

# **Impacto del agotamiento emocional en los estudiantes del grado de medicina.**

**TRABAJO DE FIN DE GRADO  
GRADO EN ESTADÍSTICA EMPRESARIAL  
FACULTAD DE CIENCIAS SOCIALES Y JURÍDICAS DE ELCHE**



**Curso académico 2022 - 2023**

**Autor: Vidal Ocaña Alonso**

**Tutora: María Asunción Martínez Mayoral**

## **Índice de contenidos**

### 1. Resumen

### 2. Palabras clave

### 3. Objetivos

#### 3.1 Antecedentes

#### 3.2 Objetivos generales

#### 3.3 Objetivos específicos

### 4. Información disponible

#### Variables relacionadas con la salud

#### Variables relacionadas con características físicas

#### Variables relacionadas con características emocionales y motivacionales

### 5. Técnicas y métodos

#### 5.1 Procesado de los datos

#### 5.2 Análisis descriptivo

#### 5.3 Metodología

##### 5.3.1 Árboles de regresión

##### 5.3.2 Modelo lineal normal

#### 5.4 Software y hardware

### 6. Resultados

#### 6.1 Análisis descriptivo

#### 6.2 Árbol de regresión

#### 6.3 Modelo lineal

### Conclusiones

### Referencias

### Anexos

## 1. Resumen

Hemos llevado a cabo un estudio exhaustivo y riguroso con el objetivo de comprender en profundidad el impacto del agotamiento emocional en los estudiantes del grado de medicina en Suiza. Para lograr este objetivo, hemos implementado una variedad de metodologías y técnicas de investigación.

En primer lugar, hemos realizado un análisis descriptivo detallado que nos ha permitido examinar de manera minuciosa las variables observadas y las asociaciones entre ellas, presentes en nuestra muestra de estudiantes. A través de este análisis, hemos podido identificar y cuantificar la prevalencia del agotamiento emocional, así como su relación con diferentes variables demográficas y académicas.

Además, hemos planteado y aplicado sendos modelos estadísticos con el propósito de comprender mejor el comportamiento de los sujetos al respecto del agotamiento emocional que sufren mientras realizan sus estudios de medicina y en función de diversas variables observadas. Estos modelos nos han permitido examinar las influencias y los factores de riesgo asociados al agotamiento emocional en los estudiantes de medicina, aportando una perspectiva más profunda y analítica a nuestro estudio, a la par que construir modelos de predicción con los que predecir a priori el agotamiento emocional en función de las variables relevantes observadas. Son modelos, pues, que permitirían en un momento dado, prevenir problemas serios al respecto de evitar un agotamiento excesivo de los estudiantes

En el estudio se han considerado tanto variables numéricas como categóricas, que nos han permitido investigar el agotamiento emocional en diversos grupos de población según el nivel académico, el género o la carga de trabajo, que en principio podrían afectar a los estudiantes en mayor o menor grado.

A través del árbol de regresión observamos que el agotamiento emocional está estrechamente ligado a la depresión psicológica. Además, se relaciona con individuos que experimentan una mayor fatiga y tienen una credibilidad baja.

Mediante el modelo lineal apreciamos que todos los factores analizados están directamente relacionados con el agotamiento emocional, excepto la eficacia académica. Los estudiantes que experimentan un mayor agotamiento emocional presentan un rendimiento académico más bajo.

## **2. Palabras clave**

En el marco de este estudio, hemos identificado una serie de palabras clave que son fundamentales para comprender y analizar el fenómeno en cuestión. Estas palabras clave son las siguientes:

- agotamiento emocional
- depresión
- estudiantado
- árbol de regresión
- modelo lineal normal.

## **3. Objetivos**

En la presente sección, se expone de manera detallada una amplia serie de propósitos establecidos en el presente estudio, los cuales abarcan tanto metas generales como específicas. Antes de sumergirnos en la presentación exhaustiva de dichos objetivos, se realiza una introducción meticulosa y completa acerca de los antecedentes relevantes relacionados con la base de datos en cuestión. Además, se comentan brevemente las investigaciones y trabajos previamente llevados a cabo en relación con esta base de datos, con el fin de establecer una sólida base de conocimientos sobre el tema. Esta introducción permite situar el contexto y proporciona una comprensión más profunda de los desafíos y oportunidades que se presentan en el estudio. A partir de este enfoque, se establecen unos objetivos de investigación claros y específicos.

### **3.1 Antecedentes**

El estudio pionero realizado por Carrard et al. (2022) empleó por primera vez una base de datos de gran relevancia con el propósito de indagar en la conexión existente entre los problemas de salud mental de los estudiantes del grado de medicina y el tipo de empatía que poseen.

El estudio contó con una muestra amplia y representativa, compuesta por un total de 886 estudiantes de medicina, abarcando desde el primer hasta el sexto año del programa de grado. La evaluación de las diferentes dimensiones cognitivas, afectivas y conductuales de la empatía se llevó a cabo mediante la utilización de cuestionarios de autoinforme y pruebas de reconocimiento de emociones, lo que permitió obtener datos sólidos y confiables. Además, se emplearon técnicas de regresión, para examinar la relación entre las dimensiones de la empatía, los síntomas depresivos, la ansiedad y el agotamiento. En este análisis, se tuvo en cuenta tanto la influencia del año curricular en el que se encontraban los participantes como el género, lo que proporcionó una visión más completa y precisa de los resultados obtenidos.

Mediante este enfoque novedoso, se logró descubrir diversos factores que hasta entonces habían sido pasados por alto, abriendo nuevas perspectivas en el campo. Por ejemplo, se constató la relevancia del curso académico en el que se encontraba cada estudiante, ya que se observaron diferencias significativas en los niveles de empatía y los problemas de salud mental entre los diferentes años de estudio. Además, se exploró la influencia del género, revelando que las mujeres presentaban una mayor incidencia de problemas de salud mental, agotamiento y niveles de empatía en comparación con los hombres.

Los hallazgos obtenidos son valiosos no solo para el ámbito académico, sino también para la comunidad médica en general, ya que aportan información relevante para abordar de manera más efectiva la salud mental de los futuros profesionales de la salud.

### **3.2 Objetivos generales**

Dado que en el estudio previo de Carrard et al (2022) se focalizó la empatía, en nuestro estudio hemos tratado de complementar aquella investigación, estudiando otra variable que puede ser crítica en los sujetos mientras cursan sus estudios universitarios: el agotamiento emocional. El objetivo general de la presente investigación consiste en analizar en profundidad el impacto que las características observadas en la muestra de estudiantes de medicina ejercen sobre el agotamiento emocional. A través de un enfoque exhaustivo, se busca comprender de manera precisa y detallada qué factores pueden contribuir al agotamiento experimentado por los estudiantes en el ámbito emocional, como la carga académica, el estrés, las demandas emocionales y otros elementos relevantes.

Se pretende explorar las diversas dimensiones del agotamiento emocional, identificando sus manifestaciones y los posibles desencadenantes dentro del contexto educativo de la carrera de medicina. Asimismo, se busca analizar cómo estos factores interactúan y se relacionan entre sí, permitiendo así establecer una visión integral sobre el fenómeno y sus implicaciones en el bienestar y la salud mental de los estudiantes.

Con este objetivo general, se espera generar conocimiento significativo que contribuya a la mejora de las condiciones de estudio y formación de los estudiantes de medicina, así como a la implementación de estrategias y políticas que promuevan su bienestar emocional y prevengan el agotamiento en este exigente campo académico.

### **3.3 Objetivos específicos**

Los objetivos específicos que se persiguen en este estudio se centran en el análisis detallado de la relación existente entre el agotamiento emocional y diversas variables de interés. Estas variables han sido agrupadas en nuestro

estudio en tres categorías principales: las relacionadas con las características físicas, las relacionadas con las características emocionales y motivacionales, y las relacionadas con la salud.

En primer lugar, nos interesamos en analizar la influencia sobre el agotamiento de las variables relacionadas con la salud, incluyendo la depresión, la ansiedad y la satisfacción que el sujeto tiene con su propia salud. Se estudian las relaciones entre ellas y cómo se interrelacionan con el agotamiento emocional e influyen en la percepción de bienestar de los participantes.

En segundo lugar, nos proponemos investigar la relación entre el agotamiento emocional y las variables asociadas a las características físicas de los individuos. Estas variables incluyen el curso en el que se encuentran los participantes, su sexo, idioma y edad. A través de un análisis estadístico buscamos identificar posibles patrones y diferencias relacionadas con estas variables físicas sobre el agotamiento emocional.

Por último, nos enfocamos en examinar las variables relacionadas con las características emocionales y motivacionales. Estas variables abarcan aspectos como la situación de pareja, el trabajo, la consulta con un psicólogo, la motivación, la incoherencia, la eficacia, las horas de estudio y las diferentes dimensiones de la empatía (cognitiva, afectiva y media). Nuestro objetivo es comprender cómo estas variables se relacionan con el agotamiento emocional y qué impacto tienen sobre los participantes.

#### **4. Información disponible**

En el contexto de este estudio, se ha recopilado y utilizado una base de datos amplia y detallada que aborda diversos aspectos relacionados con la salud mental, el agotamiento y la empatía de los estudiantes de medicina. Esta base de datos

cuenta con una notable extensión, incluyendo un total de 886 registros y un total de 20 variables distintas, además de un identificador anonimizado y único para cada sujeto que facilita su seguimiento y análisis individualizado.

Es importante destacar que esta base de datos se encuentra disponible públicamente en la plataforma de Kaggle, y en concreto en la url <https://www.kaggle.com/datasets/thedevastator/medical-student-mental-health>, lo que implica un acceso amplio y abierto para los investigadores interesados en explorar y utilizar dicha información. La disponibilidad de esta base de datos en un dominio público contribuye a fomentar la transparencia y la colaboración en el ámbito de la investigación científica.

La amplitud y diversidad de las variables disponibles en esta base de datos permite un acercamiento integral y detallado a los estudiantes de medicina en relación con su salud mental, agotamiento y empatía. Como hecho diferencial e innovador respecto del análisis original, en este estudio se ha llevado a cabo una clasificación de las variables en tres categorías principales: salud, características físicas y características emocionales y motivacionales. Esta clasificación permite una exploración y análisis más específico y estructurado de los datos, facilitando la identificación de patrones, correlaciones y tendencias relevantes.

La base de datos utilizada en este estudio representa pues, una valiosa herramienta para investigaciones futuras, brindando una oportunidad única para profundizar en la comprensión de la salud mental y el bienestar de los estudiantes de medicina, así como para desarrollar estrategias y políticas efectivas que promuevan su bienestar mental, integral y garantizar su éxito académico.

La clasificación referida de las variables disponibles se presenta a continuación.

### **Variables relacionadas con la salud**

Se trata de variables indicadoras del estado de salud del sujeto, medidas unas por un experto en escalas estandarizadas, y también referidas por el propio sujeto a modo de autopercepción. Son:

- Depresión: puntuación de depresión en una escala numérica.
- Ansiedad: puntuación de ansiedad en una escala numérica.
- Agotamiento: agotamiento emocional en una escala numérica.
- Salud: grado de satisfacción con su salud. Se trata de una pregunta de autopercepción con una respuesta categorizada en 5 niveles: Muy insatisfecho/ Insatisfecho /Ni satisfecho ni insatisfecho / Satisfecho / Muy satisfecho.

### **Variables relacionadas con características físicas**

Son variables objetivas que refieren la condición de cada sujeto en lo relativo a:

- Curso: curso académico del participante. Es una variable categórica con niveles Primero/Segundo/Tercero/Cuarto/Quinto/Sexto.
- Sexo: género del participante. Es una variable categórica con niveles Hombre/Mujer/No binario.
- Idioma: idioma materno hablado por el participante. Originariamente medido en 42 categorías, pero recodificado en este estudio en una variable dicotómica con niveles Francés/Otro.
- Edad: edad del participante. Es una variable numérica.

### **Variables relacionadas con características emocionales y motivacionales**

Se han diferenciado en tercer lugar aquellas variables relacionadas con el estado emocional de los sujetos y posibles motivaciones que pueden tener repercusión sobre la variable objetivo en el estudio. Se trata de las variables:

- Pareja: si el participante tiene pareja. Es una variable dicotómica con niveles No/Sí.

- Trabajo: si el participante tiene un trabajo remunerado. Es una variable dicotómica con niveles No/Sí.
- Psicólogo: si el participante acude al psicólogo. Es una variable dicotómica con niveles No/Sí.
- Motivación: puntuación de motivación académica del participante. Es una variable numérica.
- Incoherencia: grado de incoherencia cuantificado en la encuesta en función de las respuestas congruentes de los sujetos a varias preguntas de validación. Es una variable numérica.
- Eficacia: eficacia académica. Es una variable numérica.
- Horas de estudio: número de horas estudiadas a la semana. Es una variable numérica.
- Empatía cognitiva: Puntuación de empatía cognitiva, medida en una escala numérica.
- Empatía afectiva: Puntuación de empatía afectiva, medida en una escala numérica.
- Empatía media: Puntuación de empatía media del participante, medida en una escala numérica.

## **5. Técnicas y métodos**

En el desarrollo de este trabajo distribuimos las técnicas y métodos analíticos en tres bloques fundamentales: el procesado de los datos, el análisis exploratorio o descriptivo, y el análisis inferencial, que presentamos a continuación en detalle.

En primer lugar, en lo que respecta al procesado de los datos, hemos llevado a cabo una serie de procedimientos para organizar, limpiar y preparar los datos obtenidos de nuestra muestra. Esto incluye la revisión de la calidad de los datos, la identificación y eliminación de valores atípicos o errores, la codificación y recodificación de variables y la creación de nuevas variables derivadas cuando es necesario.

En segundo lugar, en el análisis exploratorio o descriptivo, hemos utilizado diversas técnicas para examinar y resumir los datos de manera detallada. Esto implica el cálculo de medidas estadísticas descriptivas diferenciadas en función de si las variables se han medido en una escala numérica o categórica. Así mismo, se han empleado diversos gráficos y visualizaciones para representar los datos de manera efectiva y facilitar la comprensión de los patrones y las tendencias presentes en ellos.

Por último, en el análisis inferencial, dados los objetivos predictivos descritos en el estudio, hemos optado por plantear dos aproximaciones distintas: los árboles de regresión y el modelo lineal normal. Ambos surgen del modelo de regresión, si bien tanto su funcionamiento como sus resultados e interpretación son bien distintos.

En conjunto, el uso de estas técnicas y métodos nos ha permitido realizar un análisis completo y exhaustivo de nuestros datos. El procesado de los datos nos ha asegurado la calidad y la consistencia de la información, el análisis exploratorio nos ha brindado una comprensión profunda de los datos y el análisis inferencial nos ha permitido realizar inferencias y generalizaciones válidas y confiables. El empleo de estos tres bloques metodológicos nos ha proporcionado una base sólida para responder a nuestras preguntas de investigación y alcanzar los objetivos propuestos en este estudio.

## **5.1 Procesado de los datos**

En el procesamiento de los datos hemos llevado a cabo una serie de pasos para garantizar la integridad y la adecuada preparación de la base de datos. Es de remarcar la importancia de esta etapa, que garantiza (o no), la fiabilidad de los análisis y conclusiones que se deriven a continuación.

En primer lugar, hemos realizado una comprobación exhaustiva del estado de la base de datos para identificar la presencia de valores faltantes. Para ello, hemos

utilizado la función anyNA(). Al constatar que no existen datos faltantes en nuestra muestra, no se han propuesto medidas correctoras de compleción o eliminación.

A continuación, nos hemos centrado en verificar el tipo de cada variable presente en la base de datos. Durante este proceso, hemos identificado siete variables que se clasifican como categóricas, estas son: curso académico, género del participante, idioma materno, si el sujeto tiene pareja, si el sujeto tiene trabajo, si el sujeto acude al psicólogo y el grado de satisfacción con su salud.

Con el fin de facilitar su manejo y análisis, hemos convertido estas variables en factores y hemos recodificado todos los niveles asociados a ellas en formato de texto. Inicialmente, estas variables estaban codificadas con valores numéricos, pero dado que los modelos utilizados en nuestra inferencia no requieren una codificación numérica específica para estas variables, hemos optado por convertirlas en texto para una mejor comprensión y manipulación.

Un ejemplo concreto de recodificación lo encontramos en la variable relacionada con el idioma materno. Inicialmente, esta variable presentaba 19 categorías distintas. Sin embargo, debido a las bajas frecuencias observadas en la muestra para idiomas distintos al francés, hemos decidido simplificar su clasificación a únicamente dos niveles: Francés y Otros. Esta decisión se basa en la necesidad de reducir la complejidad y mejorar la interpretación de los resultados, sin comprometer la integridad de los datos.

En resumen, el procesamiento de datos no ha requerido de un tratamiento excesivamente exhaustivo, puesto que la base de datos original disponible se había proporcionado ya depurada y sin errores. Con todo, hemos verificado la ausencia de valores faltantes, asegurado el funcionamiento de las variables categóricas como factores de clasificación, recodificado sus niveles a formato de texto y simplificado la clasificación de una variable en particular, recodificándola en solo dos niveles. Estas acciones nos han permitido tener una base de datos preparada y adecuada para los objetivos de nuestro análisis, asegurando la calidad y la coherencia de los resultados obtenidos en nuestro estudio.

## 5.2 Análisis descriptivo

En esta sección se detallan los descriptivos utilizados para describir la base de datos, en función del tipo de información disponible, esto es, de si la información se ha recopilado en variables categóricas o numéricas.

Para representar visualmente la distribución de las variables categóricas, utilizamos gráficos de barras con frecuencias relativas superpuestas, lo que permite apreciar de manera clara el reparto de la muestra en las diferentes categorías, y en consecuencia comparar estas.

Por otro lado, para visualizar la distribución de las variables numéricas, recurrimos a histogramas que nos proporcionan el rango de variación y nos dan una visión del reparto de la muestra, y otras características como la simetría de los datos.

Además de las representaciones gráficas, realizamos un análisis descriptivo detallado de las variables. Para las variables categóricas, presentamos las frecuencias relativas en escala porcentual, lo que nos permite examinar la proporción de cada categoría dentro de la muestra. En cuanto a las variables numéricas, calculamos diversos descriptivos que incluyen cuartiles, media, desviación típica y coeficientes de variación. Este último valor es especialmente útil para comprender la variabilidad de cada variable en referencia a la escala en que ha sido medida, hecho importante cuando disponemos de diferentes escalas de medición, como es el caso.

Con el fin de explorar la relación entre el agotamiento emocional y el resto de variables en la base de datos, empleamos diferentes técnicas. Para cuantificar la relación lineal del agotamiento emocional con las variables numéricas en el área de salud, utilizamos gráficos de correlaciones. Estos nos permiten visualizar de manera clara y cuantitativa la relación entre las variables.

Cuando investigamos la relación del agotamiento emocional con variables categóricas, como la satisfacción con la salud, curso académico, idioma y género, recurrimos a gráficos de cajas. Estos gráficos nos ayudan a identificar si existen

diferencias en el agotamiento emocional entre los diversos niveles de respuesta en cada una de las variables mencionadas.

Asimismo, utilizamos diagramas de dispersión condicionados para visibilizar algún tipo de tendencia en la relación entre el agotamiento emocional y las variables numéricas relacionadas con características emocionales y motivacionales, como la eficacia, empatía afectiva, horas de estudio, incoherencia y motivación, considerando a la vez diversas variables categóricas y sus niveles, como ..... Estos diagramas nos permiten evaluar si existe alguna relación lineal entre las variables numéricas y visualizar si las variables categóricas influyen en dicha relación.

Hemos aplicado pues, diversas técnicas y representaciones gráficas, adaptadas al tipo de información disponible, para comprender y explorar la muestra respecto de cada una de las variables observadas, además de su relación con el agotamiento emocional. Esta aproximación nos proporciona una visión completa y detallada de la muestra, así como de las relaciones existentes con la variable objetivo, obteniendo de este modo una base sólida para el análisis y la interpretación de los resultados.

### **5.3 Metodología**

Para llevar a cabo un análisis riguroso de la base de datos y alcanzar los objetivos propuestos, hemos planteado dos modelos alternativos que se fundamentan en técnicas de regresión, justificado por el enfoque predictivo planteado en el estudio. Estos modelos son los siguientes:

- Árboles de regresión.
- Modelo lineal normal.

Al considerar estos dos modelos alternativos, buscamos aprovechar sus fortalezas para predecir el agotamiento emocional y abordar posibles limitaciones. Los árboles de regresión nos brindan flexibilidad para capturar relaciones complejas con simplificaciones relevantes de las variables predictoras en grupos, e identificar perfiles distintos de sujetos, que reaccionan emocionalmente de forma distinta. El

modelo lineal normal persigue obtener la relación explícita entre cada una de las variables significativas para predecir la respuesta, y así generar modelos predictivos exhaustivos que permitan predecir a priori el comportamiento emocional de un sujeto en función de los atributos identificados como relevantes. La evaluación de la validez de estos modelos y la comparación de los resultados, nos permite responder con fiabilidad a las preguntas derivadas de los objetivos planteados en el estudio.

### **5.3.1 Árboles de regresión**

Un árbol de decisión es una técnica de predicción y clasificación que se basa en generar, de modo recursivo, una estructura de árbol al seccionar en trozos contiguos las variables que influyen en la respuesta, y clasificar así a los sujetos en grupos distintos. El algoritmo se detiene cuando ya no es viable seccionar ninguna variable encontrando diferencias significativas entre los grupos que genera. Las particiones que se suceden son siempre binarias en el tipo de árboles que trabajamos. Se crean pues una serie de nodos y ramas; cada nodo interno representa una característica o atributo y cada rama representa una decisión, o seccionamiento de la variable en dos categorías, basada en ese atributo. Los nodos finales contienen una clasificación de los sujetos en la base de datos, separada en función de las sucesivas ramas y nodos previos. Los árboles de decisión se clasifican en dos tipos:

- Árboles de regresión, cuando la variable respuesta es numérica, y el objetivo es predecir la respuesta a través de las variables relevantes para separar en grupos los sujetos en la base de datos; la significatividad de las diferencias entre los grupos se basa en tests de Anova.
- Árboles de clasificación, cuando la variable respuesta es categórica y el objetivo es directamente clasificar a los sujetos en grupos distintos; las pruebas de significatividad en estos casos se suelen basar en tests Chi-cuadrado.

Puesto que en nuestro estudio la variable respuesta es numérica ajustamos un árbol de regresión. Es una técnica utilizada para predecir valores numéricos en función de un conjunto de variables o características.

Nuestro objetivo predictivo a la hora de construir un árbol de regresión es el de encontrar qué variables están relacionadas con la respuesta aportando diferencias respecto del agotamiento general que manifiestan los sujetos del estudio. Aquellas variables que generan particiones en el árbol serán variables relevantes para explicar el agotamiento emocional en la población estudiada. Creamos el árbol utilizando como método de contraste para las diferencias entre grupos el test de ANOVA. Como criterio de parada, esto es, el que utilizamos para concretar la profundidad del árbol, y en consecuencia el número de nodos terminales, utilizamos la estabilización del índice de complejidad  $C_p$ .

El índice de complejidad  $C_p$  de un modelo es una medida que proporciona una estimación de la complejidad del modelo en términos de su capacidad para representar y captar patrones en los datos. Es una métrica que se utiliza para evaluar la capacidad de generalización del modelo, es decir, su capacidad para realizar predicciones precisas en datos no vistos previamente.

El objetivo al evaluar el índice de complejidad de un modelo es encontrar un equilibrio adecuado entre la capacidad del modelo para ajustarse a los datos y su capacidad para generalizar a nuevos datos. Un modelo demasiado complejo puede sobreajustar los datos de entrenamiento y tener un desempeño deficiente en datos nuevos, mientras que un modelo demasiado simple puede no captar adecuadamente los patrones y relaciones en los datos.

El índice de error  $C_p$  se calcula utilizando la suma de los cuadrados de los residuos del modelo ajustado y se ajusta por el número de parámetros del modelo. Se calcula de la manera siguiente:

1. Se construye el árbol de regresión utilizando el conjunto de datos de entrenamiento.
2. Para cada nodo terminal del árbol, se calcula la suma de los cuadrados de los residuos (SSE) en ese nodo. La suma de los cuadrados de los residuos se define como la suma de las diferencias al cuadrado entre los valores observados y los valores predichos por el modelo.
3. Se calcula el número de nodos terminales del árbol ( $k$ ).
4. Se calcula el índice  $C_p$  para el árbol de regresión utilizando la siguiente fórmula:

$$C_p = \frac{SSE + 2 d MSE}{n}$$

Donde:

- $SSE$  es la suma de los cuadrados de los residuos para el árbol.
- $d$  es el número de variables predictoras utilizadas en el árbol.
- $MSE$  es el error cuadrático medio (mean squared error), que se calcula dividiendo SSE por el número de observaciones en el conjunto de datos de entrenamiento.
- $n$  es el número de observaciones en el conjunto de datos de entrenamiento.

Se proporciona finalmente el árbol final, se identifican como relevantes para predecir el agotamiento emocional, las variables que participan en las particiones generadas, se describen los perfiles de sujetos definidos según los nodos terminales, y se proporcionan diversas medidas de bondad de ajuste que comentamos a continuación.

Para cuantificar la bondad del ajuste, utilizamos métricas clásicas en modelos de regresión, como el error cuadrático medio, definido por las distancias entre las observaciones y las predicciones:

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

Puesto que finalmente el árbol genera una partición de sujetos en grupos o nodos terminales, podemos utilizar también métricas de bondad de ajuste habituales en problemas de clasificación. Para ello consideramos una categorización previa de la variable respuesta, agotamiento emocional, en dos categorías: sujetos con agotamiento bajo y niveles de respuesta por debajo de 17, y sujetos con agotamiento medio y alto, con valor en la respuesta de al menos 17. Con esta categorización, podemos decidir si el árbol ajustado ha clasificado correctamente a los sujetos en estos dos grupos: agotamiento bajo y agotamiento medio/alto. Con ello, es viable calcular el índice de exactitud, dado por el ratio de registros correctamente clasificados con el árbol:

$$Exactitud = \frac{TP+TN}{TP+TN+FP+FN}$$

donde:

- TP representa el número de observaciones bien clasificadas con un agotamiento emocional medio alto.
- FP representa el número de observaciones mal clasificadas con un agotamiento emocional medio alto.
- TN representa el número de observaciones bien clasificadas con un agotamiento emocional bajo.
- FN representa el número de observaciones mal clasificadas con un agotamiento emocional bajo.

Estas medidas de bondad de ajuste también se evalúan sobre los resultados obtenidos aplicando el modelo lineal normal.

### **5.3.2 Modelo lineal normal**

Dada la simetría y normalidad que se aprecia en la variable agotamiento emocional, planteamos un modelo lineal normal para predecir esta variable en función del resto de variables observadas en el estudio original. Partimos para ello de un modelo que incluye como predictoras a todas las variables de la base de

datos. A partir de él procedemos con la selección de variables, que llevamos a cabo mediante un algoritmo recursivo que extrae las variables no significativas (según Anova) que menos contribuyen al modelo en términos de AIC. El algoritmo utilizado para la selección de variables se describe a continuación:

1. Inicialmente, ajustamos el modelo utilizando todas las variables regresoras disponibles. Esto nos permite capturar la posible influencia conjunta de todas las variables en la variable de respuesta.

2. A continuación, obtenemos la tabla de Anova, que nos proporciona información sobre la significancia estadística de cada variable regresora en el modelo. Identificamos aquellas variables con p-valores no significativos, es decir, aquellos valores superiores a 0.05. Estas variables se consideran candidatas a ser eliminadas del modelo.

3. Procedemos a ajustar nuevos modelos, excluyendo una variable a la vez de entre las candidatas previamente identificadas. Para cada modelo obtenido, realizamos los cálculos y análisis correspondientes.

4. Para evaluar la calidad de los modelos, calculamos el criterio de información de Akaike (AIC) para cada uno de ellos, excluyendo una variable. El AIC es una medida que penaliza la complejidad del modelo, teniendo en cuenta el ajuste a los datos. Buscamos aquel modelo que presente el mínimo valor de AIC, lo que indica una mejor relación entre ajuste y complejidad.

5. Actualizamos el modelo, incorporando la exclusión de la variable que proporcionó el mínimo valor de AIC en el paso anterior. Esto implica eliminar dicha variable del conjunto de variables regresoras.

6. Repetimos los pasos 2 a 5 iterativamente hasta que todas las variables incluidas en el modelo resulten significativas en la tabla de Anova, es decir, cuando todos los p-valores sean menores a 0.05. Esto nos permite llegar a un modelo final donde todas las variables restantes tienen una relación estadísticamente significativa con la variable de respuesta.

Este proceso iterativo de selección de variables nos ayuda a reducir la complejidad del modelo y mejorar su capacidad de generalización al eliminar variables que no aportan información significativa. Al final del proceso, obtenemos un modelo simplificado y más robusto que se ajusta de manera adecuada a los datos y permite una interpretación más concisa de las relaciones entre las variables regresoras y la variable de respuesta.

En cuanto al AIC (Akaike's Information Criterion), es un criterio de selección de modelos utilizado en el campo de la estadística y migrado al aprendizaje automático. Fue propuesto por el estadístico japonés Hirotugu Akaike y se utiliza para comparar y seleccionar entre diferentes modelos estadísticos. El AIC se basa en la idea de encontrar un equilibrio entre el ajuste del modelo a los datos y la complejidad del modelo. Proporciona una medida de la calidad relativa de un modelo al considerar tanto su capacidad para ajustarse a los datos observados como el número de parámetros que utiliza.

El cálculo del AIC se realiza utilizando la siguiente fórmula:  $AIC = -2 * \log\text{-verosimilitud} + 2 * \text{número de parámetros}$ . La log-verosimilitud es el logaritmo de la función de verosimilitud del modelo, que mide lo bien que el modelo se ajusta a los datos. El número de parámetros representa la cantidad de parámetros estimados en el modelo, y por lo tanto representa la complejidad, que juega en contraposición a la verosimilitud con signo distinto. Persigue pues este índice, proporcionar el modelo más simple posible y con máxima capacidad explicativa.

El objetivo de un criterio de selección basado en el AIC es obtener el modelo que minimiza el valor del AIC. Al comparar varios modelos utilizando el AIC, se selecciona aquel con el valor más bajo, lo que implica que es el modelo que proporciona la mejor aproximación a los datos con la menor cantidad posible de parámetros.

Finalmente presentamos el modelo ajustado, el cual representa una síntesis de las variables más relevantes para predecir el agotamiento emocional.

Valoramos la contribución de estas variables en términos de su capacidad predictiva y su impacto en la variable de interés. Para evaluar la validez y el rendimiento del modelo, utilizamos diferentes técnicas gráficas.

En primer lugar, empleamos gráficos de valores ajustados versus valores predichos, lo que nos permite visualizar la precisión del modelo al comparar las respuestas reales con las estimadas. Si los puntos en el gráfico se encuentran cercanos a una línea de referencia, indicaría que el modelo tiene una buena capacidad para predecir el agotamiento emocional.

Además, utilizamos gráficos de residuos para examinar la homogeneidad de las varianzas y la normalidad de los residuos. Si los residuos se distribuyen de manera uniforme alrededor de cero y no presentan patrones sistemáticos, podemos concluir que se cumplen los supuestos del modelo. Además, podemos verificar si los residuos siguen una distribución normal, lo que es fundamental para asegurar la validez de las inferencias estadísticas realizadas.

Por último, utilizamos gráficos *leverage* para identificar valores anómalos o atípicos que puedan influir de manera significativa en el modelo. Estos gráficos nos permiten identificar observaciones que tienen un impacto desproporcionado en los resultados del modelo, lo que puede requerir una mayor atención y consideración en el análisis.

A través de esta evaluación gráfica, buscamos asegurarnos de que el modelo ajustado sea válido, preciso y robusto en su capacidad de predicción del agotamiento emocional. Al comprender la contribución de las variables seleccionadas y al identificar posibles problemas o desviaciones en el modelo, podemos obtener resultados confiables y significativos que respalden nuestras conclusiones y recomendaciones.

## 5.4 Software y hardware

El software utilizado ha sido Rstudio (vs 2022.12.0 Build 353) como interfaz de R (vs 4.2.2) (RCoreTeam, 2023). Hemos empleado las siguientes librerías:

- tidyverse: para trabajar de manera conjunta y facilitar la manipulación, visualización y modelado de datos.
- readr: para importar datos estructurados en formato de texto plano, como archivos CSV.
- RColorBrewer: proporciona paletas de colores predefinidas para la visualización de datos.
- gridExtra: para la creación y personalización de gráficos en R. Permite combinar múltiples gráficos en una sola visualización.
- ggcorrplot: para crear gráficos de correlación visualmente atractivos y personalizables.
- dplyr: para realizar operaciones de limpieza, transformación, filtrado y resumen de datos.
- rpart: para realizar análisis de árboles de decisión mediante el algoritmo de partición recursiva.
- KableExtra: para personalizar y mejorar la presentación de tablas en R Markdown y R Notebooks.
- rattle: proporciona una interfaz gráfica de usuario para el análisis y modelado de datos.
- rpart.plot: para representar de manera visual la estructura y las divisiones realizadas por el modelo de árbol de decisión.
- sjPlot: para generar visualizaciones atractivas y descriptivas de datos en el entorno de programación R.

## 6. Resultados

A continuación, presentamos los resultados obtenidos del análisis descriptivo de manera detallada y estructurada. En primer lugar, realizamos los análisis descriptivos univariantes, que nos permiten extraer conclusiones y observaciones relevantes sobre cada una de las variables estudiadas. Mediante este enfoque, hemos podido analizar en profundidad las características y patrones de cada variable por separado.

Posteriormente, procedemos a examinar la relación entre el agotamiento emocional y las variables relacionadas con la salud, las características físicas y, por último, las características emocionales y motivacionales, en ese orden específico. A través de análisis y pruebas adecuadas, exploramos las posibles asociaciones y correlaciones entre el agotamiento emocional y estas diferentes dimensiones, lo que nos proporciona una comprensión más completa de las interacciones entre estas variables.

Continuando con nuestros análisis, planteamos diversos modelos para profundizar en nuestro entendimiento. En primer lugar, utilizamos un árbol de regresión, que nos permite clasificar a los estudiantes en función del nivel de agotamiento emocional y su caracterización según las variables que el modelo identifica como relevantes para diferenciar perfiles. Este modelo nos brinda una perspectiva discriminatoria de los diferentes perfiles de agotamiento emocional en los sujetos de nuestro estudio.

Finalmente, implementamos un modelo lineal normal, el cual nos permite llegar a un modelo más simple e identificar las variables influyentes y significativas que contribuyen a la explicación del agotamiento emocional. A través de este modelo, obtenemos información valiosa sobre cómo contribuyen a explicar el agotamiento emocional cada una de las variables incluidas en el modelo.

Finalizamos el estudio presentando medidas de bondad de ajuste en cada uno de los modelos y comparando sendos ajustes. Destacamos las ventajas y

limitaciones de cada uno de ellos y proponemos alternativas de estudio para el futuro.

Hemos aplicado pues, modelos como el árbol de regresión y el modelo lineal normal para profundizar en nuestra comprensión de los factores que influyen en el agotamiento emocional. Estos enfoques nos han proporcionado una respuesta completa y detallada de los objetivos que planteamos en nuestro estudio.

## 6.1 Análisis descriptivo

A continuación mostramos el análisis descriptivo de las diferentes variables con la finalidad de poder extraer conclusiones.

Mostramos en la Figura 1 los descriptivos las variables relacionadas con las características físicas.

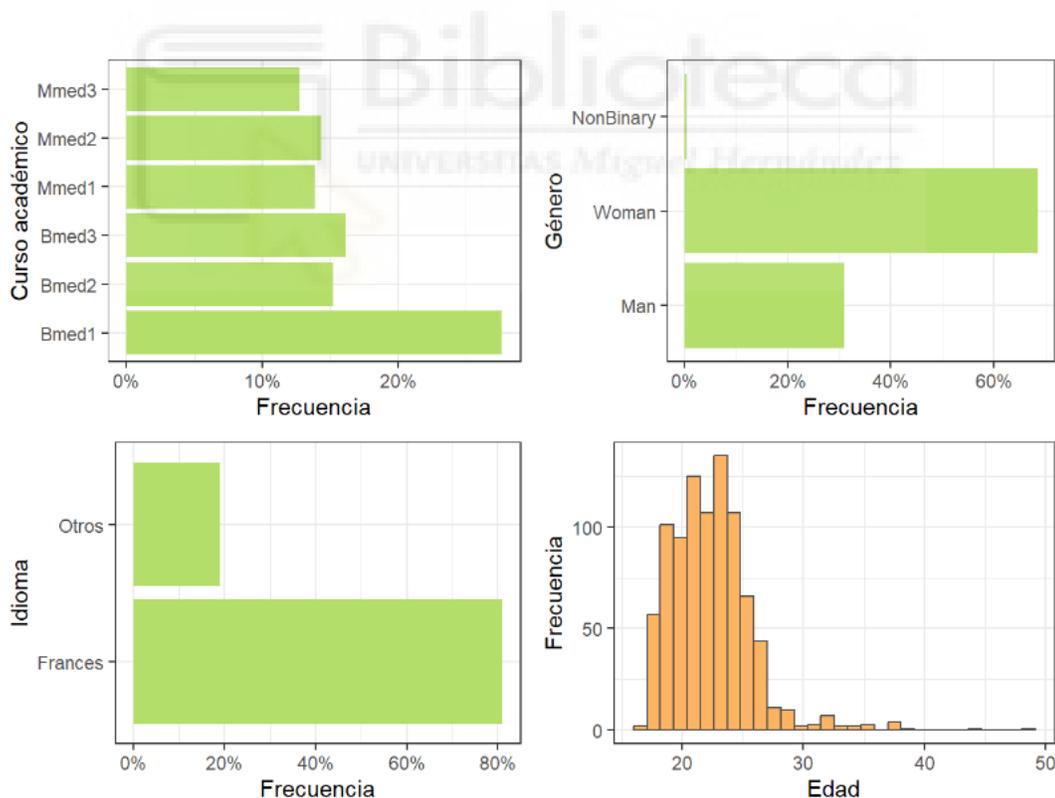


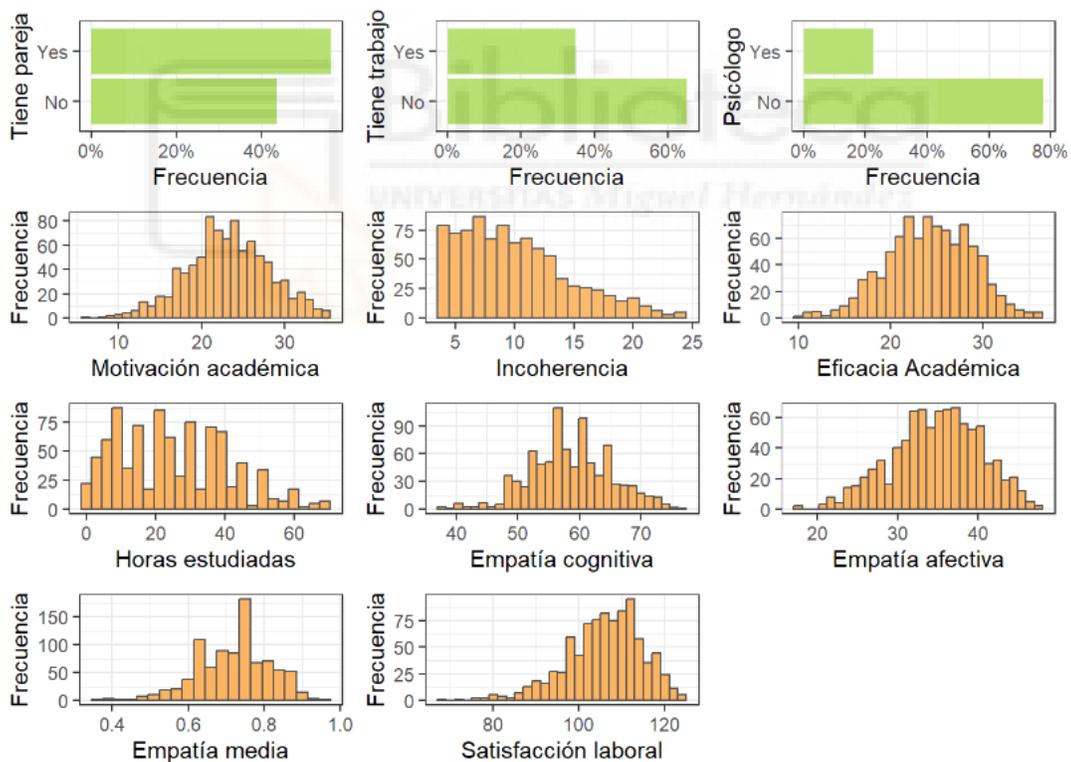
Figura 1: Descriptivos de las variables relacionadas con las características físicas.

Fuente: elaboración propia.

En base a los diagramas de barras y el histograma que se muestran en la Figura 1, destacamos los siguientes aspectos:

- El mayor porcentaje de la muestra proviene del primer curso de medicina, con un 27.65% de la muestra.
- La muestra está constituida mayoritariamente por mujeres, en torno al 68.4%.
- La gran mayoría de los estudiantes hablan francés como lengua materna, en concreto, un 80.93% del total.
- El 90% de los estudiantes tienen menos de 26 años.

Mostramos en la Figura 2 los descriptivos de las variables relacionadas con las características emocionales y motivacionales.



*Figura 2: Descriptivos de las variables relacionadas con las características emocionales y motivacionales. Fuente: elaboración propia.*

En base a los diagramas de barras y los histogramas que se muestran en la Figura 2, destacamos los siguientes aspectos:

- El 56.32% de los estudiantes del grado de medicina tienen pareja.
- El 65.12% de los estudiantes no tienen un trabajo.
- Acuden solamente el 22.46% de los estudiantes al psicólogo.
- Apreciamos cierta simetría en la variable motivación académica en torno al valor 22-23, lo que implica que prácticamente es similar el número de estudiantes motivados como desmotivados, y además se distribuyen con una campana de Gauss: hay pocos con motivaciones extremas (poca o mucha).
- Los sujetos se distribuyen de forma asimétrica hacia la derecha en la variable incoherencia y el 75% toman valores superiores a 6. Puesto que desconocemos los valores de referencia, no es posible concluir sobre si en general los estudiantes en la muestra son honestos en sus respuestas.
- Los sujetos se distribuyen de forma simétrica-campana de Gauss en la variable eficacia académica.
- El 50% de los alumnos estudian de media más de 25,29 horas semanales.
- Respecto a la empatía cognitiva se aprecia simetría respecto de la media, en torno a 58.53.
- Lo mismo ocurre respecto a la simetría en la variable empatía afectiva.
- La empatía media, al parecer en una escala estandarizada del 0 al 1, toma mayoritariamente valores por encima de 0.5 (lo cual es positivo respecto a la capacidad empática de la muestra) y tiene un rango de variación pequeño (su coeficiente de variación es de 0.13).
- Respecto a satisfacción con el trabajo, la distribución del indicador es asimétrica a la izquierda, y mayoritariamente la muestra está concentrada hacia valores altos. Puesto que no tenemos valores de referencia, no podemos ir más allá en nuestras conclusiones.

Mostramos en la Figura 3 los descriptivos de las variables relacionadas con la salud.

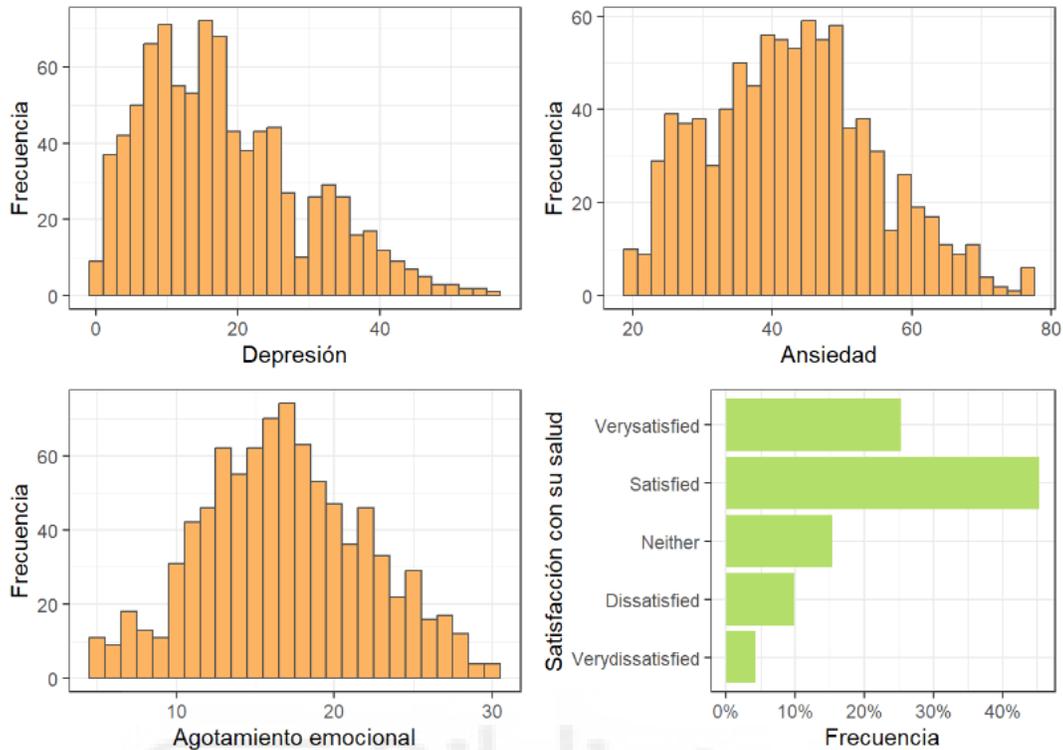


Figura 3: Descriptivos de las variables relacionadas con la salud. Fuente: elaboración propia.

En base al diagrama de barras y los histogramas que se muestran en la Figura 3, destacamos los siguientes aspectos:

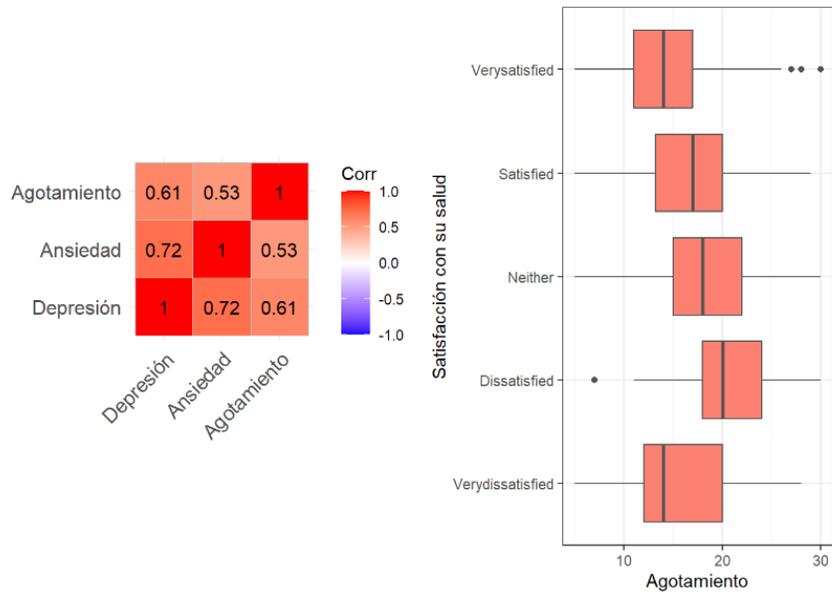
- La muestra estudiada se distribuye de forma asimétrica hacia la derecha en la variable depresión, esto es que el 36.12% de los sujetos tienen índices de depresión altos.
- El 25% de los sujetos en la muestra tienen un índice de ansiedad superior a 51.
- Acerca del agotamiento emocional, los sujetos se reparten de forma simétrica en torno a su media 16.9.
- El 70.65% de los estudiantes de medicina están satisfechos o muy satisfechos con su salud.

A continuación en la Tabla 1 mostramos los descriptivos de las variables numéricas: cuartiles, media, desviación típica y coeficiente de variación.

	Min	P0.25	Mediana	Media	P0.75	Max	Sd	Coef.var
Edad	17.00	20.00	22.00	22.38	24.00	49.00	3.30	0.15
Horas estudiadas	0.00	12.00	25.00	25.29	36.00	70.00	15.93	0.63
Satisfacción laboral	67.00	101.00	107.00	106.37	113.00	125.00	8.78	0.08
Empatía cognitiva	37.00	54.00	58.00	58.53	63.00	76.00	6.57	0.11
Empatía afectiva	18.00	31.00	35.00	34.78	39.00	48.00	5.38	0.15
Motivación académica	6.00	20.00	23.00	23.15	26.75	35.00	4.99	0.22
Empatía media	0.36	0.67	0.73	0.72	0.79	0.95	0.09	0.13
Depresión	0.00	9.00	16.00	18.05	25.00	56.00	11.48	0.64
Ansiedad	20.00	34.00	43.00	42.90	51.00	77.00	11.98	0.28
Agotamiento emocional	5.00	13.00	17.00	16.88	20.00	30.00	5.26	0.31
Incoherencia	4.00	6.00	9.00	10.08	13.00	24.00	4.59	0.46
Eficacia académica	10.00	21.00	24.00	24.21	28.00	36.00	4.63	0.19

*Tabla 1: Descriptivos de las variables numéricas. Fuente: elaboración propia.*

A continuación en la Figura 4 mostramos un gráfico de correlaciones para visualizar la relación entre el agotamiento emocional y las variables depresión, ansiedad y salud. A la derecha el diagrama de cajas nos permite diferenciar el agotamiento de sujetos en función de lo que han respondido a su percepción sobre su estado de salud.

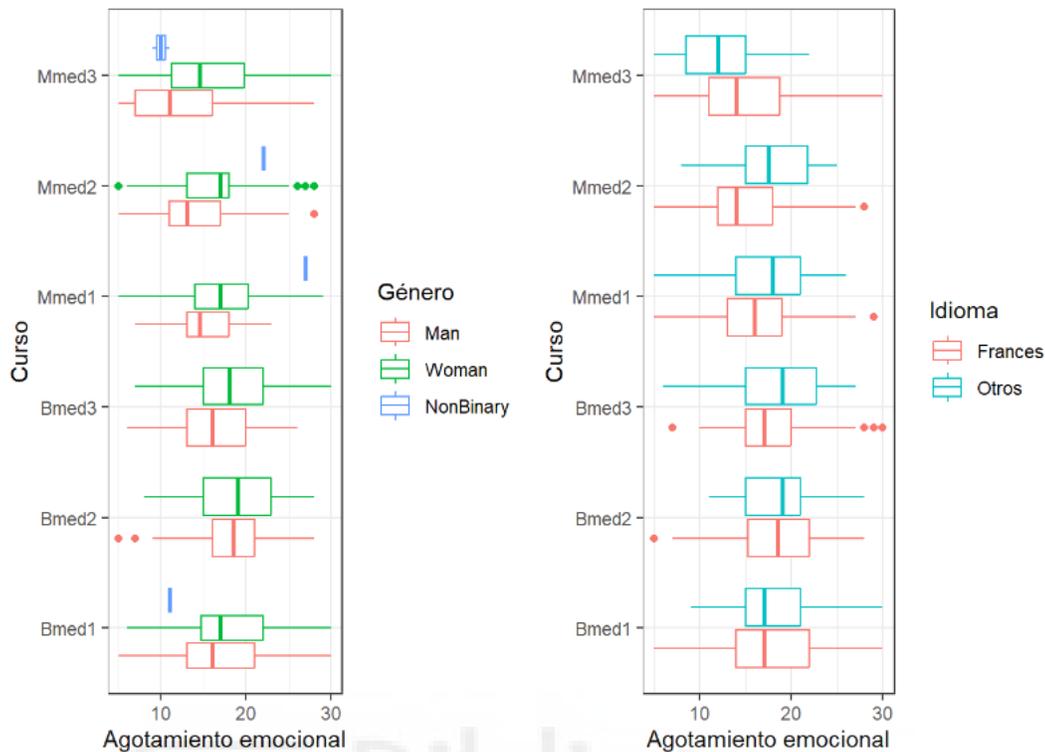


*Figura 4: Correlación y diagrama de cajas entre el agotamiento emocional, depresión, ansiedad y salud. Fuente: elaboración propia.*

En base al gráfico de correlaciones y el diagrama de cajas que se muestran en la Figura 4, destacamos los siguientes aspectos:

- Las variables depresión y ansiedad presentan correlaciones altas con el agotamiento emocional, entre 0.53 y 0.61 además ansiedad y depresión están muy correlacionadas, con una correlación de 0.72.
- En general se aprecia una relación indirecta entre el grado de satisfacción con su salud y el agotamiento emocional (a mayor agotamiento emocional menor satisfacción), excepto en el grupo de los que se manifiestan muy insatisfechos, donde hay mucha dispersión.

A continuación en la Figura 5 mostramos una serie de diagramas de cajas para mostrar la relación que existe entre el agotamiento emocional y las variables relacionadas con las características físicas.



*Figura 5: Relación entre el agotamiento emocional y las variables relacionadas con las características físicas. Fuente: elaboración propia.*

En base los diagramas de cajas que se muestran en la Figura 5, destacamos los siguientes aspectos:

- En general parece que el agotamiento emocional crece hasta segundo-tercer curso, y luego decrece.
- Independientemente del curso, las mujeres (en verde) están más agotadas emocionalmente que los hombres (en rojo).
- A medida que avanza el curso académico los alumnos que hablan otro idioma diferente al francés (azul) se distancian más respecto del agotamiento emocional de los de habla francesa (rojo), y acusan más el agotamiento, excepto en el último curso del grado.

A continuación en la Figura 6 mostramos los diagramas de dispersión con los que pretendemos observar qué relación existe entre el agotamiento emocional y las variables relacionadas con las características emocionales y motivacionales.

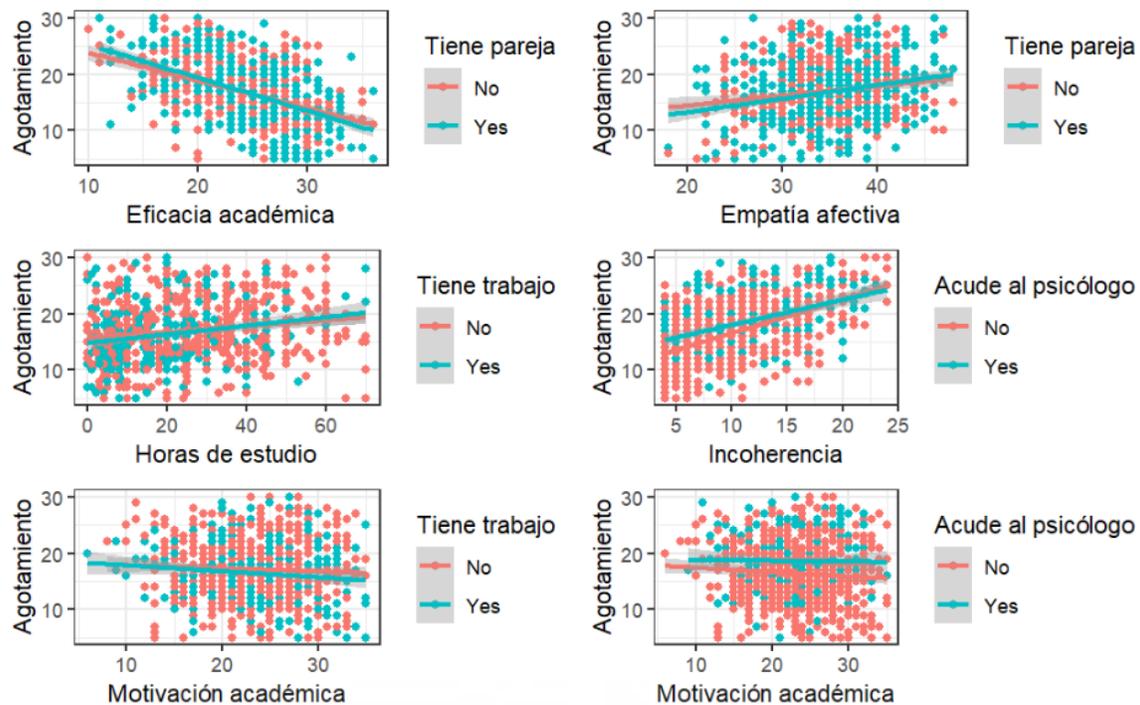


Figura 6: Relación entre el agotamiento emocional y las variables relacionadas con las características emocionales y motivacionales. Fuente: elaboración propia.

En base a las nubes de puntos que se muestran en la Figura 6, destacamos los siguientes aspectos:

- Aquellos estudiantes que están menos agotados emocionalmente presentan mayor eficacia académica, ya que la tendencia es negativa (a mayor eficacia académica menor agotamiento emocional) ; además, el hecho de tener pareja no parece influir en un cambio de tendencia.
- Los alumnos que tienen mayor empatía afectiva también tienen mayor agotamiento emocional. De nuevo el hecho de tener pareja no afecta a esta relación.
- Los estudiantes que están más agotados emocionalmente son aquellos que estudian más horas. No se aprecian diferencias entre los estudiantes que trabajan y los que no.
- El grado de incoherencia de los estudiantes está relacionado directamente con su agotamiento emocional. Esto nos puede llevar a pensar que suelen

mentir más sobre su estado de salud, aquellos que están más agotados emocionalmente. De nuevo no se observa variación respecto al hecho de tener trabajo o no.

- La motivación académica no parece estar relacionada con el agotamiento emocional de los estudiantes, si bien el agotamiento medio de los que no van al psicólogo se aprecia menor al de los que sí lo hacen.

Concluyendo, a la vista del análisis descriptivo podemos derivar que parecen estar relacionadas con el agotamiento emocional, las siguientes variables:

Del grupo de las relacionadas con la salud: depresión, ansiedad y satisfacción con la salud.

Del grupo de las relacionadas con las características físicas: curso académico, género e idioma materno.

Del grupo de las relacionadas con las características emocionales y motivacionales: eficacia académica, empatía afectiva, horas de estudio, grado de incoherencia y psicólogo.

## **6.2 Árbol de regresión**

Abordamos ahora el modelo basado en el árbol de regresión. Tras ajustar el árbol de regresión a los datos, incluyendo todos los predictores, estudiamos el índice de complejidad  $C_p$  para decidir dónde podarlo y simplificar así el modelo. Como observamos en la Figura 7, el índice  $C_p$  se estabiliza a partir de un modelo con 4 nodos terminales. Por ello, optamos por truncar el árbol para garantizar 4 nodos terminales.

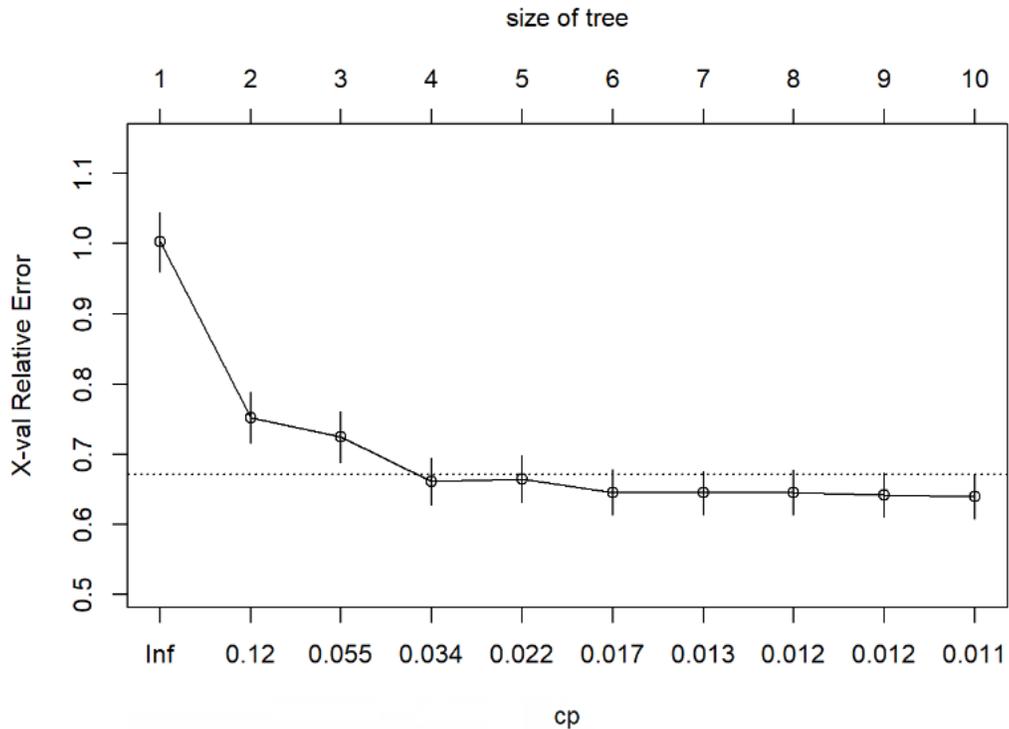


Figura 7: Gráfico del índice CP obtenido sobre un árbol de regresión con todas las variables predictoras. Fuente: elaboración propia.

La Figura 8 muestra el árbol de decisión ajustado con 4 nodos terminales. Extraemos de dicho árbol que las únicas variables relacionadas con el agotamiento emocional, y que por consiguiente diferencian sujetos con distintos niveles de agotamiento, son 'depresión' e 'incoherencia', ambas incluidas entre las variables del grupo salud.

Los perfiles de sujetos que se derivan son los siguientes:

- perfil 1 (nodo4): sujetos (un total de n=88 registros) con índices de depresión muy bajos ( $cesd < 4.5$ ); identificarían sujetos nada deprimidos, que muestran un agotamiento mínimo (media estimada de 11);
- perfil 2 (nodo5): sujetos (n=367) con depresión medio-baja ( $cesd$  superior a 4.5 e inferior a 17), que muestran un agotamiento medio-bajo (media estimada de 15);

- perfil 3 (nodo6): sujetos (n=193) con depresión alta (cesd superior a 17) y creíbles (índice de incoherencia bajo, inferior a 11), que muestran un agotamiento medio-alto (media estimada de 18);
- perfil 4 (nodo7): sujetos (n=238) con depresión alta (cesd superior a 17) y poco creíbles (índice de incoherencia alto, superior a 11), que muestran un agotamiento emocional alto (media estimada de 21).

Concluimos pues que el agotamiento emocional está directamente relacionado con la depresión psicológica y también se corresponde, para los sujetos con mayor agotamiento, con aquellos de credibilidad baja, que posiblemente traten de disimular su estado de salud, con tendencia a manifestar autopercepciones positivas, contrarias a la realidad.

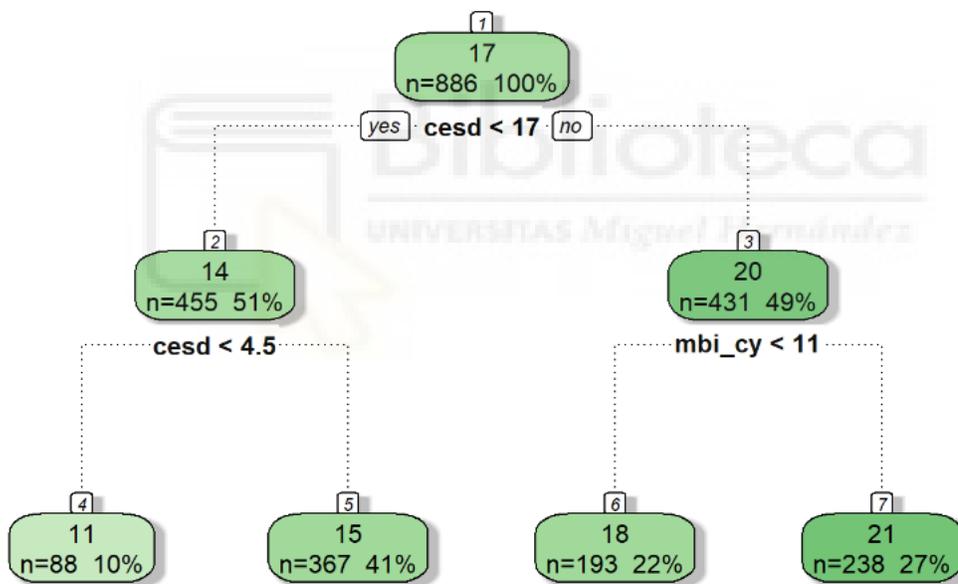


Figura 8: Resultado del árbol de regresión. Fuente: elaboración propia.

En cuanto a las medidas de bondad del ajuste, concluir que obtenemos un error cuadrático medio de predicción de 17,27 y que el 72,12% de la muestra se ha clasificado correctamente, esto es, sujetos con agotamiento bajo (< 17) y con agotamiento alto (≥ 17) han sido correctamente clasificados en el árbol.

### 6.3 Modelo lineal

Tras realizar la selección de variables partiendo del modelo global, llegamos al modelo ajustado, que visualizamos en la Tabla 2 a través de los coeficientes estimados, errores estándar y p-valores correspondientes.

Predictores	Coefficiente estimado	Error estándar	Pvalor
Intercept	10.38	1.53	<0.001
Curso 2	1.58	0.39	<0.001
Curso 3	1.20	0.40	0.003
Curso 4	-0.05	0.43	0.911
Curso 5	-1.33	0.43	0.002
Curso 6	-1.56	0.47	0.001
Género Mujer	0.49	0.29	0.092
Género No binario	-0.13	1.65	0.938
Horas estudiadas	0.03	0.01	0.001
Insatisfecho con su salud	0.98	0.73	0.180
Ni satisfecho ni insatisfecho	0.32	0.68	0.642
Satisfecho con su salud	0.45	0.62	0.468
Muy satisfecho con su salud	-0.48	0.65	0.459
Acude al psicólogo	-0.22	0.31	0.468
Empatía afectiva	0.04	0.03	0.094
Depresión	0.12	0.02	<0.001
Ansiedad	0.04	0.02	0.012
Incoherencia	0.34	0.03	<0.001
Eficacia académica	-0.15	0.03	<0.001
Observaciones	886		
R <sup>2</sup> / R <sup>2</sup> ajustado	0.539 / 0.529		
AIC	4.808.428		

*Tabla 2: Coeficientes estimados en el modelo lineal ajustado, junto con sus errores estándar y p-valores asociados. Fuente: elaboración propia.*

Las variables que influyen en el agotamiento emocional según el modelo ajustado son las siguientes:

Variables relacionadas con la salud: depresión, ansiedad y satisfacción con la salud.

Variables relacionadas con las características físicas: curso académico y género.

Variables relacionadas con las características emocionales y motivacionales: horas de estudio, psicólogo, empatía afectiva, grado de incoherencia y eficacia académica.

Respecto a los predictores numéricos, observamos que todos ellos, salvo la eficacia académica, están relacionados directamente con el agotamiento emocional (coeficientes positivos). Los estudiantes que están más agotados emocionalmente tienen un rendimiento académico inferior.

Respecto a los predictores categóricos, resaltamos las siguientes conclusiones:

En cuanto a las variables relacionadas con la salud:

- salud: el nivel de agotamiento varía según el grado de satisfacción con la salud de los alumnos. Aquellos que están insatisfechos experimentan un mayor agotamiento, con un incremento de 0.98 puntos en comparación con los que se sienten muy insatisfechos. Por otro lado, los estudiantes que se encuentran en un estado neutro, ni satisfechos ni insatisfechos, experimentan una disminución leve del agotamiento en 0.66 puntos. En contraste, los alumnos que están satisfechos con su salud experimentan un ligero aumento en el agotamiento, con un incremento de 0.13 puntos. Por último, aquellos que se sienten muy satisfechos con su salud presentan una reducción en el agotamiento emocional de 0.48 puntos en comparación con aquellos que están muy insatisfechos.

En cuanto a las variables relacionadas con las características físicas:

- curso académico: el nivel de agotamiento sube en el segundo curso respecto del primero (1.58 puntos más), el tercero baja algo pero continúa por encima del primer curso (0.38 puntos), sigue bajando durante los cursos siguientes a niveles inferiores al del primer curso, y ya en el último curso, el nivel de agotamiento es 1.56 puntos inferior al del primer curso.

- género: el nivel de agotamiento emocional es mayor en las mujeres que en los hombres (0.49 puntos) y solo algo superior para el género no binario (0.13 puntos respecto de los hombres).

En cuanto a las variables relacionadas las características emocionales y motivacionales:

- psicólogo: el nivel de agotamiento emocional de los estudiantes que acuden al psicólogo es 0.22 puntos inferior al de los que no van. Parece tener un efecto positivo el hecho de acudir a un psicólogo.

Por último, en cuanto a la bondad del ajuste, observamos que el modelo ajustado presenta un coeficiente de determinación del 53%, que no es especialmente bueno, obtenemos un error cuadrático medio de predicción de 12,73% y que el 75,4% de la muestra se ha clasificado correctamente.

A continuación procedemos con la validación del modelo ajustado, cuyos gráficos se muestran en la Figura 9.

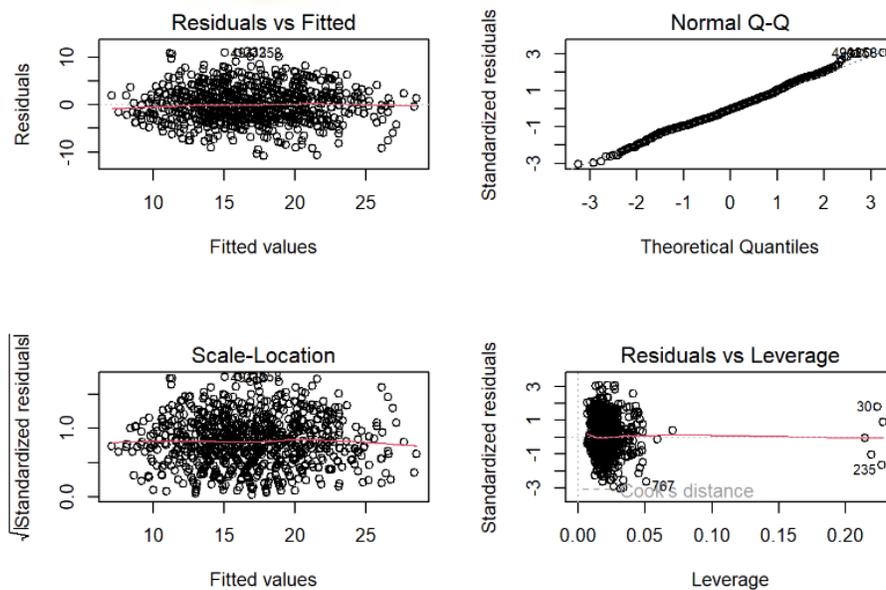


Figura 9: Gráficos de validación. Fuente: elaboración propia.

Observamos que los residuos estandarizados se encuentran en torno a la media, presentan normalidad y no aparecen valores atípicos relevantes u observaciones excesivamente influyentes. Damos pues por validado el modelo ajustado.

## **Conclusiones**

En el estudio pionero de Carrard et al. (2022) se utilizó esta base de datos, con información relativa a las características físicas y psicológicas de un grupo de estudiantes de Medicina en Suiza, para explorar la relación entre los problemas de salud mental de los estudiantes y el tipo de empatía que manifestaban. Se descubrió que factores como el curso académico y el género influyen en los problemas de salud mental, el agotamiento y los niveles de empatía. Se observó que las mujeres tienen una mayor prevalencia de problemas de salud mental y agotamiento, así como niveles más altos de empatía en comparación con los hombres. Nuestros hallazgos respaldan estas conclusiones previas y aportan nuevas perspectivas.

El propósito principal de nuestro estudio es examinar detalladamente cómo las variables recopiladas en la muestra de alumnos de medicina afectan al agotamiento emocional. Para ello realizamos una agrupación temática de las variables presentes en la base de datos, con el fin de indagar, de forma diferenciada, las relaciones del agotamiento emocional con características relativas a salud, aspectos físicos, y aspectos emocionales y motivacionales. A partir de esta diferenciación de variables planteamos los objetivos relativos a estudiar su relación con el agotamiento, y proponemos dos modelos: un árbol de regresión y un modelo lineal normal.

Mediante estos modelos, identificamos las variables que desempeñan un papel importante para explicar el agotamiento emocional de los estudiantes.

A través del árbol de regresión, hemos identificado que la variable depresión está asociada con el agotamiento emocional en el grupo de la salud. Además, hemos

encontrado que la variable incoherencia, incluida entre las características emocionales y motivacionales, también influye en el agotamiento emocional de los estudiantes.

A través del modelo lineal ajustado, hemos identificado diferentes variables que muestran una relación significativa con el agotamiento emocional. Clasificamos estas variables en tres grupos:

- Variables relacionadas con la salud: incluyen la depresión, la ansiedad y la satisfacción autopercibida con su salud.
- Variables relacionadas con las características físicas: el curso académico y el género de los estudiantes.
- Variables relacionadas con las características emocionales y motivacionales: las horas de estudio dedicadas, la asistencia a un psicólogo, la empatía afectiva, el grado de incoherencia y la eficacia académica.

Con el fin de realizar una comparación objetiva de los modelos, empleamos dos medidas de bondad de ajuste que nos permiten determinar lo bien que se ajustan cada uno de los modelos a los datos, y dan pues indicios de su capacidad predictiva para predecir nuevos datos. Los resultados del análisis comparativo se muestran en la Tabla 3 a continuación, que calculan el error cuadrático medio MSE de cada modelo (definido en la Sección 5.3) .

*Tabla 3: Medidas de bondad de ajuste. Fuente: elaboración propia.*

<b>Modelo \ Medida</b>	<b>MSE</b>	<b>Exactitud</b>
<b>Árbol de regresión</b>	17,27	72,12
<b>Modelo lineal normal</b>	12,73	75,4

Podemos observar pues, en la Tabla 3, que el modelo lineal muestra un nivel de ajuste superior al árbol de regresión, ya que presenta un error de predicción MSE

más bajo, específicamente de 12,73. Además, el modelo lineal demuestra una mayor precisión al clasificar correctamente a los individuos con tasa baja de agotamiento (menor a 17) y tasa media-alta (mayor o igual a 17), con una exactitud del 75,4%. Estos resultados indican que el modelo lineal es más efectivo para predecir y clasificar el agotamiento emocional en la muestra estudiada de estudiantes de Medicina.

Este enfoque nos permite realizar un estudio detallado de las relaciones específicas entre el agotamiento emocional y las variables relacionadas con la salud, características físicas, emocionales y motivacionales de los estudiantes.

Por otra parte, mediante el uso del modelo lineal ajustado, se ha determinado que todos los predictores numéricos analizados se pueden agrupar en dos temáticas principales: salud y características emocionales y motivacionales. En el grupo relacionado con la salud, encontramos la depresión y la ansiedad, mientras que en el grupo de características emocionales y motivacionales se incluyen las horas de estudio, la empatía afectiva, la incoherencia y la eficacia académica.

Es importante destacar que todas estas variables, excepto la eficacia académica, tienen una relación directa con el agotamiento emocional. Esto significa que los estudiantes que experimentan niveles más altos de agotamiento emocional tienden a presentar una mayor presencia de depresión, ansiedad, un mayor número de horas de estudio, una menor empatía afectiva y una mayor sensación de incoherencia. Además, se observa que aquellos estudiantes que experimentan un mayor agotamiento emocional tienden a tener un rendimiento académico inferior.

Respecto a los predictores categóricos, en cuanto a las variables del grupo de la salud extraemos lo siguiente:

El nivel de agotamiento emocional varía en relación con la satisfacción que los alumnos tienen respecto a su salud. Aquellos que están insatisfechos experimentan un mayor agotamiento, mientras que los estudiantes en un estado

neutro muestran una disminución leve. Por otro lado, los alumnos satisfechos con su salud experimentan un ligero aumento en el agotamiento.

En cuanto a las variables relacionadas con las características físicas extraemos lo siguiente:

El nivel de agotamiento emocional varía según el curso académico. En general, se observa un aumento en el agotamiento durante el segundo curso en comparación con el primero. A partir del tercer curso, el agotamiento disminuye gradualmente, alcanzando niveles inferiores al del primer curso en los cursos siguientes. Por otro lado, se nota que el nivel de agotamiento emocional es mayor en las mujeres en comparación con los hombres, y existe una ligera diferencia adicional en el género no binario en comparación con los hombres.

En cuanto a las variables relacionadas con las características emocionales y motivacionales extraemos lo siguiente:

El hecho de acudir a un psicólogo parece tener un efecto positivo en el nivel de agotamiento emocional de los estudiantes. Aquellos que reciben ayuda profesional presentan un nivel de agotamiento emocional inferior en comparación con aquellos que no buscan apoyo psicológico.

En términos de relación, se puede concluir que el agotamiento emocional tiene un impacto negativo tanto en la salud mental (como podemos apreciar a través del árbol de regresión, con la relación entre agotamiento emocional y depresión) como en el rendimiento académico (como podemos apreciar a través del modelo lineal ajustado, con la relación entre agotamiento emocional y la eficacia académica). Por lo que a partir de ambos modelos podemos extraer la conclusión de que el agotamiento emocional puede tener repercusiones negativas en diferentes aspectos de la vida de los estudiantes.

En definitiva, el proceso de investigación y análisis desarrollado ha ampliado nuestra comprensión sobre los factores que influyen en el agotamiento emocional de los estudiantes. Estos conocimientos son fundamentales para desarrollar estrategias de apoyo y promover entornos educativos saludables. Al considerar la

salud mental, la satisfacción con la salud, el entorno académico y las características individuales, podremos generar intervenciones más efectivas que fomenten el bienestar integral de los estudiantes.

## Referencias

Valerie Carrard, Céline Bourquin, Sylvie Berney, Katja Schlegel, Jacques Gaume, Pierre-Alexandre Bart, Martin Preisig, Marianne Schmid Mast & Alexandre Berney (2022): The relationship between medical students' empathy, mental health, and burnout: A cross-sectional study, *Medical Teacher*, DOI: 10.1080/0142159X.2022.2098708

EFSA (European Food Safety Authority) 2014. Guidance on Statistical Reporting. *EFSA Journal* 2014 12 (12): 3908. doi: 10.2903/j.efsa.2014.3908 <http://www.efsa.europa.eu/en/efsajournal/pub/3908>

Kaggle. (2022). Medical Student Mental Health. Kaggle. Retrieved March 21, 2023, from <https://www.kaggle.com/datasets/thedevastator/medical-student-mental-health>

R Core Team (2023). *\_R: A Language and Environment for Statistical Computing\_*. R Foundation for Statistical Computing, Vienna, Austria. <<https://www.R-project.org/>>.

Posit team (2023). *RStudio: Integrated Development Environment for R*. Posit Software, PBC, Boston, MA. URL <http://www.posit.co/>.

Peña Sánchez de Rivera, D. (1993, Julio 20). *Estadística. modelos y metodos 2. modelos lineales y series temporales* : Peña Sanchez De Rivera, Daniel. Amazon. Retrieved June 9, 2023, from <https://www.amazon.es/Estadistica-modelos-metodos-lineales-temporales/dp/8420681105>

## **Anexos**

El código utilizado para llevar a cabo este estudio se encuentra disponible públicamente en este [documento](#).

