tabla_dolor <- datos %>%
 tbl_summary(
  by = dolor12m,
  include = c(GENERO, EDAD, TALLA_m, mano_dominante, BMI, ambito,
              guardias, experiencia, actividad_fisica, Total_Cirugias_Año),
  statistic = list(all_continuous() ~ "{median} ({p25}, {p75})"),
  missing = "no"
) %>%
  add_n() %>%
  add_p(test.args = all_tests("fisher.test") ~ list(simulate.p.value = TRUE
)) %>%
  modify_header(label = "***Variable***") %>%

```

```

modify_caption("**Tabla 2: Características de la muestra según Dolor en
los Últimos 12 Meses**") %>%
bold_labels()

tabla_dolor

calcular_odds_ratios <- function(modelo, nombre_variable) {
  # Extraer los coeficientes y p-valores del modelo
  coeficientes <- summary(modelo)$coefficients

  # Omitir el intercepto si se desea
  variables_modelo <- rownames(coeficientes)
  variables_modelo <- variables_modelo[variables_modelo != "(Intercept)"]

  # Calcular OR e Intervalos de Confianza del 95%
  OR <- exp(coef(modelo)[variables_modelo])
  CI <- exp(confint(modelo)[variables_modelo, ])

  # Crear un data.frame con los resultados para todas las variables
  resultados_OR <- data.frame(
    Variable = variables_modelo,
    OR = round(OR, 2),
    `CI 2.5%` = round(CI[, 1], 2),
    `CI 97.5%` = round(CI[, 2], 2),
    `P-Valor` = round(coeficientes[variables_modelo, "Pr(>|z|)"], 4),
    row.names = NULL
  )
  print(resultados_OR)
  return(resultados_OR)
}

```

Tabla descriptiva de la muestra según el dolor en los últimos 12 meses

```

#CARACTERISTICAS DE LA MUESTRA SEGUN EL GENERO
# Calcular los porcentajes reales desde la base de datos
tabla_dolor <- datos %>%
  group_by(GENERO) %>%
  summarise(
    Cuello = mean(Cuello == "Sí") * 100,
    Hombros = mean(Hombros == "Sí") * 100,
    Zona_Dorsal = mean(Zona_dorsal == "Sí") * 100,
    Lumbares = mean(Lumbares == "Sí") * 100,
    Muñeca = mean(Muñeca == "Sí") * 100,
    Manos_Dedos = mean(Manos_y_dedos == "Sí") * 100,
    Rodillas = mean(Rodillas == "Sí") * 100,
    Tobillo_Pie = mean(Tobillo_pie == "Sí") * 100
  ) %>%
  pivot_longer(-GENERO, names_to = "Zona del Cuerpo", values_to = "Por
centaje") %>%
  pivot_wider(names_from = GENERO, values_from = Porcentaje, names_pre
fix = "Género_") %>%

```

```

rename(`Hombres (%)` = Género_Hombre, `Mujeres (%)` = Género_Mujer)

# Definir los colores según los valores
color_map <- function(x) {
  case_when(
    x > 75 ~ "#E74C3C", # Rojo (>75%)
    x > 50 ~ "#E67E22", # Naranja (50-75%)
    x > 25 ~ "#F7DC6F", # Amarillo (25-50%)
    TRUE ~ "#2ECC71" # Verde (<25%)
  )
}

# Crear la tabla formateada

tabla_dolor <- datos %>%
  group_by(GENERO) %>%
  summarise(
    Cuello = mean(Cuello == "Sí") * 100,
    Hombros = mean(Hombros == "Sí") * 100,
    Zona_Dorsal = mean(Zona_dorsal == "Sí") * 100,
    Lumbares = mean(Lumbares == "Sí") * 100,
    Muñeca = mean(Munyeca == "Sí") * 100,
    Manos_Dedos = mean(Manos_y_dedos == "Sí") * 100,
    Rodillas = mean(Rodillas == "Sí") * 100,
    Tobillo_Pie = mean(Tobillo_pie == "Sí") * 100
  ) %>%
  pivot_longer(-GENERO, names_to = "Zona del Cuerpo", values_to = "Porcentaje") %>%
  pivot_wider(names_from = GENERO, values_from = Porcentaje, names_prefix = "Género_") %>%
  rename(`Hombres (%)` = Género_Hombre, `Mujeres (%)` = Género_Mujer)

# Crear la tabla sin colores
tabla_dolor %>%
  gt() %>%
  tab_header(
    title = md("**Prevalencia de Dolor Musculoesquelético por Zona y Género**")
  ) %>%
  cols_label(
    `Zona del Cuerpo` = "Zona del Cuerpo",
    `Hombres (%)` = "Hombres (%)",
    `Mujeres (%)` = "Mujeres (%)"
  ) %>%
  fmt_number(
    columns = c(`Hombres (%)`, `Mujeres (%)`),
    decimals = 1
  ) %>%
  tab_options(
    table.font.color = "black",
    table.border.top.color = "black",
    table.border.bottom.color = "black"
  )

```

Cuello

Modelo

```
vars_modelo <- c("Cuello", "EDAD", "GENERO", "BMI_categoria", "PESO_kg",
  "dias_quirofano",
  "Duracion_de_las_intervenciones",
  "Ubicación_de_pantallas_laparoscopia",
  "Total_Cirugias_Año", "menor", "cirugiaEndo",
  "Instrumental_quirurgico", "factores_causa_molestias")

datos_modelo <- datos %>%
  select(all_of(vars_modelo))

modelo_inicial <- glm(Cuello ~ ., data = datos_modelo, family = binomial)

# Seleccionar el mejor modelo eliminando variables no significativas (p > 0.05)
modelo_optimo_Cuello <- stepAIC(modelo_inicial,
  direction = "both",
  scope = list(lower = Cuello ~ GENERO, upper = modelo_inicial),
  trace = FALSE)

# Mostrar resumen del modelo final
summary(modelo_optimo_Cuello)
```

```
## 
## Call:
## glm(formula = Cuello ~ EDAD + GENERO + Duracion_de_las_intervenciones +
##     Ubicación_de_pantallas_laparoscopia, family = binomial,
##     data = datos_modelo)
## 
## Coefficients:
##                               Estimate Std. Error
## (Intercept)                2.86183   0.51573
## EDAD                     -0.05360   0.01014
## GENEROHombre               -0.26795   0.24772
## Duracion_de_las_intervencionesModerada-Importante 1.30059   0.28562
## Ubicación_de_pantallas_laparoscopiaModerada-Importante 0.93109   0.25570
##                               z value Pr(>|z|)
## (Intercept)                5.549 2.87e-08 ***
## EDAD                      -5.285 1.25e-07 ***
## GENEROHombre               -1.082 0.279395
## Duracion_de_las_intervencionesModerada-Importante 4.554 5.28e-06 ***
## Ubicación_de_pantallas_laparoscopiaModerada-Importante 3.641 0.000271 ***
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for binomial family taken to be 1)
## 
## Null deviance: 564.55  on 648  degrees of freedom
## Residual deviance: 479.90  on 644  degrees of freedom
## AIC: 489.9
## 
## Number of Fisher Scoring iterations: 5
```

Odds Ratio

```
tabla_cuello <- modelo_optimo_Cuello %>%
 tbl_regression(
  exponentiate = TRUE,                                # mostrar OR
  estimate_fun = ~style_sigfig(.x, digits = 2), # OR con 2 cifras
  conf.int = TRUE,                                    # incluir IC
  conf.level = 0.95)                                 # nivel 95%
```

```

) %>%
  modify_table_body(
    ~ .x %>%
      # 1) crear '95% CI', vacía si no hay conf.Low
      mutate(
        `95% CI` = ifelse(
          !is.na(conf.low),
          paste0("(", sprintf("%.2f", conf.low), ", ", sprintf("%.2f",
          conf.high), ")"),
          ""
        )
      ) %>%
      # 2) quitar conf.Low/conf.high
      select(-conf.low, -conf.high) %>%
      # 3) reordenar columnas: label (o variable), estimate, 95% CI,
      p.value, luego resto
      select(
        label,          # el texto de la característica
        estimate,       # el OR
        `95% CI`,       # nuestro intervalo formateado
        p.value,         # p-value
        everything()    # AIC, BIC, etc. quedarán abajo
      )
    ) %>%
    # 4) renombrar encabezados
    modify_header(
      label ~ "***Characteristic***",
      estimate ~ "***Cuello (OR)***",
      p.value ~ "***p-value***",
      `95% CI` ~ "***95 % CI***"
    ) %>%
    # 5) añadir métricas de modelo al pie
    add_glance_table(
      include = c("AIC", "BIC", "logLik", "deviance", "nobs")
    )
  )

tabla_cuello

```

Random Forest

```

variables_seleccionadas <- all.vars(formula(modelo_optimo_Cuello))[-1]

# Asegurar que La variable dependiente está incluida
vars_modelo_final <- c("Cuello", variables_seleccionadas)

# Seleccionar solo las variables finales
datos_modelo <- datos %>% select(all_of(vars_modelo_final))

# Convertir variables categóricas en factores
datos_modelo <- datos_modelo %>% mutate(across(where(is.character), as.factor))

# Dividir en datos de entrenamiento (70%) y prueba (30%)
set.seed(123) # Para reproducibilidad
trainIndex <- createDataPartition(datos_modelo$Cuello, p = 0.7, list = FALSE)

```

```

datos_train <- datos_modelo[trainIndex, ]
datos_test <- datos_modelo[-trainIndex, ]

# Entrenar el modelo Random Forest
set.seed(123)
modelo_rf <- randomForest(Cuello ~ ., data = datos_train, ntree = 500, mtry = 3,
importance = TRUE)

# Predicciones en datos de prueba
predicciones <- predict(modelo_rf, datos_test, type = "class")

# Matriz de confusión
tabla_confusion <- confusionMatrix(predicciones, datos_test$Cuello)
tabla_confusion

# Extraer la matriz de confusión
conf_mat <- tabla_confusion$table

rownames(conf_mat) <- paste("Predicho", rownames(conf_mat))
colnames(conf_mat) <- paste("Referencia", colnames(conf_mat))

# Extraer métricas importantes del modelo
accuracy <- round(tabla_confusion$overall["Accuracy"], 4)
kappa <- round(tabla_confusion$overall["Kappa"], 4)
sensitivity <- round(tabla_confusion$byClass["Sensitivity"], 4)
specificity <- round(tabla_confusion$byClass["Specificity"], 4)

# Crear tabla de métricas
metrics_table <- data.frame(
  Valor = c(accuracy, kappa, sensitivity, specificity)
)

# Imprimir tabla de matriz de confusión
kable(conf_mat, caption = "Matriz de confusión", align = "c") %>%
  kable_styling(full_width = FALSE, position = "center")

# Imprimir tabla de métricas
kable(metrics_table, caption = "Métricas del modelo", align = "c") %>%
  kable_styling(full_width = FALSE, position = "center")

```

Hombros

Modelo

```

# Definir todas las variables iniciales
vars_modelo <- c("HOMBROS", "EDAD", "GENERO", "BMI_categoria", "PESO_kg",
  "dias_quirofano",
  "Duración_de_las_intervenciones",
  "Ubicación_de_pantallas_laparoscopia",
  "Total_Cirugías_Año", "menor", "cirugiaEndo",
  "Instrumental_quirúrgico", "factores_causa_molestias")

```

```

# Seleccionar datos del modelo
datos_modelo <- datos %>%
  select(all_of(vars_modelo))

# Crear modelo inicial con todas las variables
modelo_inicial <- glm(Hombres ~ ., data = datos_modelo, family = binomial)

# Aplicar stepAIC() para optimizar el modelo
modelo_optimo_Hombres <- stepAIC(modelo_inicial,
                                    direction = "both",
                                    scope = list(lower = Hombres ~ GENERO, upper = m
odelo_inicial),
                                    trace = FALSE)

# Mostrar resumen del modelo final
summary(modelo_optimo_Hombres)

```

```

## 
## Call:
## glm(formula = Hombres ~ EDAD + GENERO + PESO_kg + Duracion_de_las_intervenciones +
##      Total_Cirugias_Año + Iinstrumental_quirurgico, family = binomial,
##      data = datos_modelo)
## 
## Coefficients:
##                               Estimate Std. Error z value
## (Intercept)                0.876141  0.783804  1.119
## EDAD                     -0.015352  0.008093 -1.897
## GENEROHomble               -0.209449  0.253606 -0.826
## PESO_kg                    -0.013992  0.008153 -1.716
## Duracion_de_las_intervencionesModerada-Importante 1.071028  0.254855  4.202
## Total_Cirugias_Año25-50      0.356841  0.549152  0.650
## Total_Cirugias_Año>50       0.974732  0.486254  2.005
## Iinstrumental_quirurgicoModerada-Importante      0.406321  0.194595  2.088
## Pr(>|z|)
## (Intercept)                  0.2632
## EDAD                         0.0578 .
## GENEROHomble                  0.4089
## PESO_kg                       0.0861 .
## Duracion_de_las_intervencionesModerada-Importante 2.64e-05 ***
## Total_Cirugias_Año25-50       0.5158
## Total_Cirugias_Año>50        0.0450 *
## Iinstrumental_quirurgicoModerada-Importante      0.0368 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for binomial family taken to be 1)
## 
## Null deviance: 738.04  on 648  degrees of freedom
## Residual deviance: 686.20  on 641  degrees of freedom
## AIC: 702.2
## 
## Number of Fisher Scoring iterations: 4

```

Odds Ratio

```

tabla_hombres <- modelo_optimo_Hombres %>%
 tbl_regression(
  exponentiate = TRUE,                                # mostrar OR
  estimate_fun = ~style_sigfig(.x, digits = 2), # OR con 2 cifras
  conf.int     = TRUE,                                # incluir IC
  conf.level   = 0.95                                # nivel 95%
) %>%
  modify_table_body(
  ~ .x %>%
    # 1) crear '95% CI', vacía si no hay conf.Low
    mutate(
      `95% CI` = ifelse(
        !is.na(conf.low),

```

```

            paste0(", sprintf("%.2f", conf.low), ", ", sprintf("%.2f
", conf.high), ")"),
        ""
    )
) %>%
# 2) quitar conf.Low/conf.high
select(-conf.low, -conf.high) %>%
# 3) reordenar columnas: label (o variable), estimate, 95% CI,
p.value, luego resto
select(
  label,          # el texto de la característica
  estimate,       # el OR
  `95% CI`,      # nuestro intervalo formateado
  p.value,        # p-value
  everything()   # AIC, BIC, etc. quedarán abajo
)
) %>%
# 4) renombrar encabezados
modify_header(
  label ~ "***Characteristic***",
  estimate ~ "***Cuello (OR)***",
  p.value ~ "***p-value***",
  `95% CI` ~ "***95 % CI***"
) %>%
# 5) añadir métricas de modelo al pie
add_glance_table(
  include = c("AIC", "BIC", "logLik", "deviance", "nobs")
)

```

tabla_hombres

Random Forest

```

variables_seleccionadas <- all.vars(formula(modelo_optimo_Hombres))[-1] # Quita "Hombres"

# Asegurar que la variable dependiente está incluida
vars_modelo_final <- c("Hombres", variables_seleccionadas)

# Seleccionar solo las variables finales
datos_modelo <- datos %>% select(all_of(vars_modelo_final))

# Dividir en datos de entrenamiento (70%) y prueba (30%)
set.seed(123) # Para reproducibilidad
trainIndex <- createDataPartition(datos_modelo$Hombres, p = 0.7, list = FALSE)
datos_train <- datos_modelo[trainIndex, ]
datos_test <- datos_modelo[-trainIndex, ]

# Entrenar el modelo Random Forest con las variables seleccionadas
set.seed(123)
modelo_rf <- randomForest(Hombres ~ ., data = datos_train, ntree = 500, mtry = 3
, importance = TRUE)

```

```

# Predicciones en datos de prueba
predicciones <- predict(modelo_rf, datos_test, type = "class")

# Matriz de confusión
tabla_confusion <- confusionMatrix(predicciones, datos_test$Hombres)
tabla_confusion

conf_mat <- tabla_confusion$table

rownames(conf_mat) <- paste("Predicho", rownames(conf_mat))
colnames(conf_mat) <- paste("Referencia", colnames(conf_mat))

# Extraer métricas importantes del modelo
accuracy <- round(tabla_confusion$overall["Accuracy"], 4)
kappa <- round(tabla_confusion$overall["Kappa"], 4)
sensitivity <- round(tabla_confusion$byClass["Sensitivity"], 4)
specificity <- round(tabla_confusion$byClass["Specificity"], 4)

# Crear tabla de métricas
metrics_table <- data.frame(
  Valor = c(accuracy, kappa, sensitivity, specificity)
)

# Imprimir tabla de matriz de confusión
kable(conf_mat, caption = "Matriz de confusión", align = "c") %>%
  kable_styling(full_width = FALSE, position = "center")

# Imprimir tabla de métricas
kable(metrics_table, caption = "Métricas del modelo", align = "c") %>%
  kable_styling(full_width = FALSE, position = "center")

```

Zona Dorsal

Modelo

```

vars_modelo <- c("Zona_dorsal", "EDAD", "GENERO", "BMI_categoria", "PESO_kg",
  "dias_quirofano",
  "Duracion_de_las_intervenciones",
  "Ubicación_de_pantallas_laparoscopia",
  "Total_Cirugias_Año", "menor", "cirugiaEndo",
  "Instrumental_quirurgico", "factores_causa_molestias")

datos_modelo <- datos %>%
  select(all_of(vars_modelo))
modelo_inicial <- glm(Zona_dorsal ~ ., data = datos_modelo, family = binomial)
# Seleccionar el mejor modelo eliminando variables no significativas (p > 0.05)
modelo_optimo_Dorsal <- stepAIC(modelo_inicial,
  direction = "both",
  scope = list(lower = Zona_dorsal ~ GENERO, upper
  = modelo_inicial),
  trace = FALSE)# Mostrar resumen del modelo f

```

inal
 summary(modelo_optimo_Dorsal)

```
## 
## Call:
## glm(formula = Zona_dorsal ~ EDAD + GENERO + BMI_categoria + Duracion_de_las_intervenciones +
##     Ubicación_de_pantallas_laparoscopia + Iinstrumental_quirurgico +
##     factores_causa_molestias, family = binomial, data = datos_modelo)
## 
## Coefficients:
##                               Estimate Std. Error
## (Intercept)                0.888357  0.423380
## EDAD                     -0.032392  0.007671
## GENEROHomme               -0.051710  0.193519
## BMI_categoriaoverweight   -0.330504  0.205438
## BMI_categoriaobese        0.395298  0.358975
## Duracion_de_las_intervencionesModerada-Importante  0.417133  0.276288
## Ubicación_de_pantallas_laparoscopiaModerada-Importante 0.443240  0.194142
## Iinstrumental_quirurgicoModerada-Importante        0.363372  0.192025
## factores_causa_molestiasModerada-Importante        0.598903  0.254749
##                               z value Pr(>|z|)
## (Intercept)                2.098   0.0359 *
## EDAD                     -4.223  2.41e-05 ***
## GENEROHomme               -0.267   0.7893
## BMI_categoriaoverweight   -1.609   0.1077
## BMI_categoriaobese        1.101   0.2708
## Duracion_de_las_intervencionesModerada-Importante  1.510   0.1311
## Ubicación_de_pantallas_laparoscopiaModerada-Importante 2.283   0.0224 *
## Iinstrumental_quirurgicoModerada-Importante        1.892   0.0584 .
## factores_causa_molestiasModerada-Importante        2.351   0.0187 *
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for binomial family taken to be 1)
## 
## Null deviance: 843.86 on 648 degrees of freedom
## Residual deviance: 779.29 on 640 degrees of freedom
## AIC: 797.29
## 
## Number of Fisher Scoring iterations: 4
```

Odds Ratio

```
tabla_dorsal <- modelo_optimo_Dorsal %>%
 tbl_regression(
  exponentiate = TRUE,                               # mostrar OR
  estimate_fun = ~style_sigfig(.x, digits = 2), # OR con 2 cifras
  conf.int     = TRUE,                               # incluir IC
  conf.level   = 0.95                               # nivel 95%
) %>%
  modify_table_body(
  ~ .x %>%
    # 1) crear '95% CI', vacía si no hay conf.Low
    mutate(
      `95% CI` = ifelse(
        !is.na(conf.low),
        paste0("(", sprintf("%.2f", conf.low), ", ", sprintf("%.2f",
        conf.high), ")"),
        ""))
    )
  ) %>%
  # 2) quitar conf.Low/conf.high
  select(-conf.low, -conf.high) %>%
  # 3) reordenar columnas: label (o variable), estimate, 95% CI,
  p.value, luego resto
  select(
```

```

        label,      # el texto de la característica
        estimate,   # el OR
        `95% CI`,   # nuestro intervalo formateado
        p.value,    # p-value
        everything() # AIC, BIC, etc. quedarán abajo
    )
) %>%
# 4) renombrar encabezados
modify_header(
    label ~ "**Characteristic**",
    estimate ~ "**Cuello (OR)**",
    p.value ~ "**p-value**",
    `95% CI` ~ "**95 % CI**"
) %>%
# 5) añadir métricas de modelo al pie
add_glance_table(
    include = c("AIC", "BIC", "logLik", "deviance", "nobs")
)

tabla_dorsal

```

Random Forest

```

variables_seleccionadas <- all.vars(formula(modelo_optimo_Dorsal))[-1] # Quita "Hombros"

# Asegurar que la variable dependiente está incluida
vars_modelo_final <- c("Zona_dorsal", variables_seleccionadas)

# Seleccionar solo las variables finales
datos_modelo <- datos %>% select(all_of(vars_modelo_final))

# Convertir variables categóricas en factores
datos_modelo <- datos_modelo %>% mutate(across(where(is.character), as.factor))

# Dividir en datos de entrenamiento (70%) y prueba (30%)
set.seed(123) # Para reproducibilidad
trainIndex <- createDataPartition(datos_modelo$Zona_dorsal, p = 0.7, list = FALSE)
)
datos_train <- datos_modelo[trainIndex, ]
datos_test <- datos_modelo[-trainIndex, ]

# Entrenar el modelo Random Forest
set.seed(123)
modelo_rf <- randomForest(Zona_dorsal ~ ., data = datos_train, ntree = 500, mtry = 3, importance = TRUE)

# Predicciones en datos de prueba
predicciones <- predict(modelo_rf, datos_test, type = "class")

# Matriz de confusión
tabla_confusion <- confusionMatrix(predicciones, datos_test$Zona_dorsal)
tabla_confusion

```

```

conf_mat <- tabla_confusion$table

rownames(conf_mat) <- paste("Predicho", rownames(conf_mat))
colnames(conf_mat) <- paste("Referencia", colnames(conf_mat))

# Extraer métricas importantes del modelo
accuracy     <- round(tabla_confusion$overall["Accuracy"], 4)
kappa        <- round(tabla_confusion$overall["Kappa"], 4)
sensitivity  <- round(tabla_confusion$byClass["Sensitivity"], 4)
specificity  <- round(tabla_confusion$byClass["Specificity"], 4)

# Crear tabla de métricas
metrics_table <- data.frame(
  Valor = c(accuracy, kappa, sensitivity, specificity)
)

# Imprimir tabla de matriz de confusión
kable(conf_mat, caption = "Matriz de confusión", align = "c") %>%
  kable_styling(full_width = FALSE, position = "center")

# Imprimir tabla de métricas
kable(metrics_table, caption = "Métricas del modelo", align = "c") %>%
  kable_styling(full_width = FALSE, position = "center")

```

Lumbares

Modelo

```

vars_modelo <- c("Lumbares", "EDAD", "GENERO", "BMI_categoria", "PESO_kg",
                 "dias_quirofano",
                 "Duracion_de_las_intervenciones",
                 "Ubicación_de_pantallas_laparoscopia",
                 "Total_Cirugias_Año", "menor", "cirugiaEndo",
                 "Instrumental_quirurgico", "factores_causa_molestias")
datos_modelo <- datos %>%
  select(all_of(vars_modelo))
modelo_inicial <- glm(Lumbares ~ ., data = datos_modelo, family = binomial)
# Seleccionar el mejor modelo eliminando variables no significativas (
p > 0.05)
modelo_optimo_Lumbar <- stepAIC(modelo_inicial,
                                     direction = "both",
                                     trace = FALSE)# Mostrar resumen del modelo f
inal
summary(modelo_optimo_Lumbar)

```

```

## 
## Call:
## glm(formula = Lumbares ~ EDAD + PESO_kg + Duracion_de_las_intervenciones +
##      Ubicación_de_pantallas_laparoscopia + factores_causa_molestias,
##      family = binomial, data = datos_modelo)
## 
## Coefficients:
##                               Estimate Std. Error
## (Intercept)                0.352388  0.644687
## EDAD                     -0.023209  0.009379
## PESO_kg                   0.014387  0.007948
## Duracion_de_las_intervencionesModerada-Importante  0.862964  0.301686
## Ubicación_de_pantallas_laparoscopiaModerada-Importante  0.395794  0.231270
## factores_causa_molestiasModerada-Importante          0.505566  0.294022
##                               z value Pr(>|z|)
## (Intercept)                0.547  0.58465
## EDAD                     -2.474  0.01334 *
## PESO_kg                   1.810  0.07025 .
## Duracion_de_las_intervencionesModerada-Importante  2.860  0.00423 **
## Ubicación_de_pantallas_laparoscopiaModerada-Importante  1.711  0.08701 .
## factores_causa_molestiasModerada-Importante          1.719  0.08553 .
## --- 
## Signif. codes:  0 '****' 0.001 '***' 0.01 '**' 0.05 '*' 0.1 ' ' 1
## 
## (Dispersion parameter for binomial family taken to be 1)
## 
## Null deviance: 584.29  on 648  degrees of freedom
## Residual deviance: 548.47  on 643  degrees of freedom
## AIC: 560.47
## 
## Number of Fisher Scoring iterations: 4

```

Odds Ratio

```

tabla_lumbar <- modelo_optimo_Lumbar %>%
 tbl_regression(
  exponentiate = TRUE,                                     # mostrar OR
  estimate_fun = ~style_sigfig(.x, digits = 2),        # OR con 2 cifras
  conf.int     = TRUE,                                     # incluir IC
  conf.level   = 0.95                                     # nivel 95%
) %>%
  modify_table_body(
  ~ .x %>%
    # 1) crear '95% CI', vacía si no hay conf.Low
    mutate(
      `95% CI` = ifelse(
        !is.na(conf.low),
        paste0("(", sprintf("%.2f", conf.low), ", ", sprintf("%.2f",
        conf.high), ")"),
        ""
      )
    ) %>%
    # 2) quitar conf.Low/conf.high
    select(-conf.low, -conf.high) %>%
    # 3) reordenar columnas: label (o variable), estimate, 95% CI,
    p.value, luego resto
    select(
      label,        # el texto de la característica
      estimate,    # el OR
      `95% CI`,    # nuestro intervalo formateado
      p.value,     # p-value
      everything() # AIC, BIC, etc. quedarán abajo
    )
) %>%

```

```

# 4) renombrar encabezados
modify_header(
  label ~ "**Characteristic**",
  estimate ~ "**Cuello (OR)**",
  p.value ~ "**p-value**",
  `95% CI` ~ "**95 % CI**"
) %>%
# 5) añadir métricas de modelo al pie
add_glance_table(
  include = c("AIC", "BIC", "logLik", "deviance", "nobs")
)

tabla_lumbar

```

Random Forest

```

variables_seleccionadas <- all.vars(formula(modelo_optimo_Lumbar))[-1] # Quita "Hombros"

# Asegurar que la variable dependiente está incluida
vars_modelo_final <- c("Lumbares", variables_seleccionadas)

# Seleccionar solo las variables finales
datos_modelo <- datos %>% select(all_of(vars_modelo_final))

# Convertir variables categóricas en factores
datos_modelo <- datos_modelo %>% mutate(across(where(is.character), as.factor))

# Dividir en datos de entrenamiento (70%) y prueba (30%)
set.seed(123) # Para reproducibilidad
trainIndex <- createDataPartition(datos_modelo$Lumbares, p = 0.7, list = FALSE)
datos_train <- datos_modelo[trainIndex, ]
datos_test <- datos_modelo[-trainIndex, ]

# Entrenar el modelo Random Forest
set.seed(123)
modelo_rf <- randomForest(Lumbares ~ ., data = datos_train, ntree = 500, mtry = 3, importance = TRUE)

# Predicciones en datos de prueba
predicciones <- predict(modelo_rf, datos_test, type = "class")

# Matriz de confusión
tabla_confusion <- confusionMatrix(predicciones, datos_test$Lumbares)
tabla_confusion

conf_mat <- tabla_confusion$table

rownames(conf_mat) <- paste("Predicho", rownames(conf_mat))
colnames(conf_mat) <- paste("Referencia", colnames(conf_mat))

# Extraer métricas importantes del modelo

```

```

accuracy      <- round(tabla_confusion$overall["Accuracy"], 4)
kappa         <- round(tabla_confusion$overall["Kappa"], 4)
sensitivity   <- round(tabla_confusion$byClass["Sensitivity"], 4)
specificity   <- round(tabla_confusion$byClass["Specificity"], 4)

# Crear tabla de métricas
metrics_table <- data.frame(
  Valor = c(accuracy, kappa, sensitivity, specificity)
)

# Imprimir tabla de matriz de confusión
kable(conf_mat, caption = "Matriz de confusión", align = "c") %>%
  kable_styling(full_width = FALSE, position = "center")

# Imprimir tabla de métricas
kable(metrics_table, caption = "Métricas del modelo", align = "c") %>%
  kable_styling(full_width = FALSE, position = "center")

```

Muñeca

Modelo

```

vars_modelo <- c("Munyeca", "EDAD", "GENERO", "BMI_categoria", "PESO_kg",
                 "dias_quirofano",
                 "Duracion_de_las_intervenciones",
                 "Ubicación_de_pantallas_laparoscopia",
                 "Total_Cirugias_Año", "menor", "cirugiaEndo",
                 "Iinstrumental_quirurgico", "factores_causa_molestias")

datos_modelo <- datos %>%
  select(all_of(vars_modelo))
modelo_inicial <- glm(Munyeca ~ ., data = datos_modelo, family = binomial)
# Seleccionar el mejor modelo eliminando variables no significativas (
p > 0.05)
modelo_optimo_Muñeca <- stepAIC(modelo_inicial,
                                      direction = "both",
                                      scope = list(lower = Munyeca ~ GENERO, upper = mo
delo_inicial),
                                      trace = FALSE)# Mostrar resumen del modelo f
inal
summary(modelo_optimo_Muñeca)

```

```

## 
## Call:
## glm(formula = Munyeca ~ EDAD + GENERO + Duracion_de_las_intervenciones +
##     Ubicación_de_pantallas_laparoscopia + cirugiaEndo + Iinstrumental_quirurgico,
##     family = binomial, data = datos_modelo)
## 
## Coefficients:
##                               Estimate Std. Error
## (Intercept)                0.119385  0.441321
## EDAD                     -0.032872  0.007668
## GENEROHombre              -0.112948  0.178807
## Duracion_de_las_intervencionesModerada-Importante 0.429359  0.266906
## Ubicación_de_pantallas_laparoscopiaModerada-Importante 0.299792  0.184252
## cirugiaEndoSí              0.642736  0.190751
## Iinstrumental_quirurgicoModerada-Importante 0.615482  0.183148
##                               z value Pr(>|z|)
## (Intercept)                0.271 0.786762
## EDAD                     -4.287 1.81e-05 ***
## GENEROHombre              -0.632 0.527597
## Duracion_de_las_intervencionesModerada-Importante 1.609 0.107692
## Ubicación_de_pantallas_laparoscopiaModerada-Importante 1.627 0.103721
## cirugiaEndoSí              3.369 0.000753 ***
## Iinstrumental_quirurgicoModerada-Importante 3.361 0.000778 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for binomial family taken to be 1)
## 
## Null deviance: 898.89 on 648 degrees of freedom
## Residual deviance: 823.51 on 642 degrees of freedom
## AIC: 837.51
## 
## Number of Fisher Scoring iterations: 4

```

Odds Ratio

```

tabla_muñeca <- modelo_optimo_Muñeca %>%
 tbl_regression(
  exponentiate = TRUE,                                     # mostrar OR
  estimate_fun = ~style_sigfig(.x, digits = 2), # OR con 2 cifras
  conf.int = TRUE,                                       # incluir IC
  conf.level = 0.95                                      # nivel 95%
) %>%
  modify_table_body(
  ~ .x %>%
    # 1) crear '95% CI', vacía si no hay conf.Low
    mutate(
      `95% CI` = ifelse(
        !is.na(conf.low),
        paste0("(", sprintf("%.2f", conf.low), ", ", sprintf("%.2f",
        conf.high), ")"),
        ""
      )
    ) %>%
    # 2) quitar conf.low/conf.high
    select(-conf.low, -conf.high) %>%
    # 3) reordenar columnas: label (o variable), estimate, 95% CI,
    p.value, Luego resto
    select(
      label,          # el texto de la característica
      estimate,       # el OR
      `95% CI`,       # nuestro intervalo formateado
      p.value,        # p-value
      everything()   # AIC, BIC, etc. quedarán abajo
    )

```

```

) %>%
# 4) renombrar encabezados
modify_header(
  label ~ "**Characteristic**",
  estimate ~ "**Cuello (OR)**",
  p.value ~ "**p-value**",
  `95% CI` ~ "**95 % CI**"
) %>%
# 5) añadir métricas de modelo al pie
add_glance_table(
  include = c("AIC", "BIC", "logLik", "deviance", "nobs")
)

```

tabla_muñeca

Random Forest

```

variables_seleccionadas <- all.vars(formula(modelo_optimo_Muñeca))[-1] # Quita "Hombros"

# Asegurar que La variable dependiente está incluida
vars_modelo_final <- c("Munyeca", variables_seleccionadas)

# Seleccionar solo las variables finales
datos_modelo <- datos %>% select(all_of(vars_modelo_final))

# Convertir variables categóricas en factores
datos_modelo <- datos_modelo %>% mutate(across(where(is.character), as.factor))

# Dividir en datos de entrenamiento (70%) y prueba (30%)
set.seed(123) # Para reproducibilidad
trainIndex <- createDataPartition(datos_modelo$Munyeca, p = 0.7, list = FALSE)
datos_train <- datos_modelo[trainIndex, ]
datos_test <- datos_modelo[-trainIndex, ]

# Entrenar el modelo Random Forest
set.seed(123)
modelo_rf <- randomForest(Munyeca ~ ., data = datos_train, ntree = 500, mtry = 3
, importance = TRUE)

# Predicciones en datos de prueba
predicciones <- predict(modelo_rf, datos_test, type = "class")

# Matriz de confusión
tabla_confusion <- confusionMatrix(predicciones, datos_test$Munyeca)
tabla_confusion

conf_mat <- tabla_confusion$table

rownames(conf_mat) <- paste("Predicho", rownames(conf_mat))
colnames(conf_mat) <- paste("Referencia", colnames(conf_mat))

```

```

# Extraer métricas importantes del modelo
accuracy <- round(tabla_confusion$overall["Accuracy"], 4)
kappa <- round(tabla_confusion$overall["Kappa"], 4)
sensitivity <- round(tabla_confusion$byClass["Sensitivity"], 4)
specificity <- round(tabla_confusion$byClass["Specificity"], 4)

# Crear tabla de métricas
metrics_table <- data.frame(
  Valor = c(accuracy, kappa, sensitivity, specificity)
)

# Imprimir tabla de matriz de confusión
kable(conf_mat, caption = "Matriz de confusión", align = "c") %>%
  kable_styling(full_width = FALSE, position = "center")

# Imprimir tabla de métricas
kable(metrics_table, caption = "Métricas del modelo", align = "c") %>%
  kable_styling(full_width = FALSE, position = "center")

```

Manos y dedos

Modelo

```

#Para la variable Manos_y_dedos
vars_modelo <- c("Manos_y_dedos", "EDAD", "GENERO", "BMI_categoria", "PESO_kg",
  "dias_quirofano",
  "Duracion_de_las_intervenciones",
  "Ubicación_de_pantallas_laparoscopia",
  "Total_Cirugias_Año", "menor", "cirugiaEndo",
  "Instrumental_quirurgico", "factores_causa_molestias")

datos_modelo <- datos %>%
  select(all_of(vars_modelo))
modelo_inicial <- glm(Manos_y_dedos ~ ., data = datos_modelo, family = binomial)
# Seleccionar el mejor modelo eliminando variables no significativas (p > 0.05)
modelo_optimo_Manos <- stepAIC(modelo_inicial,
  direction = "both",
  scope = list(lower = Manos_y_dedos ~ GENERO, upper = modelo_inicial),
  trace = FALSE)# Mostrar resumen del modelo final
summary(modelo_optimo_Manos)

```

```

## 
## Call:
## glm(formula = Manos_y_dedos ~ EDAD + GENERO + Duracion_de_las_intervenciones +
##      Ubicación_de_pantallas_laparoscopia + Iinstrumental_quirurgico,
##      family = binomial, data = datos_modelo)
## 
## Coefficients:
##                               Estimate Std. Error
## (Intercept)                0.277448  0.389164
## EDAD                     -0.016801  0.007269
## GENEROHombre               -0.338207  0.178240
## Duracion_de_las_intervencionesModerada-Importante  0.596542  0.253351
## Ubicación_de_pantallas_laparoscopiaModerada-Importante  0.362259  0.185689
## Iinstrumental_quirurgicoModerada-Importante          0.740222  0.184688
##                               z value Pr(>|z|)
## (Intercept)                0.713   0.4759
## EDAD                     -2.312   0.0208 *
## GENEROHombre               -1.897   0.0578 .
## Duracion_de_las_intervencionesModerada-Importante  2.355   0.0185 *
## Ubicación_de_pantallas_laparoscopiaModerada-Importante  1.951   0.0511 .
## Iinstrumental_quirurgicoModerada-Importante          4.008 6.12e-05 ***
## ...
## Signif. codes:  0 '****' 0.001 '***' 0.01 '**' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for binomial family taken to be 1)
## 
## Null deviance: 870.57 on 648 degrees of freedom
## Residual deviance: 811.73 on 643 degrees of freedom
## AIC: 823.73
## 
## Number of Fisher Scoring iterations: 4

```

Odds Ratio

```

tabla_manos <- modelo_optimo_Manos %>%
 tbl_regression(
  exponentiate = TRUE,                               # mostrar OR
  estimate_fun = ~style_sigfig(.x, digits = 2), # OR con 2 cifras
  conf.int = TRUE,                                 # incluir IC
  conf.level = 0.95                                # nivel 95%
) %>%
  modify_table_body(
  ~ .x %>%
    # 1) crear '95% CI', vacía si no hay conf.Low
    mutate(
      `95% CI` = ifelse(
        !is.na(conf.low),
        paste0("(", sprintf("%.2f", conf.low), ", ", sprintf("%.2f",
        conf.high), ")"),
        ""
      )
    ) %>%
    # 2) quitar conf.low/conf.high
    select(-conf.low, -conf.high) %>%
    # 3) reordenar columnas: label (o variable), estimate, 95% CI,
    p.value, luego resto
    select(
      label,          # el texto de la característica
      estimate,       # el OR
      `95% CI`,      # nuestro intervalo formateado
      p.value,        # p-value
      everything()   # AIC, BIC, etc. quedarán abajo
    )
) %>%

```

```

# 4) renombrar encabezados
modify_header(
  label ~ "**Characteristic**",
  estimate ~ "**Cuello (OR)**",
  p.value ~ "**p-value**",
  `95% CI` ~ "**95 % CI**"
) %>%
# 5) añadir métricas de modelo al pie
add_glance_table(
  include = c("AIC", "BIC", "logLik", "deviance", "nobs")
)

tabla_manos

```

Random Forest

```

variables_seleccionadas <- all.vars(formula(modelo_optimo_Manos))[-1] # Quita "Hombros"

# Asegurar que la variable dependiente está incluida
vars_modelo_final <- c("Manos_y_dedos", variables_seleccionadas)

# Seleccionar solo las variables finales
datos_modelo <- datos %>% select(all_of(vars_modelo_final))
# Convertir variables categóricas en factores
datos_modelo <- datos_modelo %>% mutate(across(where(is.character), as.factor))

# Dividir en datos de entrenamiento (70%) y prueba (30%)
set.seed(123) # Para reproducibilidad
trainIndex <- createDataPartition(datos_modelo$Manos_y_dedos, p = 0.7, list = FALSE)
datos_train <- datos_modelo[trainIndex, ]
datos_test <- datos_modelo[-trainIndex, ]

# Entrenar el modelo Random Forest
set.seed(123)
modelo_rf <- randomForest(Manos_y_dedos ~ ., data = datos_train, ntree = 500, mtry = 3, importance = TRUE)

# Predicciones en datos de prueba
predicciones <- predict(modelo_rf, datos_test, type = "class")

# Matriz de confusión
tabla_confusion <- confusionMatrix(predicciones, datos_test$Manos_y_dedos)
tabla_confusion

conf_mat <- tabla_confusion$table

rownames(conf_mat) <- paste("Predicho", rownames(conf_mat))
colnames(conf_mat) <- paste("Referencia", colnames(conf_mat))

# Extraer métricas importantes del modelo

```

```

accuracy      <- round(tabla_confusion$overall["Accuracy"], 4)
kappa         <- round(tabla_confusion$overall["Kappa"], 4)
sensitivity   <- round(tabla_confusion$byClass["Sensitivity"], 4)
specificity   <- round(tabla_confusion$byClass["Specificity"], 4)

# Crear tabla de métricas
metrics_table <- data.frame(
  Valor = c(accuracy, kappa, sensitivity, specificity)
)

# Imprimir tabla de matriz de confusión
kable(conf_mat, caption = "Matriz de confusión", align = "c") %>%
  kable_styling(full_width = FALSE, position = "center")

# Imprimir tabla de métricas
kable(metrics_table, caption = "Métricas del modelo", align = "c") %>%
  kable_styling(full_width = FALSE, position = "center")

```

Rodillas

Modelo

```

vars_modelo <- c("Rodillas", "EDAD", "GENERO", "BMI_categoria", "PESO_kg",
                 "dias_quirofano",
                 "Duracion_de_las_intervenciones",
                 "Ubicación_de_pantallas_laparoscopia",
                 "Total_Cirugias_Año", "menor", "cirugiaEndo",
                 "Iinstrumental_quirurgico", "factores_causa_molestias")

datos_modelo <- datos %>%
  select(all_of(vars_modelo))
modelo_inicial <- glm(Rodillas ~ .,
                        data = datos_modelo, family = binomial)
# Seleccionar el mejor modelo eliminando variables no significativas (
p > 0.05)
modelo_optimo_Rodillas <- stepAIC(modelo_inicial,
                                      direction = "both",
                                      scope = list(lower = Rodillas ~ GENERO, upper = mo
delo_inicial),
                                      trace = FALSE)# Mostrar resumen del modelo f
inal

summary(modelo_optimo_Rodillas)

conf_mat <- tabla_confusion$table

rownames(conf_mat) <- paste("Predicho", rownames(conf_mat))
colnames(conf_mat) <- paste("Referencia", colnames(conf_mat))

# Extraer métricas importantes del modelo
accuracy      <- round(tabla_confusion$overall["Accuracy"], 4)
kappa         <- round(tabla_confusion$overall["Kappa"], 4)
sensitivity   <- round(tabla_confusion$byClass["Sensitivity"], 4)
specificity   <- round(tabla_confusion$byClass["Specificity"], 4)

```

```

# Crear tabla de métricas
metrics_table <- data.frame(
  Valor = c(accuracy, kappa, sensitivity, specificity)
)

# Imprimir tabla de matriz de confusión
kable(conf_mat, caption = "Matriz de confusión", align = "c") %>%
  kable_styling(full_width = FALSE, position = "center")

# Imprimir tabla de métricas
kable(metrics_table, caption = "Métricas del modelo", align = "c") %>%

```

```

## 
## Call:
## glm(formula = Rodillas ~ GENERO + Duracion_de_las_intervenciones +
##     Ubicación_de_pantallas_laparoscopia, family = binomial,
##     data = datos_modelo)
## 
## Coefficients:
##                               Estimate Std. Error
## (Intercept)                 -1.3350   0.2826
## GENEROHombre                 -0.2597   0.1690
## Duracion_de_las_intervencionesModerada-Importante      0.7984   0.2908
## Ubicación_de_pantallas_laparoscopiaModerada-Importante  0.3556   0.1680
##                               z value Pr(>|z|)
## (Intercept)                 -4.724 2.31e-06 ***
## GENEROHombre                 -1.537 0.12433
## Duracion_de_las_intervencionesModerada-Importante      2.745 0.00604 **
## Ubicación_de_pantallas_laparoscopiaModerada-Importante  2.117 0.03424 *
## --- 
## Signif. codes:  0 '****' 0.001 '***' 0.01 '**' 0.05 '*' 0.1 ' ' 1
## 
## (Dispersion parameter for binomial family taken to be 1)
## 
## Null deviance: 851.93  on 648  degrees of freedom
## Residual deviance: 833.80  on 645  degrees of freedom
## AIC: 841.8
## 
## Number of Fisher Scoring iterations: 4

```

Odds Ratio

```

tabla_rodillas <- modelo_optimo_Rodillas %>%
 tbl_regression(
  exponentiate = TRUE,                               # mostrar OR
  estimate_fun = ~style_sigfig(.x, digits = 2), # OR con 2 cifras
  conf.int     = TRUE,                               # incluir IC
  conf.level   = 0.95                               # nivel 95%
) %>%
  modify_table_body(
  ~ .x %>%
    # 1) crear '95% CI', vacía si no hay conf.Low
    mutate(
      `95% CI` = ifelse(
        !is.na(conf.low),
        paste0("(", sprintf("%.2f", conf.low), ", ", sprintf("%.2f",
        conf.high), ")"),
        ""))
  )
) %>%
  # 2) quitar conf.Low/conf.high

```

```

        select(-conf.low, -conf.high) %>%
    # 3) reordenar columnas: label (o variable), estimate, 95% CI,
p.value, luego resto
    select(
        label,          # el texto de la característica
        estimate,       # el OR
        `95% CI`,      # nuestro intervalo formateado
        p.value,        # p-value
        everything()   # AIC, BIC, etc. quedarán abajo
    )
) %>%
# 4) renombrar encabezados
modify_header(
    label ~ "**Characteristic**",
    estimate ~ "**Cuello (OR)**",
    p.value ~ "**p-value**",
    `95% CI` ~ "**95 % CI**"
) %>%
# 5) añadir métricas de modelo al pie
add_glance_table(
    include = c("AIC", "BIC", "logLik", "deviance", "nobs")
)

tabla_rodillas

```

Random Forest

```

variables_seleccionadas <- all.vars(formula(modelo_optimo_Rodillas))[-1] # Quita "Hombros"

# Asegurar que la variable dependiente está incluida
vars_modelo_final <- c("Rodillas", variables_seleccionadas)

# Seleccionar solo las variables finales
datos_modelo <- datos %>% select(all_of(vars_modelo_final))

# Convertir variables categóricas en factores
datos_modelo <- datos_modelo %>% mutate(across(where(is.character), as.factor))

# Dividir en datos de entrenamiento (70%) y prueba (30%)
set.seed(123) # Para reproducibilidad
trainIndex <- createDataPartition(datos_modelo$Rodillas, p = 0.7, list = FALSE)
datos_train <- datos_modelo[trainIndex, ]
datos_test <- datos_modelo[-trainIndex, ]

# Entrenar el modelo Random Forest
set.seed(123)
modelo_rf <- randomForest(Rodillas ~ ., data = datos_train, ntree = 500, mtry = 3, importance = TRUE)

# Predicciones en datos de prueba
predicciones <- predict(modelo_rf, datos_test, type = "class")

# Matriz de confusión

```

```

tabla_confusion <- confusionMatrix(predicciones, datos_test$Rodillas)
tabla_confusion

conf_mat <- tabla_confusion$table

rownames(conf_mat) <- paste("Predicho", rownames(conf_mat))
colnames(conf_mat) <- paste("Referencia", colnames(conf_mat))

# Extraer métricas importantes del modelo
accuracy <- round(tabla_confusion$overall["Accuracy"], 4)
kappa <- round(tabla_confusion$overall["Kappa"], 4)
sensitivity <- round(tabla_confusion$byClass["Sensitivity"], 4)
specificity <- round(tabla_confusion$byClass["Specificity"], 4)

# Crear tabla de métricas
metrics_table <- data.frame(
  Valor = c(accuracy, kappa, sensitivity, specificity)
)

# Imprimir tabla de matriz de confusión
kable(conf_mat, caption = "Matriz de confusión", align = "c") %>%
  kable_styling(full_width = FALSE, position = "center")

# Imprimir tabla de métricas
kable(metrics_table, caption = "Métricas del modelo", align = "c") %>%
  kable_styling(full_width = FALSE, position = "center")

```

Tobillo/Pie

Modelo

```

vars_modelo <- c("Tobillo_pie", "EDAD", "GENERO", "BMI_categoria", "PESO_kg",
  "dias_quirofano",
  "Duracion_de_las_intervenciones",
  "Ubicación_de_pantallas_laparoscopia",
  "Total_Cirugias_Año", "menor", "cirugiaEndo",
  "Instrumental_quirurgico", "factores_causa_molestias")

datos_modelo <- datos %>%
  select(all_of(vars_modelo))
modelo_inicial <- glm(Tobillo_pie ~ ., data = datos_modelo, family = binomial)
# Seleccionar el mejor modelo eliminando variables no significativas (p > 0.05)
modelo_optimo_Tobillo <- stepAIC(modelo_inicial,
  direction = "both",
  scope = list(lower = Tobillo_pie ~ GENERO, upper = modelo_inicial),
  trace = FALSE)# Mostrar resumen del modelo final
summary(modelo_optimo_Tobillo)

```

```

## 
## Call:
## glm(formula = Tobillo_pie ~ EDAD + GENERO + PESO_kg + Duracion_de_las_intervenciones +
##     Ubicación_de_pantallas_laparoscopia + factores_causa_molestias,
##     family = binomial, data = datos_modelo)
## 
## Coefficients:
##                               Estimate Std. Error
## (Intercept)                -4.820775  0.937002
## EDAD                     -0.035169  0.009716
## GENEROHomme                -0.528671  0.269541
## PESO_kg                      0.036658  0.008858
## Duracion_de_las_intervencionesModerada-Importante 1.636728  0.614759
## Ubicación_de_pantallas_laparoscopiaModerada-Importante 0.626828  0.211140
## factores_causa_molestiasModerada-Importante          0.753724  0.413392
##                               z value Pr(>|z|)
## (Intercept)                -5.145 2.68e-07 ***
## EDAD                      -3.620 0.000295 ***
## GENEROHomme                -1.961 0.049835 *
## PESO_kg                      4.139 3.50e-05 ***
## Duracion_de_las_intervencionesModerada-Importante 2.662 0.007759 **
## Ubicación_de_pantallas_laparoscopiaModerada-Importante 2.969 0.002990 **
## factores_causa_molestiasModerada-Importante          1.823 0.068263 .
## 
## ---
## Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for binomial family taken to be 1)
## 
## Null deviance: 650.08 on 648 degrees of freedom
## Residual deviance: 590.98 on 642 degrees of freedom
## AIC: 604.98
## 
## Number of Fisher Scoring iterations: 6

```

Odds Ratio

```

tabla_tobillo <- modelo_optimo_Tobillo %>%
 tbl_regression(
  exponentiate = TRUE,                               # mostrar OR
  estimate_fun = ~style_sigfig(.x, digits = 2), # OR con 2 cifras
  conf.int     = TRUE,                               # incluir IC
  conf.level   = 0.95                               # nivel 95%
) %>%
  modify_table_body(
  ~ .x %>%
    # 1) crear '95% CI', vacía si no hay conf.Low
    mutate(
      `95% CI` = ifelse(
        !is.na(conf.low),
        paste0("(", sprintf("%.2f", conf.low), ", ", sprintf("%.2f",
        conf.high), ")"),
        ""
      )
    ) %>%
    # 2) quitar conf.low/conf.high
    select(-conf.low, -conf.high) %>%
    # 3) reordenar columnas: label (o variable), estimate, 95% CI,
    p.value, luego resto
    select(
      label,          # el texto de la característica
      estimate,       # el OR
      `95% CI`,      # nuestro intervalo formateado
      p.value,        # p-value
      everything()   # AIC, BIC, etc. quedarán abajo
    )

```

```

) %>%
# 4) renombrar encabezados
modify_header(
  label ~ "**Characteristic**",
  estimate ~ "**Cuello (OR)**",
  p.value ~ "**p-value**",
  `95% CI` ~ "**95 % CI**"
) %>%
# 5) añadir métricas de modelo al pie
add_glance_table(
  include = c("AIC", "BIC", "logLik", "deviance", "nobs")
)

```

tabla_tobillo

Random Forest

```

variables_seleccionadas <- all.vars(formula(modelo_optimo_Tobillo))[-1] # Quita "Hombros"

# Asegurar que La variable dependiente está incluida
vars_modelo_final <- c("Tobillo_pie", variables_seleccionadas)

# Seleccionar solo las variables finales
datos_modelo <- datos %>% select(all_of(vars_modelo_final))

# Convertir variables categóricas en factores
datos_modelo <- datos_modelo %>% mutate(across(where(is.character), as.factor))

# Dividir en datos de entrenamiento (70%) y prueba (30%)
set.seed(123) # Para reproducibilidad
trainIndex <- createDataPartition(datos_modelo$Tobillo_pie, p = 0.7, list = FALSE)
) 
datos_train <- datos_modelo[trainIndex, ]
datos_test <- datos_modelo[-trainIndex, ]

# Entrenar el modelo Random Forest
set.seed(123)
modelo_rf <- randomForest(Tobillo_pie ~ ., data = datos_train, ntree = 500, mtry = 3, importance = TRUE)

# Predicciones en datos de prueba
predicciones <- predict(modelo_rf, datos_test, type = "class")

# Matriz de confusión
tabla_confusion <- confusionMatrix(predicciones, datos_test$Tobillo_pie)
tabla_confusion

conf_mat <- tabla_confusion$table

rownames(conf_mat) <- paste("Predicho", rownames(conf_mat))
colnames(conf_mat) <- paste("Referencia", colnames(conf_mat))

```

```

# Extraer métricas importantes del modelo
accuracy      <- round(tabla_confusion$overall["Accuracy"], 4)
kappa         <- round(tabla_confusion$overall["Kappa"], 4)
sensitivity   <- round(tabla_confusion$byClass["Sensitivity"], 4)
specificity   <- round(tabla_confusion$byClass["Specificity"], 4)

# Crear tabla de métricas
metrics_table <- data.frame(
  Valor = c(accuracy, kappa, sensitivity, specificity)
)

# Imprimir tabla de matriz de confusión
kable(conf_mat, caption = "Matriz de confusión", align = "c") %>%
  kable_styling(full_width = FALSE, position = "center")

# Imprimir tabla de métricas
kable(metrics_table, caption = "Métricas del modelo", align = "c") %>%
  kable_styling(full_width = FALSE, position = "center")

```

Calculadora

```

library(shiny)
library(shinythemes)
library(dplyr)

# Función para calcular la probabilidad de dolor
calcular_probabilidad <- function(coeficientes, datos) {
  eta <- sum(coeficientes * datos)
  prob <- 1 / (1 + exp(-eta))
  return(prob)
}

ui <- fluidPage(
  theme = shinytheme("flatly"),

  tags$head(
    tags$style(HTML("
      html, body {
        background-color: #dfefff;
        min-height: 100vh;
        height: auto;
        overflow-x: hidden;
      }
      .panel {
        background-color: white;
        padding: 30px;
        border-radius: 15px;
        box-shadow: 0px 0px 15px rgba(0, 0, 0, 0.2);
        margin-bottom: 20px;
      }
      .btn {
        background-color: #0077b6 !important;
        color: white !important;
        width: 100%;
        font-size: 18px;
        padding: 10px;
      }
      .btn:hover {
        background-color: #006699 !important;
      }
    "))
  )
)
```

```

        background-color: #005f8a !important;
    }
    #resultado {
        background-color: #f8f9fa;
        padding: 15px;
        border-radius: 10px;
        font-size: 18px;
        font-weight: bold;
        text-align: center;
        width: 100% !important;
        word-wrap: break-word;
    }
    .container-fluid {
        max-width: 1400px;
    }
    "))
),
titlePanel(
    HTML("Calculadora")
),
fluidRow(
    column(12,
        # Panel de inputs básicos
        div(class = "panel",
            h4("[[ Información Básica"),
            fluidRow(
                column(3, numericInput("edad", "Edad:", value = NULL, min = 25, max = 70, step = 1)),
                column(3, radioButtons("genero", "Género:", choices = c("Hombre", "Mujer"), inline = TRUE) ,
                    selected = character(0)),
                column(3, numericInput("peso", "Peso (kg):", value = NULL, min = 40, max = 120, step = 1)),
                column(3, numericInput("talla", "Talla (m):", value = NULL, min = 1.40, max = 2.00, step = 0.01))
            )
        ),
        # Panel de ergonomía
        div(class = "panel",
            h4("※ Medidas para reducir el dolor"),
            fluidRow(
                column(6,
                    radioButtons("duracion_interv", "Reducir tiempo en quirófano :",
                        choices = c("Nunca"=1, "A veces"=1, "A menudo"=0, "Siempre"=0) ,
                        selected = character(0))
                ),
                column(6,
                    radioButtons("ubicacion_pantallas", "Monitor a altura adecuada :",
                        choices = c("Nunca"=1, "A veces"=1, "A menudo"=0, "Siempre"=0) ,
                        selected = character(0))
                )
            ),
            fluidRow(
                column(6,
                    radioButtons("instrumental", "Instrumental adecuado:",
                        choices = c("Nunca"=1, "A veces"=1, "A menudo"=0, "Siempre"=0) ,
                        selected = character(0))
                )
            )
        )
    )
),

```

```

            column(6,
                    radioButtons("factores_molestias","Micro-descansos en cirugía
s largas:",
                                choices = c("Nunca"=1,"A veces"=1,"A menudo"=0,
"Siempre"=0) ,
                                selected = character(0))
                )
            ),
            # Panel operaciones
            div(class = "panel",
                h4("📊 Información sobre Cirugías"),
                fluidRow(
                    column(6,
                            radioButtons("total_cirugias","Total de Cirugías al Año:",
                                choices = c("<25","25-50",>50"),
                                selected = character(0))
                ),
                column(6,
                            radioButtons("cirugia_endo","Cirugía Endoscópica:",
                                choices = c("No"=0,"Sí"=1), inline = TRUE) ,
                            selected = character(0))
                )
            ),
            # Botón de cálculo
            actionButton("calcular","🔍 Calcular", class="btn btn-lg"),
            # Resultado
            hr(),
            h3("📊 Resultados", style="text-align:center;"),
            div(id="resultado",
                verbatimTextOutput("resultado")
            )
        )
    )
}

server <- function(input, output) {
  observeEvent(input$calcular, {
    # Calcula BMI y categoría
    bmi <- input$peso / (input$talla^2)
    bmi_cat <- case_when(
      bmi < 25 ~ "healthy",
      bmi >= 25 & bmi<30 ~ "overweight",
      TRUE ~ "obese"
    )
    # Vector de datos usuario
    datos_usuario <- c(
      1, # intercepto
      ifelse(input$genero=="Hombre",1,0), # masculino
      input$edad,
      input$peso,
      ifelse(bmi_cat=="overweight",1,0),
      ifelse(bmi_cat=="obese", 1,0),
      as.numeric(input$duracion_interv),
      as.numeric(input$ubicacion_pantallas),
      as.numeric(input$instrumental),
      as.numeric(input$factores_molestias),
      ifelse(input$total_cirugias=="25-50",1,0),
      ifelse(input$total_cirugias==">50", 1,0),
      as.numeric(input$cirugia_endo)
    )
  })
}

```

```

# Coeficientes (ejemplo simplificado)
coeficientes <- list(
  Cuello    = c(2.8618,-0.2679,-0.0536,0,0,0,1.3006,0.9311,0,0,0,0,0),
  Hombros   = c(0.8761,-0.2094,-0.0154,-0.014,0,0,1.0710,0,0.4063,0,0.3568,0.9747,
0),
  Dorsal    = c(0.8884,-0.0517,-0.0324,0,-0.3305,0.3953,0.4171,0.4432,0.3634,0.598
9,0,0,0),
  Lumbares  = c(0.3524,0,-0.0232,0.014,0,0,0.8629,0.3958,0,0.5056,0,0,0),
  Muñeca    = c(0.119,-0.113,-0.0329,0,0,0,0.4294,0.2998,0.6155,0,0,0,0.6427),
  Manos     = c(0.2774,-0.3382,-0.0168,0,0,0,0.5965,0.3623,0.7402,0,0,0,0),
  Rodillas  = c(-1.3350,-0.2597,0,0,0,0.7984,0.3556,0,0,0,0,0),
  Tobillo   = c(-4.8208,-0.5287,-0.0352,0.0367,0,0,1.6367,0.6268,0,0.7537,0,0,0)
)

# Calcula todas las probabilidades
probs <- lapply(coeficientes, calcular_probabilidad, datos = datos_usuario)

output$resultado <- renderText({
  líneas <- paste0(names(probs), ":", round(unlist(probs) * 100,2), "%")
  paste(líneas, collapse = "\n")
})
})
}
}

shinyApp(ui, server)

```

Encuesta ErgoSurgeon

Estudio sobre riesgos musculoesqueléticos asociados al trabajo en quirófano - Cirugía General

Se solicita su autorización para participar en el proyecto de investigación titulado "Estudio sobre riesgos musculoesqueléticos asociados al trabajo en quirófano" cuyo investigador/a principal es D. Antonio Arroyo en colaboración con la Universidad Miguel Hernández, y cuyo objetivo es valorar las circunstancias que predisponen a los/las cirujanos/as a sufrir dolor y trastornos musculoesqueléticos. Su participación en el estudio, consiste en contestar a una serie de preguntas en el siguiente formulario online.

Los beneficios que se esperan de este trabajo son el conocimiento de los factores de riesgo para estas dolencias y su potencial prevención en un futuro. La recogida de datos del estudio se realizará desde el 1/10/2021 hasta 28/2/2022.

La participación en éste estudio es totalmente voluntaria, si usted no desea participar en el estudio, no habrá ninguna consecuencia negativa para usted. En cualquier momento puede retirarse del estudio dejando de cumplimentar la encuesta, sin que ello tenga ninguna consecuencia. La respuesta es completamente anónima, por lo que no se dispondrá de ningún dato que pueda identificarle.

Si tienen alguna pregunta sobre este proyecto de investigación, puede consultar en cualquier momento al investigador Luis Sánchez-Guillén cuyo e-mail es drsanchezguillen@gmail.com. Si usted responde a las cuestiones que se le propone, se entiende de forma tácita que ha comprendido el objetivo del presente estudio, que ha podido preguntar y aclarar las dudas que se le hubieran planteado inicialmente y que acepta participar en el estudio.

El equipo investigador le agradece su tiempo y su valiosa participación.

Sección 1. Variables individuales

¿Es soci@ de la AEC? Si No

Indique por favor, su EDAD (años): _____

¿Cuál es su género? Masculino Femenino No binario

¿Cuál es su PESO (kg)? _____

¿Y su ESTATURA (m)?: _____

¿Cuál es su MANO DOMINANTE? Derecha Izquierda Ambidiestro

¿En qué ámbito desempeña su labor asistencial? Público Privado Mixto

Si es usted residente, indique el año de residencia en que se encuentra:

MIR 1 MIR 2 MIR 3 MIR 4 MIR 5

Si es usted especialista, ¿Cuántos años lleva trabajando como CIRUJANO?

Menos de 10 años Entre 10 y 20 años Más de 20 años

¿Realiza usted habitualmente guardias? Si No

Indique la media de días a la semana que realiza al menos 30 minutos de actividad física/deportiva:

1 2 3 4 5 6 7

¿Cuantos días a la semana (media) tiene actividad en quirófano?:

1 2 3 4 5

¿Cuántos de los siguientes procedimientos realiza al año aproximadamente?:

	<10	10-25	26-40	41-50	51-75	76-100	>100
Cirugía laparoscópica							
Cirugía menor							
Cirugía abierta							
Cirugía endoscópica							
Cirugía robótica							

Sección 2: Molestias musculo-esqueléticas asociadas a la práctica quirúrgica

En los últimos 12 meses, ¿ha experimentado usted DOLOR MUSCULOESQUELÉTICO que atribuya a su trabajo en quirófano?

Si No

Señale si ha tenido antecedentes de alguna de las patologías que se citan a continuación:

- Epicondilitis
- Patología y dolor cervical
- Tendinopatía de hombro
- Hombro congelado
- Dolor lumbar
- Síndrome del túnel carpiano
- Tendinopatía de De Quervain
- Tendinopatía del bíceps

¿Alguna vez sintió que su cuerpo estaba en una posición incómoda en la sala de operaciones? causando malestar, dolores musculares o dolores mientras se retrae, ayuda u opera?

Si No

En caso afirmativo, podría identificar las áreas en las que se produjo ese dolor, y calificar su intensidad (de 0 a 10 o cualitativo):

	Ningún dolor	Dolor leve	Dolor moderado	Dolor severo
Cuello				
Hombros				
Zona dorsal				
Lumbares				
Muñeca				
Manos y dedos				
Rodillas				
Tobillo/pie				

Y respecto a la duración de estos síntomas...podría decir que fue:

	Horas	Varios días	Unas semanas	Meses
Cuello				
Hombros				
Zona dorsal				
Lumbares				
Muñeca				
Manos y dedos				
Rodillas				
Tobillo/pie				

Por favor, describa otro tipo de molestias o lesiones no reflejadas en este cuestionario que atribuya a su práctica quirúrgica?

¿Ha necesitado asistencia en alguno de estos episodios de dolor/molestias tras salir de quirófano?

Sí No

En caso afirmativo, ¿de qué tipo?

- Farmacológico
- Sesión de fisioterapia
- Ortesis/ortopédica
- Cirugía

Esas molestias, ¿le han obligado en alguna ocasión a estar de baja/perder días de trabajo?

Sí No

Sección 3: Ergonomía

¿Ha recibido alguna formación sobre ergonomía en el quirófano o en procedimientos quirúrgicos?

Sí No

¿Ha tenido recibido alguna formación reglada, conferencias o cursos sobre ergonomía en las siguientes áreas? Marque todo lo que corresponda.

- Postura / mecánica corporal en la posición de la cabeza en quirófano, hombros, columna, pelvis
- Configuración ergonómica de mesa de operaciones, microscopio, robot
- Técnicas de transferencia de pacientes

¿Identifica o es consciente de alguna mala postura/vicio postural que pueda tener en el quirófano?

Sí No

¿Qué factores relacionados con la práctica quirúrgica/quirófano, y en qué medida considera que han podido influir en la aparición de sus molestias o síntomas musculoesqueléticos?

	Ninguna	Mínimamente	Moderada	Importante	No procede
Horas seguidas de cirugía /ausencia de breaks					
Tipos de cirugía que realizo					
Duración de las intervenciones					
Instrumental quirúrgico					
Ubicación de pantallas de laparoscopia					

¿Conoce las recomendaciones que parten del campo de la ergonomía quirúrgica y de sus estudios e investigaciones?

Sí No

Si es así, ¿cómo obtuvo esta información?

- Boca a boca, de otros compañeros
- Búsqueda en Internet
- Cursos o conferencias

- Formación específica
- Durante mi formación MIR

¿Ha aplicado esa información parte de esta a su práctica quirúrgica?

- Sí
- No

¿Implementa habitualmente estrategias de reducción de riesgos en el quirófano para ayudar a aliviar los problemas musculoesqueléticos?

- Taburetes
- Almohadillas para bipedestación
- Descansos cronometrados (minibreaks)
- Estiramiento intraoperatorio
-

¿Considera que la formación de estudiantes de medicina y residentes de cirugía en estas áreas es apropiado o suficiente?

- Sí
- No

¿Qué sugerencias tiene para mejorar el bienestar físico de los futuros cirujanos?

- Educación aplicada en ergonomía y postura
- Grabación de video y evaluación/análisis de hábitos en quirófano
- Acceso a fisioterapia en el hospital
- Disponer de personal técnico dedicado al traslado y movilización de pacientes
- Planes de prevención basados en el ejercicio físico

Para finalizar...

¿Tiene algún comentario o consejo para los nuevos cirujanos para evitar / minimizar estas posibles dolencias laborales asociadas al trabajo en quirófano?

.....
.....
.....
.....

Si desea recibir resultados de este estudio, así como material audiovisual que se pueda generar a partir de los mismo, escriba su correo electrónico:

¡MUCHAS GRACIAS POR SU PARTICIPACIÓN!

