

INVESTIGACIÓN SOBRE LAS VARIABLES QUE AFECTAN LA CANTIDAD DE HORAS DE SUEÑO DE UN ESTUDIANTE UNIVERSITARIO

Jose Carlos Salinas, Universidad del Pacífico, jc.salinasma@alum.up.edu.pe

César Cabezas, Universidad del Pacífico, cm.cabezag@alum.up.edu.pe

Franz Figueroa, Universidad del Pacífico, f.figueroag@alum.up.edu.pe

Gonzalo Alvis, Universidad del Pacífico, gr.alvisb@alum.up.edu.pe

Alejandro Ramirez, Universidad del Pacífico, ae.ramirez@alum.up.edu.pe

Resumen

El sueño desempeña un papel fundamental en nuestra rutina diaria y en nuestra salud general. Estudios han demostrado que la calidad del sueño está relacionada con la calidad de vida y la salud en general. La falta de sueño adecuado puede tener efectos negativos en el sistema inmunológico, el rendimiento académico, la salud mental, las relaciones interpersonales y la productividad laboral. Es importante tener entre 7 y 9 horas de sueño de calidad para mantener un equilibrio adecuado. Sin embargo, muchos estudiantes universitarios presentan hábitos de sueño deficientes, lo cual puede tener repercusiones negativas en su rendimiento académico y su salud en general. En el caso de los estudiantes peruanos, se ha observado una falta de sueño de calidad, lo que está relacionado con niveles altos de ansiedad y estrés. Además, el mal uso de dispositivos electrónicos también afecta negativamente la calidad del sueño en los estudiantes. Por tanto, es necesario investigar los factores que influyen en las horas de sueño de los estudiantes de la Universidad del Pacífico. Esta investigación tiene como objetivo analizar los factores que afectan la duración del sueño y crear modelos predictivos que ayuden a los estudiantes a mantener un sueño adecuado, lo cual contribuirá a su bienestar y rendimiento académico.

Palabras Clave: Sueño, Regresión lineal, Regresión logística, training, testing

INTRODUCCIÓN

El sueño desempeña un papel fundamental en el desarrollo de nuestra rutina diaria y es esencial para mantener una vida saludable, ya que está estrechamente relacionado con nuestra salud en general. Durante el periodo de sueño, nuestro cuerpo realiza procesos vitales de reparación y regeneración que fortalecen nuestro sistema inmunológico, el cual desempeña un papel clave en la prevención de enfermedades y en la promoción de la resistencia a las infecciones. En un estudio llevado a cabo por Chen et al. (2014), publicado en el Quality of Life Research, se comenta sobre la importancia del sueño, estableciendo una relación entre la calidad de sueño y la calidad de vida relacionada a la salud (HRQOL por sus siglas en inglés). Este estudio recae en la importancia del sueño para el correcto funcionamiento del sistema inmune, desempeño escolar o universitario, salud mental entre otros HRQOL.

Sin embargo, no cualquier tipo de sueño es suficiente. Como mencionan Chen et al. (2014) y Wang & Biro (2020), existen diferentes estándares de calidad del sueño que se utilizan en los estudios para medir esta actividad de manera adecuada. Según un estudio realizado por Hirshkowitz et al. (2015), se considera que el rango de horas adecuado para un buen sueño es de 7 a 9 horas en promedio. La calidad del sueño puede verse afectada por diversos factores,

como problemas mentales y el estilo de vida. Wang & Biro (2020) mencionan que uno de estos factores es el nivel de ejercicio físico que una persona realiza, lo cual puede tener tanto efectos positivos como negativos en la calidad del sueño percibida por estudiantes universitarios.

Como se ha mencionado anteriormente, el sueño tiene un impacto en los procesos del sistema inmunológico del cuerpo humano, los cuales son esenciales para el bienestar general. Además, existe una estrecha relación entre el sueño y el desempeño en actividades físicas como el deporte, tal como señalan Wang & Biro (2020). Durante el sueño, se liberan hormonas de crecimiento que ayudan a reparar y hacer crecer los tejidos musculares. Asimismo, un sueño adecuado mejora la coordinación, la velocidad de reacción y la precisión, lo cual puede marcar la diferencia en el rendimiento deportivo, con rendimiento deportivo el autor se refiere al nivel de desempeño que un deportista logra en su disciplina. Un estudio llevado a cabo en Corea por Jin Ban & Jin Lee (2001), con un grupo mayoritariamente conformado por estudiantes universitarios, encontró que aquellos con una calidad de sueño deficiente mostraban una menor velocidad de reacción, lo cual es especialmente peligroso en actividades como la conducción. Este mismo estudio también evidenció una relación entre malos hábitos de sueño y una alta mortalidad.

La relación entre el sueño y la salud mental es especialmente importante. La falta de sueño puede afectar negativamente a la salud mental, contribuyendo a la aparición de trastornos del estado de ánimo como la depresión y la ansiedad. Las personas que sufren de insomnio crónico tienen un mayor riesgo de desarrollar problemas de salud mental. Dormir lo suficiente puede marcar una gran diferencia en la salud mental y emocional. En la revisión literaria llevada a cabo por Wang & Biró (2020), se señala que los problemas relacionados a la salud mental como la ansiedad, depresión o TDAH además de tener un impacto negativo en la calidad de sueño, son un resultado de este. El estudio de Jin Ban & Jin Lee (2001) comenta también que los estudiantes con peor calidad de sueño tienden a presentar depresión en algún grado. Esto se complementa con el estudio realizado por Joo et. al. (2022). En dicho estudio, la calidad de sueño se mide según el Pittsburgh Sleep Quality Index, en el que un puntaje menor a 5 equivale a buena calidad y mayor de 5 equivale a una peor calidad de sueño. El 72.3% de los encuestados presentaban un puntaje mayor a 5, y de ellos, el 15% reportaban síntomas de depresión, porcentaje mucho mayor comparado con los que decían tener buena calidad de sueño, puntaje menor a 5, y presentaban síntomas de depresión, que eran solo el 1.5%.

La falta crónica de sueño también puede afectar nuestra vida social y nuestras relaciones interpersonales. Cuando no se duerme lo suficiente, es decir, en un rango fuera de las 7 a 9 horas presentadas en el estudio de Hirshkowitz et al. (2015), es más probable que nos sintamos irritables, impacientes y con una menor capacidad para manejar las emociones. Esto puede tener un impacto negativo en nuestras relaciones personales. De acuerdo a Jin Ban & Jin Lee (2001), un sueño interrumpido o con deficiencias tiene un efecto en la calidad de las interacciones personales de los implicados, describiéndolo como un entorpecimiento y

disfrute reducido de las relaciones interpersonales.

La falta de sueño puede afectar negativamente nuestra capacidad para concentrarnos, afectar nuestra memoria y disminuir nuestra productividad en el trabajo o en los estudios como lo demuestra Chen et al (2014). Jin Ban & Jin Lee (2001) también realizan una observación respecto al nivel de concentración que una persona con pobre nivel de sueño presenta. Sobre el tema, Chiang et al. (2020) realiza un trabajo a mayor profundidad en una población de estudiantes en USA. Chiang concluye que, si los administradores académicos conocieran el impacto del sueño en el desenvolvimiento académico, harían un mayor énfasis en los horarios de sueño sobre otras cosas. Para el desarrollo del trabajo, nos enfocamos en un descubrimiento revelado por el estudio de Chen et al. (2014) y respaldado por otros documentos académicos como el de Desjardins & Grandbois (2022): los estudiantes universitarios presentan hábitos de sueño deficientes. Lo afirmado anteriormente lo podemos ver en un escenario latinoamericano, ya que en un estudio realizado por Monterrosa et al (2014) sobre los estudiantes universitarios colombianos de la Universidad de Cartagena de primero a décimo semestre, la duración de sueño promedio fue $4,6 \pm 1,3$ horas lo cual no cumple con los estándares mencionados anteriormente. Además, el autor encontró que el 88.1% de los estudiantes fueron identificados como malos dormilones, es decir, que sus horas de sueño se encontraban fuera del rango de 7 a 9 horas.

Tomando como referencia los datos de los estudios previamente presentados, unas horas de sueño aceptables para un alumno universitario deben ser alrededor de 7 u 9 horas. Dichos estudios dejan en claro que las horas de sueño afectan tanto en el desempeño académico como la salud mental, sin embargo, gran parte de ellos se encuentran enfocados en universidades extranjeras donde los alumnos acostumbran vivir en el mismo campus, el cual no es el caso en la mayoría de universitarios peruanos. En el caso de los universitarios peruanos, investigaciones como la realizada en estudiantes de medicina (Rosales et al., 2007) han demostrado que la falta de sueño de calidad está directamente relacionada con altos niveles de ansiedad y estrés, afectando la salud mental. Además, el mal uso de dispositivos electrónicos, como la adicción a Internet debido al uso excesivo de teléfonos móviles, también influye negativamente en la calidad del sueño y en la salud mental de los estudiantes (Vásquez-Chacon et al., 2019). En la Universidad Católica Santo Toribio de Mogrovejo de Lambayeque, se encontró que el 55,8% de los estudiantes dormía menos de 5 horas, mientras que en la Universidad Peruana Cayetano Heredia, el 83,9% de los estudiantes de medicina dormía por debajo del rango recomendado de 7 a 9 horas (Izaguirre Corcuera et al., 2019). Ante esta situación, resulta importante investigar si la cantidad de horas de sueño de los alumnos de la Universidad del Pacífico es adecuada, con el objetivo de comprender y abordar esta problemática en el contexto específico de la institución.

La presente investigación es de gran relevancia debido a la notable escasez de estudios que aborden específicamente la duración del sueño. En su mayoría, las investigaciones existentes se enfocan en la calidad del sueño en general o no se centran en estudiantes universitarios o dejan de lado el análisis estadístico y matemático de los patrones de sueño y la cantidad de

horas de sueño en estudiantes universitarios peruanos. Esta carencia de investigación evidencia una falta de conocimiento científico sólido y riguroso en relación con la duración del sueño en estudiantes universitarios peruanos, lo que constituye una laguna significativa en la literatura científica. Con esta investigación buscamos llenar este vacío de conocimiento y aportar nuevos hallazgos relevantes. Se busca abordar esta temática desde una perspectiva estadística y matemática, con el propósito de proporcionar datos concretos y análisis rigurosos. Además, los resultados de este estudio se pueden utilizar en futuros estudios sobre el rendimiento académico de los alumnos de la Universidad del Pacífico. Es importante mencionar que existen estudios previos que respaldan indirectamente la relación entre la cantidad de horas de sueño y el rendimiento académico, como el trabajo de Izaguirre Corcuera et al. (2019). Asimismo, se considera la variable "duración del sueño" planteada por Zheng et al. (2023) en su investigación sobre la calidad del sueño. Aunque en este estudio no se tiene planeado analizar directamente el rendimiento académico, se espera que las conclusiones obtenidas puedan sentar las bases para futuras investigaciones que exploren la relación entre el rendimiento académico y las horas de sueño.

Se entiende que existen varias variables que pueden tener un impacto significativo en la cantidad de horas de sueño de los estudiantes. Por lo tanto, este trabajo tiene como objetivo principal determinar los factores que influyen significativamente en las horas de sueño de los estudiantes de la Universidad del Pacífico. Se busca identificar y analizar estos factores en profundidad con el fin de comprender su relación con la duración del sueño. Asimismo, se pretende utilizar esta información para poder definir si un estudiante tendrá la cantidad adecuada de horas de sueño, de acuerdo a lo mencionado por Hirshkowitz et al. (2015), mediante la técnica de regresión lineal. En otras palabras, se busca establecer si existen características o comportamientos específicos que pueden predecir si un estudiante está en riesgo de no dormir lo suficiente, es decir, entre 7 y 9 horas. Por otro lado, uno de los objetivos es la creación de un modelo que permita determinar si un estudiante está durmiendo las horas adecuadas mediante una regresión logística. Esto será de gran utilidad para impulsar la aplicación de estrategias preventivas que guíen a los estudiantes de la Universidad del Pacífico a cumplir con las horas de sueño recomendadas. Esto mediante charlas en conjunto con el área de Bienestar y Formación Estudiantil de la Universidad del Pacífico (BFE). Adicionalmente, otro objetivo es la detección de patrones existentes entre las características encontradas. Esto implica examinar si hay correlaciones o tendencias entre los factores identificados y la cantidad de horas de sueño. Esto con el propósito de comprender mejor cómo interactúan estos factores y cómo pueden afectar la cantidad de sueño de los estudiantes.

Asimismo, se obtendrán estadísticas descriptivas sobre los datos demográficos de la muestra que se va a analizar, como la media, la mediana y la varianza de las horas de sueño por cada facultad de la Universidad del Pacífico. Esto permitirá tener una comprensión más completa de la muestra estudiantil y su relación con las horas de sueño. Por tanto, se espera obtener modelos que nos permitan que, con un R cuadrado mayor a 0.5, se pueda predecir el número

de horas de sueño diarias de un estudiante dependiendo de los factores estudiados en el trabajo. También se espera encontrar las variables más influyentes y las menos influyentes en la cantidad de horas dormidas de los estudiantes. Los resultados que se obtengan serán de una utilidad bastante alta para que los alumnos puedan ver la importancia de tener una cantidad de sueño ideal para su correcto desenvolvimiento en la universidad. Tomando en cuenta que no muchas personas le dan la relevancia necesaria a este aspecto, se podrán aplicar distintas medidas de concientización sobre los factores relevantes a tomar en cuenta para tener una mayor cantidad de horas de sueño.

MARCO CONCEPTUAL E HIPÓTESIS

Para la realización del presente trabajo, se determinó encuestar a los alumnos de la Universidad del Pacífico sobre factores considerados relevantes que afectarían su tiempo de sueño. Además, se decidió encuestar a una cantidad específica de alumnos de cada facultad según el análisis de la muestra realizado más adelante. Una vez con la información recolectada, se utilizarán técnicas de análisis de datos como la Regresión Lineal Múltiple, Regresión Logística y el Análisis de Componentes Principales que permitirán examinar la información recopilada y la correcta predicción del número de horas dormidas de un estudiante dependiendo de distintos factores relevantes además de si se encuentra en el intervalo de horas dormidas adecuadas diarias. Estas técnicas incluirán el uso de softwares que permitirán el correcto procesado de los datos tales como SPSS, además de lenguajes de programación tales como R y Python. Se hará uso del software GPower para hallar el tamaño de muestra apropiado para la investigación, el software estadístico SPSS y el lenguaje de programación Python para el procesamiento de los datos de las muestras, la creación de los modelos de regresión y técnicas de reducción de dimensiones, además del uso de R para realizar las predicciones con la data de testing mencionada en la metodología con los modelos que se crearán. Es importante mencionar que para todas las pruebas se usa un nivel de significancia de 0.05.

Se planteó al número de horas de sueño diarias promedio como la variable dependiente debido a la importancia que presenta fundamentada en los distintos trabajos revisados previamente durante la introducción, además de 11 variables independientes: edad, sexo, créditos en el ciclo, consumo de café, consumo de cigarro, horas diarias promedio de uso del celular, trabajo, facultad, participación en una organización estudiantil, tiempo de viaje de la universidad a su casa y las horas usadas en actividades físicas:

- *Edad*: La variable de edad fue seleccionada debido a su relevancia en diversos estudios, como los llevados a cabo por Zheng et al. (2023), Ban et al. (2001) y numerosos estudios adicionales sobre los factores que afectan a la cantidad de sueño de los estudiantes.

H1: Existe una relación estadísticamente significativa entre la edad de los estudiantes universitarios y el número de horas que duermen diariamente en promedio.

- *Sexo:* La variable ha sido ampliamente investigada en estudios que buscan explicar los factores asociados al tiempo de sueño de los universitarios. Un ejemplo de ello es el estudio de Lu, L et al. (2021), donde se analizan las diferencias en las horas de sueño entre los dos sexos y se encuentran diferencias significativas. Además, esta variable ha sido abordada en estudios similares, como el de Chen et al. (2014), y otros estudios relacionados, los cuales fueron motivos para incluirla en el modelo de investigación.

H2: Existe una relación estadísticamente significativa entre el sexo de los estudiantes universitarios y el número de horas que duermen diariamente en promedio.

- *Créditos en el ciclo:* La variable de créditos en el ciclo se deriva de la variable "Academic pressure" que se encuentra en el estudio de Lu, L et al. (2021). Los valores de créditos aceptados para esta variable se encuentran en un rango específico, que va desde 12 hasta 25. Estos límites representan tanto el mínimo como el máximo de créditos que la universidad permite llevar a los estudiantes, presentado en Universidad del Pacífico. (2023) Reglamento de Estudios de Pregrado 2023.

H3: Existe una relación estadísticamente significativa entre los créditos cursados en el ciclo de los estudiantes universitarios y el número de horas que duermen diariamente en promedio.

- *Consumo de café:* La variable "Coffee consumption" ha sido identificada en los estudios realizados por Zheng et al. (2023) y Ban, D. (2001). En el estudio de Ban, el intervalo de confianza de la variable no incluye el valor 0, lo que indica una relación significativa entre el consumo de café y las horas de sueño de los universitarios. Por otro lado, en el estudio de Zheng, el valor p del coeficiente de la variable es inferior a 0.01, lo que respalda aún más la importancia de considerar esta variable para explicar la variación en las horas de sueño.

Estos hallazgos respaldan la inclusión de la variable "Consumo de café" en el modelo. Además, se encuentra respaldada también por la investigación de Valenzuela, A. (2010), donde se menciona que el consumo de café puede afectar el sueño de las personas.

H4: Existe una relación estadísticamente significativa entre el consumo de café de los estudiantes universitarios y el número de horas que duermen diariamente en promedio.

- *Consumo de cigarro:* En los estudios de Ban (2001) y Lu, L et al. (2021), se investiga el impacto de una variable en las horas de sueño de los estudiantes. En el estudio de Ban, la variable se presenta como métrica, representando el número de cigarrillos que una persona fuma al día, mientras que en el estudio de Lu, la variable se mide de forma no métrica nominal, indicando si la persona fuma o no.

H5: Existe una relación estadísticamente significativa entre el consumo de cigarro de

los estudiantes universitarios y el número de horas que duermen diariamente en promedio.

- *Horas diarias promedio de uso del celular:* En el estudio de Uzunçakmak, T (2022), se analiza la variable "Daily smartphone usage duration", que es una variable métrica de escala razón medida en horas. Por otro lado, en el trabajo de Wang, F (2021), se investiga la variable "Adicción al smartphone", la cual es una variable no métrica de escala nominal. Ambas variables resultan significativas en sus respectivos modelos, lo que respalda su relevancia en investigaciones similares a la nuestra. En este modelo, hemos seleccionado la variable "Horas diarias de uso de celular".

H6: Existe una relación estadísticamente significativa entre las horas diarias promedio de uso del celular de los estudiantes universitarios y el número de horas que duermen diariamente en promedio.

- *Trabajo:* Es común que los estudiantes universitarios de tercer año en adelante comiencen a realizar prácticas en diversas empresas, e incluso hay estudiantes que trabajan para financiar sus estudios. Al dedicar tiempo al trabajo, sacrifican una cantidad considerable de tiempo que podrían destinar a otras actividades, como dormir. En el estudio realizado por Chiang, Y. (2020), se plantea la variable "Trabajo", que se presenta de manera cuantitativa y discreta.

H7: Existe una relación estadísticamente significativa entre el hecho de tener un trabajo por parte de los estudiantes universitarios y el número de horas que duermen diariamente en promedio.

- *Facultad:* La variable "Speciality" presentada por Zheng, W (2023) muestra la especialidad de cada uno de los estudiantes encuestados, por lo que se decidió categorizar a los estudiantes por su facultad. La variable presentada por Zheng se refiere a la especialidad que siguen los estudiantes, ya sea relacionada con medicina, ciencia e ingeniería o arte.

Dado que se trata de una variable nominal con más de dos posibles valores, se requiere utilizar variables ficticias (dummies). Para crear estas variables ficticias, se tomará como referencia a algún valor, ya sea la facultad de ingeniería, economía y finanzas o alguna de las otras facultades. Se elegirá a la variable de referencia según el promedio de sueño en cada facultad, la facultad con alumnos que tengan el menor promedio de sueño será la facultad elegida como referencia.

H8: Existe una relación estadísticamente significativa entre la facultad de los estudiantes universitarios y el número de horas que duermen diariamente en promedio.

- *Participación en organización estudiantil:* Antes de explicar la variable, es importante definir una organización estudiantil (OE) en la Universidad del Pacífico (UP). Una OE se crea con el propósito de beneficiar a los alumnos de la universidad

o a personas externas que requieran apoyo. Para que una OE sea reconocida por la universidad, debe cumplir con ciertos requisitos, como tener más de 40 miembros y contar con procesos para el ingreso a la organización y el desarrollo de actividades de acuerdo a lo mencionado en Memoria Participación Universitaria 2020 (2020). Actualmente, la UP cuenta con 30 OEs, lo que implica que una gran cantidad de alumnos se encuentran participando en ellas.

Teniendo ello en cuenta, la variable "Participación en organización estudiantil" es una forma de medir una parte de la carga extraacadémica que tienen algunos alumnos de la UP.

H9: Existe una relación estadísticamente significativa entre la participación en una organización estudiantil de los estudiantes universitarios y el número de horas que duermen diariamente en promedio.

- *Tiempo de viaje de la universidad a su casa:* Esta variable busca explicar la variabilidad en el tiempo de sueño según el estrés experimentado por los estudiantes debido a la duración del trayecto. Si el tiempo es largo y el alumno no cuenta con un medio de transporte propio, estará expuesto a niveles altos de estrés e inseguridad. En el estudio de Uzunçakmak, T (2022), "Prevalence and predisposing factors of smartphone addiction, sleep quality and daytime sleepiness of nursing students: A cross-sectional design", se considera la variable "School location", que hace referencia al lugar donde se encuentra la institución educativa a la que asiste el alumno. Esta variable es relevante, ya que llegar a una escuela en un área urbana tiende a ser más fácil y rápido que en una zona rural, cabe recalcar que en el estudio la variable "School location" fue catalogada como significativa.

H10: Existe una relación estadísticamente significativa entre el tiempo de viaje de la universidad a la casa de los estudiantes universitarios y el número de horas que duermen diariamente en promedio.

- *Actividad física:* En el estudio realizado por Wang, F. (2021), se examinan los principales factores que pueden influir en el sueño de los estudiantes, y uno de ellos es la variable determinante "Actividad física". Basándonos en este antecedente, se plantea la variable "Actividad física" en nuestra investigación. Esta variable es métrica, con una escala de razón y se medirá en horas. Específicamente, representa el promedio de tiempo que los alumnos dedican a realizar actividad física.

H11: Existe una relación estadísticamente significativa entre el promedio de tiempo dedicado a actividades físicas de los estudiantes universitarios y el número de horas que duermen diariamente en promedio.

METODOLOGÍA

Procedimiento

El proceso comienza con la recolección de información necesaria para el análisis. Se realiza un análisis de la muestra, identificando estadísticos descriptivos relevantes y llevando a cabo pruebas de hipótesis de interés. Luego, los datos se preprocesan, codificando variables nominales y creando variables dummies. Además, se evalúa la posibilidad de aplicar PCA.

Posteriormente, se separa el conjunto de datos en conjuntos de entrenamiento y prueba, siguiendo los porcentajes asignados por la facultad. A continuación, se procede al entrenamiento de los modelos, revisando los supuestos necesarios y presentando los resultados obtenidos. Finalmente, se prueba el desempeño de los modelos y se presentan medidas de error relevantes para evaluar su calidad en la predicción y explicación de los datos.

Instrumento de Medida

Basamos nuestro estudio en el análisis de las horas de sueño habituales de los estudiantes de la Universidad del Pacífico. Con ese fin, hemos identificado una variable dependiente llamada "Horas de sueño diarias promedio". Esta variable se caracteriza por ser métrica, tener una escala de razón y aceptar valores superiores a 0.

Ahora, se pasará a presentar la medición que se usará para cada una de las variables antes presentadas.

- *Edad*: Esta variable es métrica, posee una escala de razón y se medirá en años para los propósitos de nuestro estudio.
- *Sexo*: Es una variable no métrica y se utiliza una escala nominal con dos valores posibles. El valor "1" indica alumnos de sexo masculino, mientras que el valor "0" representa alumnos de sexo femenino.
- *Créditos en el ciclo*: Esta variable se caracteriza por ser métrica y tener una escala de razón. Los valores de créditos aceptados para esta variable se encuentran en un rango específico, que va desde 12 hasta 25.
- *Consumo de café*: Esta variable es cuantitativa, con una escala de razón, y se medirá en función de la cantidad de tazas de café que los alumnos suelen consumir por día.
- *Consumo de cigarro*: En los estudios de Ban (2001) y Lu, L et al. (2021), se investiga el impacto de una variable en las horas de sueño de los estudiantes. En el estudio de Ban, la variable se presenta como métrica, representando el número de cigarrillos que una persona fuma al día, mientras que en el estudio de Lu, la variable se mide de forma no métrica nominal, indicando si la persona fuma o no. Para esta investigación, se va a medir la variable de manera no métrica en una escala nominal, utilizando la

codificación "0" para representar a las personas que fuman y "1" para las que no fuman.

- *Horas diarias promedio de uso del celular*: Esta variable es métrica, con una escala de razón y se medirá en horas.
- *Trabajo*: Se plantea esta variable de manera no métrica y de escala nominal, donde el valor "0" indica que el estudiante trabaja y el valor "1" indica que no trabaja.
- *Facultad*: En el presente trabajo, se introduce la variable "Facultad", que es una variable no métrica de escala nominal con los siguientes valores: "1" para la facultad de Ingeniería, "2" para la facultad de Economía y finanzas, "3" para la facultad de Derecho y "4" para la facultad de Ciencias empresariales. Dado que se trata de una variable nominal con más de dos posibles valores, se requiere utilizar variables ficticias (dummies). Para crear estas variables ficticias, se tomará como referencia a algún valor, ya sea la facultad de ingeniería, economía y finanzas o alguna de las otras facultades. Se elegirá a la variable de referencia según el promedio de sueño en cada facultad, la facultad con alumnos que tengan el menor promedio de sueño será la facultad elegida como referencia. Por el momento, antes de codificar las variables dummies tenemos que la codificación de esta variable es de la siguiente manera: 1 representa a la facultad de Ingeniería, 2 representa a la facultad de economía y finanzas, 3 representa a la facultad de derecho y 4 representa a la facultad de ciencias empresariales.
- *Participación en organización estudiantil*: Esta variable es no métrica, su escala es nominal y su codificación es "0" si el alumno participa en al menos una organización estudiantil y "1" si no participa en ninguna organización estudiantil.
- *Tiempo de viaje de la universidad a su casa*: Se decidió medir esta variable en horas y se incluirá como una métrica en una escala de razón.
- *Actividad física*: Esta variable es métrica, con una escala de razón y se medirá en horas. Específicamente, representa el promedio de tiempo que los alumnos dedican a realizar actividad física.

A continuación se presenta la tabla 1, donde se resume el nombre de las variables y sus características.

Tabla 1: Variables y descripciones

Nombre de la variable	Pregunta correspondiente en la encuesta	Tipo de variable	Escala de medición	Categorías correspondientes
Horas de sueño diarias promedio	¿Cuántas horas en promedio duermes al día?	Métrica	Razón, medido en horas	No tiene categorías por ser variable métrica
Edad	¿Cuántos años tienes?	Métrica	Razón, medido en Años	No tiene categorías por ser variable métrica
Sexo	¿Cuál es tu género?	No métrica	Nominal	0=Masculino 1=Femenino
Créditos en el ciclo	¿Cuántos créditos estás matriculado en este ciclo?	Métrica	Razón, medido en número de créditos	No tiene categorías por ser variable métrica
Consumo de café	¿Cuántas tazas de café sueles consumir al día?	Métrica	Razón, medido en número de tazas de café al día	No tiene categorías por ser variable métrica
Consumo de cigarro	¿Consumes cigarrillos?	No métrica	Nominal	0=Sí consume 1= No consume
Horas diarias de uso del celular	¿Cuántas horas al día utilizas tu teléfono celular en promedio?	Métrica	Razón, medido en horas	No tiene categorías por ser variable métrica
Trabajo	¿Te encuentras trabajando, ya sea de manera formal o informal?	No métrica	Nominal	0= Sí trabaja 1= No trabaja
Facultad	¿A qué facultad perteneces?	No métrica	Nominal	1= Facultad de Ingeniería 2=Facultad de Economía y Finanzas 3= Facultad de Derecho 4= Facultad de Ciencias Empresariales

Tiempo promedio de viaje hasta la universidad	¿Cuánto tiempo sueles tardar en llegar a tu hogar?	Métrica	Razón, medida en minutos	No tiene categorías por ser variable métrica
Participación en organización estudiantil	¿Eres miembro activo de alguna organización estudiantil?	No métrica	Nominal	0 = Sí participa 1 = No participa
Actividad física	¿Cuánto tiempo dedicas al ejercicio físico a la semana?	Métrica	Razón, medido en horas	No tiene categorías por ser variable métrica

Colección de datos y muestra

El propósito de este estudio consiste en llevar a cabo un trabajo actual utilizando una fuente de datos primaria, donde se recopilarán los datos a través de encuestas individuales. El objetivo primordial de esta encuesta es obtener tanto datos cuantitativos como cualitativos que permitan comprender los diversos factores que ejercen influencia en la cantidad de horas de sueño que los estudiantes de la Universidad del Pacífico experimentan. Con el fin de lograr una muestra que nos ayude a entender a la población, se emplea un método de muestreo no aleatorio de tipo cuota de acuerdo con la cantidad de alumnos por facultad, donde cada cuota se corresponderá con cada una de las facultades presentes en la universidad, es decir, Ingeniería, Ciencias Empresariales, Derecho y Economía y Finanzas. Se buscará que el número de respuestas sea proporcional a la cantidad de alumnos por facultad, según el reporte publicado por la Universidad del Pacífico titulado “*Alumnos matriculados y egresados pregrado 2022*” (Universidad del Pacífico, 2022) y mostrado en la tabla 2. Se decidió utilizar la información de los alumnos matriculados en el 2022-2, debido a que es la información más reciente que se tiene sobre la cantidad de alumnos matriculados por facultad en la casa de estudios.

Tabla 2: Alumnos matriculados en el semestre académico 2022-2 por facultad

Facultad	Cantidad de matriculados el 2022-2
Ciencias Empresariales	1927
Economía y Finanzas	1748

Derecho	386
Ingeniería	886

A continuación se detallarán las especificaciones de la muestra:

1. *Población objetivo*: La población objetivo de este estudio está compuesta por todos los estudiantes matriculados en la Universidad del Pacífico durante el semestre académico 2023-1, que cumplan con los siguientes criterios: haber inscrito un mínimo de 12 créditos y un máximo de 25 créditos, este criterio sigue lo mencionado en el Reglamento de Estudios de Pregrado 2023 de la Universidad del Pacífico. Es importante destacar que esta población no incluye a los egresados ni a aquellos estudiantes que han obtenido un permiso especial para cursar un número inferior de créditos al requerido por la universidad.
2. *Unidad de investigación*: Alumno de la universidad del pacifico.
3. *Periodo de ejecución*: Mes de junio del 2023
4. *Periodo de evaluación*: Mes de junio de 2023
5. *Metido de colección*: Encuesta al colectivo
6. *Detalles de la encuesta*: El formato seleccionado para llevar a cabo la encuesta será en línea, utilizando Microsoft Forms. Esta elección se basa en la recomendación de Leonardo Rocco (2007), quien afirma que esta modalidad ofrece ventajas significativas, como una mayor rapidez en el análisis de los datos, una mejor validación de los mismos y la reducción de costos asociados.
7. *Tamaño de muestra*: Para calcular el tamaño de muestra, se utilizó el software "G-Power". El software "G-Power" es una herramienta estadística ampliamente utilizada que ofrece cálculos de tamaño de muestra. Permite determinar la cantidad adecuada de participantes necesarios en un estudio, y ofrece ventajas como la capacidad de especificar diferentes parámetros estadísticos y seleccionar el tipo de análisis requerido. En nuestro caso, hemos optado por utilizar el enfoque *A priori* para calcular el tamaño de muestra antes de llevar a cabo el estudio. Este enfoque nos permite determinar el número necesario de participantes para garantizar un adecuado poder estadístico. Nuestro enfoque para determinar el tamaño de muestra es *A priori*, el cual se utiliza antes de realizar el estudio o experimento para asegurar un adecuado poder estadístico. En este caso, se emplea el test estadístico *Linear multiple regression: Fixed model, R^2 deviation from zero* perteneciente a la familia de tests

F-test. Este test se utiliza en el contexto de regresión múltiple y nos permitirá evaluar las relaciones entre las variables de interés. Al especificar los parámetros requeridos, el software "G-Power" calcula el tamaño de muestra necesario con base en los mismos. En nuestro caso, se han utilizado cuatro parámetros para el cálculo del tamaño de muestra. El primero es el *Effect size* f^2 , que representa la correlación que se sospecha que existe entre las variables. De acuerdo con el artículo de Faul, F (2009), se menciona que el valor ideal para una correlación grande es de 0.35, por lo que se utilizó este valor. El segundo parámetro es el nivel de significancia, establecido en 0.05. Este valor representa la probabilidad de cometer un error de tipo I, es decir, rechazar incorrectamente la hipótesis nula. En tercer lugar, se especifica una potencia de prueba de 0.9. De acuerdo con C. Fuentelsaz (2004), este valor se considera ideal para garantizar una alta probabilidad de detectar efectos o relaciones significativas en el estudio. Por último, el cuarto parámetro es el número de predictores, que en nuestro caso es 11. Esto se refiere al número de variables predictoras que se están considerando en la regresión múltiple lineal. Al ingresar estos valores en el software "G-Power", se obtuvo un tamaño mínimo de muestra de 72 alumnos de la Universidad del Pacífico, como se muestra en el Anexo 1. Sin embargo, dado que se busca que la muestra sea lo más representativa posible, se respetarán los porcentajes de la población total que representa cada grupo, según se muestra en la tabla 3:

Tabla 3: Porcentaje y cantidad a encuestar de alumnos según facultad

Facultad	Porcentaje de alumnos según facultad	Cantidad de alumnos a encuestar según facultad
Ciencias Empresariales	39%	28
Economía y Finanzas	35%	25
Derecho	8%	6
Ingeniería	18%	13

Análisis de Datos

En el presente estudio se utilizarán las siguientes técnicas:

- 1) *PCA*: Se utilizará la técnica de componentes principales para las variables métricas que tengamos y así poder eliminar algún posible problema de multicolinealidad que exista entre las variables y de ese modo lograr una mejor predicción con el modelo de Regresión Lineal Múltiple o Regresión Logística.
- 2) *Regresión Lineal Múltiple*: Con esta técnica se busca entender el comportamiento de las horas de sueño de los alumnos de la Universidad del Pacífico y así analizar las variables que tienen más impacto en esta variable.
- 3) *Regresión Logística*: Esta técnica será usada para clasificar a los alumnos que duermen bien y a los alumnos que duermen mal. Como se menciona en el artículo publicado por Hirshkowitz et al. (2015), para que se considere que una persona duerme bien, la cantidad de horas en promedio que duerme al día se debe encontrar en el rango de 7 a 9 horas. Teniendo en cuenta eso, se recodificará la variable dependiente “Horas de sueño diarias promedio” a ceros y unos donde el cero significa que la respuesta del alumno se encuentra fuera del rango sugerido por Hirshkowitz y uno significa que la respuesta se encuentra en el rango sugerido.

El presente estudio, mediante todas estas técnicas antes mencionadas, busca entender el comportamiento de la variable dependiente “Horas de sueño diarias promedio” y clasificar a los estudiantes que duermen la cantidad de horas sugerida y a los estudiantes que no duermen la cantidad de horas sugeridas según el criterio expuesto en Hirshkowitz et al. (2015): National Sleep Foundation's updated sleep duration recommendations: final report, Sleep Health. El proceso de estudio será de la siguiente manera: Primero se hará un análisis de los datos obtenidos por la encuesta y se verificará si hay variables correlacionadas, si este es el caso, se aplicará el Análisis de Componentes Principales (PCA) para así eliminar la correlación entre variables. Luego se separará a la data de entrenamiento (training set), equivalente al 80% de la data, y a la data de testeo (testing set), equivalente al 20% de la data. Esta forma de dividir la data coincidirá con la cantidad de alumnos por facultad y así conseguir una mayor representatividad tanto en el training set como en el testing set. Una vez tengamos la data separada, en el training set se crearán los coeficientes de los modelos, tanto de Regresión Lineal Múltiple como de Regresión Logística y se evaluará el error de predicción de los modelos. Por último, al modelo ya creado se le ingresarán los datos del testing set para evaluar el error que arroja el modelo con estos datos. Para este análisis se utilizará el software SPSS y el lenguaje de programación Python.

Análisis de la muestra

Según los objetivos establecidos, se realizará un análisis de la muestra para comprender el comportamiento de los alumnos encuestados. Se encontró que en la facultad de Ciencias Empresariales, los alumnos duermen aproximadamente la misma cantidad de tiempo que utilizan sus celulares, como se muestra en la tabla 4. Es importante destacar que ambas variables están medidas en horas diarias, lo que permite una comparación válida.

Tabla 4: Medias de horas de sueño y horas de uso de teléfono por facultad

Facultad	Media de horas sueño	Media de horas de uso de teléfono
Ciencias Empresariales	5.68	5.48
Derecho	5.33	4.66
Economía y Finanzas	6.42	4.72
Ingeniería	6.31	4.58

En el análisis de la muestra, se encontró que en la facultad de Ciencias Empresariales los alumnos duermen casi el mismo tiempo que utilizan sus celulares, como se muestra en la tabla 4. Sin embargo, es aún más sorprendente que sólo en las facultades de ingeniería y economía y finanzas existe una diferencia de más de 2 horas entre el promedio de sueño y el promedio de uso del celular. Esta disparidad podría resultar contraproducente, ya que los alumnos pasan casi la misma cantidad de tiempo durmiendo que utilizando sus dispositivos móviles.

Además, otro dato relevante es que los alumnos que no forman parte activa de alguna Organización Estudiantil en la universidad duermen ligeramente más que aquellos que sí lo son, como se observa en la tabla 5. Aunque la diferencia no es muy amplia, es importante considerar que la carga adicional de ser miembro de una organización estudiantil afecta las horas de sueño de los alumnos de la Universidad del Pacífico.

Tabla 5: Media de horas de sueño de alumnos si pertenecen o no a una Organización Estudiantil

Miembro activo de una Organización Estudiantil	Horas de sueño
Sí	5.92
No	6.12

En este escenario, resulta relevante realizar una prueba de hipótesis con el objetivo de determinar si la media de sueño de los alumnos de la Universidad del Pacífico es inferior a las siete horas mínimas recomendadas por Hirshkowitz et al. (2015). Para llevar a cabo esta

evaluación, utilizaremos un nivel de significancia establecido en 0.05. Esta prueba de hipótesis nos permitirá concluir si, estadísticamente, existen pruebas suficientes para rechazar la hipótesis nula, la cual establece que la media de sueño es igual o mayor a siete horas, a favor de la hipótesis alternativa, que sugiere que la media de sueño es inferior a siete horas.

Tabla 6: Prueba de hipótesis

Prueba de hipótesis	Valores	Tabla Z
$H_0: \mu \geq 7$	$\bar{X}=6.02$	$Z_{\text{tabla}} = Z(0.05) = -1.645$
$H_1: \mu < 7$	$S=1.089$	$Z_{\text{cal}} = (6.02-7)/((1.089)(\sqrt{72})) = -7.636$

Z_{tabla} : Es el valor de la tabla Z el cual corresponde a un nivel de significancia de 0.05

Z_{cal} : Es el estadístico de prueba que es comparado con el valor obtenido de la Z_{tabla}

Decisión: Como el valor de Z_{cal} es mayor al valor de Z_{tabla} , se rechaza la hipótesis nula.

Conclusión: Con un nivel de significancia de 0.05, existe suficiente evidencia estadística para afirmar que la media de las horas de sueño de los alumnos de la Universidad del Pacífico es menor a 7 horas.

Preprocesamiento de los datos

Una vez recolectados los datos de los alumnos de la universidad del pacifico se procesarán los valores de las variables no métricas de la encuesta para que puedan ser utilizadas en las técnicas seleccionadas. Estas variables son: Sexo, consumo de cigarros , trabajo, participación en organización estudiantil y facultad. Para este procedimiento se utilizará el lenguaje de programación python. Primero se recodificara la variable sexo teniendo en cuenta que “masculino” es 0 y “femenino” es 1, luego, se recodificará consumo de cigarro, trabajo y participacion en organización estudiantil con una respuesta afirmativa a la pregunta de la respectiva pregunta como 0 y una respuesta negativa como 1. Por último, como se observa en la tabla número 4, la facultad con una media de sueño menor es la facultad de Derecho y por eso la tomaremos como el valor de referencia para la creación de las variables dummies. Estas variables se verían como lo presentado en la tabla 7.

Tabla 7: Codificación de las variables dummies que representan la facultad de los alumnos

Facultad	Dummy de Economía y Finanzas	Dummy de Ingeniería	Dummy de Ciencias Empresariales
Facultad de Derecho	0	0	0
Facultad de Ingeniería	0	1	0

Facultad de Ciencias Empresariales	0	0	1
Facultad de Economía y Finanzas	1	0	0

Adicionalmente se creó una variable llamada “logística” que se utilizará solo para la regresión logística, esta variable es no métrica, nominal y tiene 2 posibles valores: 1 si es que el alumnos encuestado duerme más de 7 horas o 7 horas y 0 si es que el alumno encuestado duerme menos de 7 horas. Para más detalle, revisar el apéndice N°1 donde se detalla el código de python utilizado para este proceso.

El siguiente paso es probar si es que la técnica de componentes principales es aplicable para evitar posibles problemas de multicolinealidad, por lo que, lo primero que debemos hacer es analizar las varianzas entre las variables para poder saber si debemos de usar una matriz de correlaciones o una matriz de correlaciones.

Tabla 8: Varianza de las variables independientes métricas

Variable	Varianza
Edad	1.937
Créditos en el ciclo	10.337
Consumo de café	2.070
Horas diarias de uso de celular	5.967
Tiempo promedio de viaje hasta la universidad	1129.042
Actividad física	8.181

Como se ve en la tabla 8, las varianzas de las variables difieren mucho y por ello se debería de usar la matriz de correlaciones. Con esta información se hizo el test de esfericidad de Bartlett en el SPSS en el cual la hipótesis nula fue: la determinante de la matriz de correlaciones es igual a 1. Por otro lado, la hipótesis alternante fue: la determinante de la matriz de correlaciones no es igual a 1. Y los resultados fueron los siguientes:

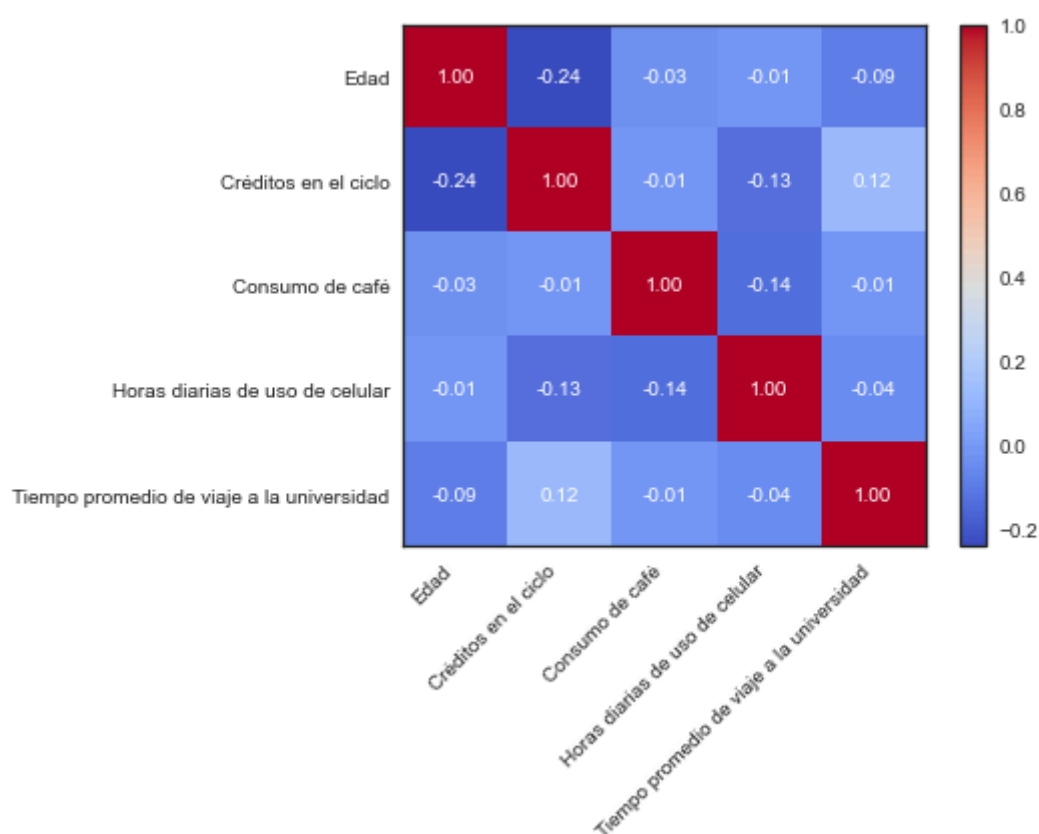
Tabla 9: Prueba de Bartlett

Prueba de esfericidad de Bartlett	Aprox. Chi-Cuadrado	0.519
	gl	15
	Sig	0.302

Como se observa en la Tabla 9, el p-value de la prueba de esfericidad de Bartlett es 0.302,

valor que es mayor al nivel de significancia con el que se hizo la prueba, 0.05. Esto quiere decir que se acepta la hipótesis nula y por ende podemos concluir que el determinante de la matriz de correlaciones de las variables expuestas en la tabla 7 es cercano a 1. Por lo mismo, no es adecuado realizar la técnica de componentes principales. Teniendo en cuenta este resultado resulta necesario comprobar la magnitud de las correlaciones entre las variables para asegurarnos que no es necesario aplicar la técnica y para ello se realizó un mapa de correlaciones.

Figura 1: Mapa de correlaciones



En la Figura 1, se muestra el mapa de correlaciones entre las variables métricas independientes. Observamos que las correlaciones son cercanas a cero y ninguna supera un valor absoluto de 0.5. Esto confirma que no es necesario aplicar la técnica de reducción de dimensiones mediante los componentes principales (PCA). El mapa de correlaciones fue generado en Python y se proporciona el código detallado en el Apéndice 2.

Para garantizar una representatividad adecuada de la población original, es necesario establecer el conjunto de datos de entrenamiento (training set) y el conjunto de datos de prueba (testing set) antes de iniciar el entrenamiento de los modelos. Los porcentajes indicados en la Tabla 3 serán seguidos para esta división. En la Tabla 10 se especifica la cantidad de alumnos por facultad que se asignarán tanto al conjunto de entrenamiento como al conjunto de prueba. El proceso de asignación se realizará utilizando el lenguaje de programación R, y se encuentran más detalles en el Apéndice 3.

Tabla 10: Cantidad de training y testing set

Facultad	Cantidad de alumnos en el training set	Cantidad de alumnos en el testing set
Ciencias Empresariales	23	5
Economía y Finanzas	20	5
Ingeniería	11	2
Derecho	4	2

Entrenamiento de modelos

1. Regresión lineal múltiple

1.1 Revisión de supuestos

- 1.1.1. *Variable dependiente métrica:* Tal y como se muestra en la tabla 1, la variable dependiente es métrica.
- 1.1.2. *Exogeneidad:* Según se observa en el anexo 2, que resume el modelo de regresión, el valor de Durbin-Watson se sitúa entre 1.5 y 2.5. En consecuencia, podemos inferir que los errores no presentan correlación entre ellos.
- 1.1.3. *Multicolinealidad:* De acuerdo con los resultados presentados en la tabla 11, se observa que las variables dummy correspondientes a las facultades de Ciencias Empresariales y Economía y Finanzas muestran problemas de multicolinealidad. Una solución común en estos casos sería transformar las variables o eliminar una de ellas. Sin embargo, ambas variables poseen una gran relevancia teórica, dado que los alumnos de estas facultades cursan diferentes asignaturas. Además, como se evidenció en la tabla 4, se observan diferencias significativas en las medias de horas de sueño entre ellos.

Tabla 11: Indicadores de cada variable

Variable	Coefficiente	Tolerance	VIF
Constante	13.809		
Edad	-0.304	0.622	1.607
Sexo	-0.010	0.624	1.603
Créditos en el ciclo	-0.66	0.774	1.292
Consumo de café	-0.090	0.632	1.583

Consumo de cigarro	-0.232	0.807	1.239
Horas en el celular	-0.061	0.751	1.332
Dummy Facultad de Economía	1.783	0.182	5.506
Dummy Facultad de Ingeniería	1.322	0.230	4.346
Dummy Facultad de Ciencias Empresariales	1.000	0.184	5.442
Trabajo	-0.397	0.686	1.458
Tiempo de viaje de la universidad a su hogar	-0.016	0.759	1.318
Participación en organización estudiantil	-0.001	0.802	1.248
Actividad física	0.003	0.843	1.186

1.1.4. *Normalidad:* Según se puede apreciar en el Anexo 3, los datos se adhieren a la distribución normal, lo cual confirma el cumplimiento de este supuesto.

1.1.5. *Homocedasticidad:* Según se observa en el Anexo 4, tanto los residuos estudiantizados como los valores predichos estandarizados no presentan una tendencia clara y exhiben un comportamiento aleatorio. Por consiguiente, podemos concluir que se satisface el supuesto de homocedasticidad.

1.1.6. *Linealidad:* Este supuesto se verificará sólo para las variables métricas.

1.1.6.1. La variable edad cumple con el supuesto, ya que a mayor edad, menores son las horas de sueño promedio, sin embargo, la relación de linealidad no es muy fuerte. Esta información se puede corroborar en el Anexo 5

1.1.6.2. La variable créditos en el ciclo cumple con el supuesto de linealidad, mientras más créditos se lleven, menos horas dormirá el alumno, este resultado se observa en el anexo 6. No

obstante, es importante tener en cuenta que hay algunas observaciones que se escapan de esta relación lineal, para algunos casos, un número de créditos elevado, veinte o más, no necesariamente significa dormir menos de 6 horas.

- 1.1.6.3. La variable consumo de café muestra una tendencia lineal negativa no muy fuerte. Los resultados se pueden observar en el Anexo 7.
- 1.1.6.4. La variable tiempo en el celular, no tiene una relación lineal con la variable dependiente, no obstante, la mantendremos porque en el estudio de Wang, F (2021) es una variable significativa y se puede analizar el peso de la variable durante la pandemia y post pandemia. Este resultado se observa en el Anexo 8
- 1.1.6.5. En el Anexo 9 se puede observar una relación lineal negativa entre la variable "Tiempo promedio de viaje hasta la universidad" y la variable dependiente. Esto implica que a medida que el tiempo de desplazamiento desde o hacia la universidad aumenta, la cantidad de horas de sueño del alumno tiende a disminuir.
- 1.1.6.6. La variable Actividad física tiene una relación lineal positiva con la variable dependiente, esto significa que mientras más actividad física realiza el alumno, mayor será la cantidad de horas de sueño que tenga el alumno. Este resultado se ve en el anexo. Lo podemos ver en el Anexo 10.

1.2. Interpretación del modelo

- 1.2.1. *Test anova*: El test de ANOVA se utiliza para determinar si hay al menos algún coeficiente significativo en el modelo. Este test evalúa las siguientes hipótesis:

$$H_0: \beta_1 = \beta_2 = \beta_3 = \dots = \beta_{13} = 0$$

$$H_1: \text{Al menos un } \beta_j \neq 0$$

La información necesaria para realizar este test se encuentra en la Tabla 12. Al examinar el valor p en la tabla, podemos llegar a conclusiones sobre las hipótesis planteadas. Dado que el valor p es mayor que el nivel de significancia especificado en la introducción (0.05), rechazamos la hipótesis nula. Por lo tanto, podemos concluir que hay al menos un coeficiente distinto de cero, lo que indica que el modelo de regresión lineal es adecuado.

Tabla 12: ANOVA

Modelo	Suma de cuadrados	Grados de libertad	Media cuadrática	Estadístico F	P-Valor
Regresión	30.486	13	2.345	3.075	0.003
Residuo	33.553	44	0.763		
Total	64.039	57			

- 1.2.2. Ecuación de regresión y significancia de las variables: Teniendo en cuenta todo lo anterior, resulta conveniente realizar la regresión lineal múltiple para examinar la significancia de las variables. La ecuación de regresión correspondiente se muestra en la Figura 2.

Figura 2: Ecuación de regresión lineal

$$Y = \beta_0 + \beta_1 * Edad + \beta_2 * Sexo + \beta_3 * Créditos + \dots + \beta_{13} * ActividadFísica.$$

El modelo de regresión lineal múltiple desarrollado revela un coeficiente de determinación (R²) de 47.6%. Esto indica que el 47.6% de la variabilidad en las horas promedio de sueño diarias de los alumnos de la Universidad del Pacífico puede ser explicada por la ecuación de regresión estimada, la cual incluye las siguientes variables independientes: edad, sexo, créditos en el ciclo, consumo de café, consumo de cigarrillos, horas diarias de uso del celular, trabajo, facultad, tiempo promedio de viaje hasta la universidad, participación en organización estudiantil y actividad física.

Tabla 13: Significancia de variables

Variable	Coefficiente	Significancia
Constante	13.809	<0.001
Edad	-0.304	0.013
Sexo	-0.010	0.971
Créditos en el ciclo	-0.66	0.098
Consumo de café	-0.090	0.354
Consumo de cigarro	-0.232	0.690
Horas en el celular	-0.061	0.338

Dummy Facultad de Economía	1.783	0.003
Dummy Facultad de Ingeniería	1.322	0.036
Dummy Facultad de Ciencias Empresariales	1.000	0.074
Trabajo	-0.397	0.328
Tiempo de viaje de la universidad a su hogar	-0.016	<0.001
Participación en organización estudiantil	-0.001	0.996
Actividad física	0.003	0.941

1.3. Significancia de las Variables

- 1.3.1. La variable "Edad" tiene una significancia de 0.013. Esto indica que con un nivel de significancia de 0.05 existe suficiente evidencia estadística para afirmar que la variable "Edad" es significativa para el modelo.
- 1.3.2. La variable "Sexo" tiene una significancia de 0.971. Esto indica que con un nivel de significancia de 0.05 no existe suficiente evidencia estadística para afirmar que la variable "Sexo" es significativa para el modelo.
- 1.3.3. La variable "Créditos en el ciclo" tiene una significancia de 0.098. Esto indica que con un nivel de significancia de 0.05 no existe suficiente evidencia estadística para afirmar que la variable "Créditos en el ciclo" es significativa para el modelo.
- 1.3.4. La variable "Consumo de Café" tiene una significancia de 0.354. Esto indica que con un nivel de significancia de 0.05 no existe suficiente evidencia estadística para afirmar que la variable "Consumo de Café" es significativa para el modelo.
- 1.3.5. La variable "Dummy Facultad de Economía" tiene una significancia de 0.003. Esto indica que con un nivel de significancia de 0.05 existe suficiente evidencia estadística para afirmar que la variable "Horas en el celular" es significativa para el modelo.
- 1.3.6. La variable "Tiempo de viaje de la universidad a su hogar" tiene una significancia de <0.001. Esto indica que con un nivel de significancia de 0.05 existe suficiente evidencia estadística para afirmar que la variable "Tiempo de viaje de la universidad a su hogar" es significativa para el modelo.

2. Regresión logística

2.1. Revisión de supuestos

- 2.1.1. *Multicolinealidad:* Los valores de VIF y tolerancia en el análisis de multicolinealidad muestran resultados similares a los obtenidos en la regresión lineal múltiple. A pesar de que se identificaron las mismas variables problemáticas, se ha decidido mantenerlas en el modelo por la misma razón mencionada durante la revisión de supuestos en la regresión lineal múltiple.

Tabla 15: Tolerance y VIF para regresión logística

Variable	Tolerance	VIF
Edad	0.622	1.607
Sexo	0.624	1.603
Créditos en el ciclo	0.774	1.292
Consumo de café	0.632	1.583
Consumo de cigarro	0.807	1.239
Horas en el celular	0.751	1.332
Dummy Facultad de Economía	0.182	5.506
Dummy Facultad de Ingeniería	0.230	4.346
Dummy Facultad de Ciencias Empresariales	0.184	5.442
Trabajo	0.686	1.458
Tiempo de viaje de la universidad a su hogar	0.759	1.318
Participación en organización estudiantil	0.802	1.248
Actividad física	0.843	1.186

- 2.1.2. *Linealidad:* Para verificar la linealidad, utilizaremos un valor de significancia corregido, que en este caso es igual a $0.05/13 = 0.00384615$. Este valor corregido de alfa nos permitirá comparar los p-valores obtenidos al multiplicar las variables por su logaritmo neperiano. Si el resultado no es significativo, podemos concluir que se

cumple la suposición de linealidad. Tal y como se ve en la tabla 16, todas las variables métricas cumplen con la linealidad para el modelo de regresión logística, ya que todos los p-valores son no significativos.

Tabla 16: Tolerance y VIF para regresión logística

Variable	Significancia
Edad * ln(Edad)	1.000
Créditos por ciclo * ln(Créditos por ciclo)	1.000
Consumo de café * ln(Consumo de café)	1.000
Tiempo de uso del celular * ln(Tiempo de uso del celular)	1.000
Tiempo de viaje de casa a la universidad * ln(Tiempo de viaje de casa a la universidad)	0.999
Actividad física * ln(Actividad física)	1.000

2.2. Interpretación del modelo

2.2.1. *Prueba de ómnibus:* Como se ve en la tabla 17, el p-valor de la prueba de ómnibus de coeficientes del modelo es 0.01. Esta prueba responde a las siguientes hipótesis:

$H_0: \beta_1 = \beta_2 = \beta_3 = \dots = \beta_{13} = 0$

$H_1: \text{Al menos un } \beta_j \neq 0$

Tabla 17: Prueba de ómnibus

	Chi-Cuadrado	G.L.	Significancia
Paso	34920	13	<0.001
Bloque	34920	13	<0.001
Modelo	34920	13	<0.001

Como el p-valor es menor al nivel de significancia, podemos concluir que hay al menos un coeficiente distinto de 0 y por ello es idóneo hacer la técnica.

2.2.2. *Prueba de Hosmer y Lemeshow:* Esta es otra prueba que respalda la necesidad de realizar el análisis es una prueba estadística relevante. Esta prueba se basa en las siguientes hipótesis:

H0: Los datos se ajustan al modelo

H1: Los datos no se ajustan al modelo

Como se ve en la tabla 18 el p-valor es mayor que el nivel de significancia, por lo que se acepta la hipótesis nula y podemos concluir que con un nivel de significancia de 0.05 los datos se ajustan al modelo de regresión.

Tabla 18: Prueba de Hosmer y Lemeshow

Paso	Chi-Cuadrado	G.L.	Significancia
1	14335	8	0.073

2.2.3. *Modelo de regresión logística y coeficiente del modelo:* El modelo de regresión logística proporciona los siguientes resultados, que se presentan en la Tabla 19. El modelo ha clasificado correctamente 53 observaciones y 5 observaciones de manera incorrecta. Esto implica que el Accuracy del modelo es del 91.4%, lo que significa que el error de clasificación (MAPE) es del 8.6%. Esto indica que el 8.6% de las observaciones ha sido clasificado incorrectamente.

Tabla 19: Tabla de clasificación

			Pronosticado		
Paso 1	Observado		Log		Porcentaje correcto
			0	1	
	Log	0	34	2	94.4
		1	3	19	86.4
	Porcentaje global				91.4

Estos resultados se complementan con la Tabla 20, donde se muestra que el área bajo la curva ROC es de 0.913. Esta cifra indica que nuestro modelo es un excelente discriminador, lo cual se confirma en dicha tabla. El gráfico de la curva ROC se encuentra disponible en el Anexo 11.

Tabla 20: Tabla de clasificación

Área	Desv. Error	Significación asintótica	95% de intervalo de confianza asintótico	
			Límite inferior	Límite superior
0.913	0.040	0.000	0.835	0.991

Y la ecuación del modelo de regresión logística sería la siguiente:

Figura 3: Modelo de regresión logística

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 edad + \beta_2 Sexo + \beta_3 Creditos\ en\ el\ ciclo + \dots + \beta_{13} Actividad\ fisica)}}$$

La ecuación tiene 11 coeficientes cuyos valores y Odd-Ratios se presentan a continuación en la tabla 21.

Tabla 21: Variables de regresión logística

Variable	Coefficiente	Significancia	Odd-Ratio
Edad	-0.489	0.236	0.614
Sexo	0.037	0.968	1.037
Créditos en el ciclo	-0.069	0.652	0.934
Consumo de café	-0.290	0.519	0.748
Consumo de cigarro	17.508	0.999	40140880.912
Horas en el celular	-0.46	0.835	0.955
Dummy Facultad de Economía	23.252	0.999	12535021744
Dummy Facultad de Ingeniería	20.801	0.999	1080944467.5
Dummy Facultad de Ciencias Empresariales	20.136	0.999	555706229.14
Trabajo	-0.508	0.740	0.602
Tiempo de viaje de la universidad a su hogar	-0.065	0.006	0.937
Participación en organización estudiantil	1.078	0.217	2.938
Actividad física	-0.094	0.527	0.910

2.3. Coeficientes de las variables

Por lo expuesto en el marco teórico, a pesar que solo una de las variables, “Tiempo de viaje de la universidad a su hogar”, sea significativa, de todas formas se incluirán en el modelo todas por su relevancia teórica y para su posterior análisis.

Testeo de modelos

Regresión lineal múltiple

Este procedimiento implica utilizar los coeficientes presentados en la tabla 13 y aplicarlos a una nueva muestra de datos, llamada testing set. El resultado será una predicción de la media de horas de sueño diario para los datos del testing set. Esta predicción se comparará con el valor real de la variable y se calculará el error cuadrático medio para determinar cuánto se desvía en promedio el valor predicho del valor real. Los valores predichos y reales se muestran en la tabla 22.

Tabla 22: Predicciones

Número de observación	Valor real de la variable dependiente	Valor predicho de la variable dependiente
1	6	5.352271
2	6	6.211644
3	6	6.138606
4	4	3.373941
5	6	5.941063
6	7	6.990994
7	8	6.335153
8	6	6.662143
9	4	6.019056
10	4	5.477385
11	7	6.969540
12	7	6.548274
13	6	3.771535
14	5	5.521805

Con los resultados de la tabla 22, la medida de error escogida, error cuadrático medio, es de 1.062066. Esto significa que los valores predichos se alejan en 1.063066 unidades en promedio de los valores reales. A pesar de no parecer un error muy amplio, según el artículo realizado por Chen et al (2014) una hora de sueño es un valor importante debido a que dormir una hora más influye en la salud. Estos resultados fueron hechos con el lenguaje de programación R, y el código se encuentra en el Apéndice 4.

Regresión Logística

Se llevará a cabo la validación del modelo mediante pruebas con conjuntos de datos independientes del conjunto de entrenamiento, con el objetivo de evaluar su capacidad de clasificación. Para tal fin, se analizarán dos métricas clave: El accuracy y el área bajo la curva ROC.

Tabla 23: Test de predicción

		Predichos	
		0	1
Observados	0	9	1
	1	0	4

Los resultados obtenidos, según se reflejan en la tabla 23, revelan un alto desempeño del modelo evaluado. De un total de 15 observaciones, se logró una correcta predicción en 14 casos, lo que se traduce en una tasa de precisión del 93.3%. Este valor demuestra la capacidad del modelo para realizar estimaciones acertadas en situaciones no previamente vistas. Además, la curva ROC, muestra un área bajo la curva de 0.925. Este indicador cuantifica la habilidad del modelo para distinguir entre clases positivas y negativas, siendo un valor superior a 0.9 que nuestro modelo discrimina de manera excelente. Con el propósito de ofrecer una comprensión visual completa, la curva ROC se encuentra disponible en el Anexo 12 de este trabajo de investigación. Para aquellos interesados en los detalles técnicos, se ha adjuntado el código utilizado en el apéndice 5, lo cual permitirá replicar y verificar los procedimientos implementados en el estudio.

DISCUSIÓN E IMPLICANCIAS

Discusión

Al evaluar los resultados obtenidos en nuestro estudio, podemos concluir que se cumplió el objetivo de reconocer características o comportamientos específicos que influyen en la cantidad de sueño que un alumno tiene. El análisis realizado para la predicción de las horas de sueño diarias de un estudiante según los factores planteados mediante una regresión lineal

múltiple demostró que las variables de “Edad”, “Dummy Facultad de Economía”, “Dummy Facultad de Ingeniería” y “Tiempo de viaje de la universidad a su hogar” son significativas, como se puede observar en la tabla 13, por lo que se aprueban las hipótesis número 1, 8 y 10.

Con respecto al análisis realizado para determinar si un estudiante dormía las horas adecuadas mediante la regresión logística, sólo se aprueba una de las hipótesis, que plantea la existencia de una relación estadísticamente significativa entre el tiempo de viaje de la universidad a la casa de los estudiantes universitarios y el número de horas que duermen diariamente en promedio. Las demás variables no resultaron significativas como se puede revisar en la tabla 21 de los coeficientes de las variables y sus significancias. Esto se puede corroborar con el área debajo de la curva ROC (AUC) observado en el Anexo 11, el cual es 0.913, indicando que es un excelente discriminador. Con estos resultados podemos concluir también que se cumplió el segundo objetivo: la creación de un modelo que pueda determinar si un estudiante está durmiendo las horas adecuadas.

En relación al tercer y último objetivo, que consistía en identificar patrones en las características encontradas, se pudo constatar que no se observaron patrones discernibles en las muestras analizadas.

En el estudio de Zheng et al. (2023) la variable de “consumo de café” tiene un impacto significativo, sin embargo, nuestra variable no lo tiene. Esto se puede deber a que el estudio de Zheng et al. (2023) analiza a los estudiantes universitarios que tengan entre 18 y 25 años de Asia Sur y China mientras que nosotros nos concentramos en los estudiantes universitarios de la universidad del pacífico sin limitar la edad.

La variable “horas diarias promedio de uso del celular” tiene un aporte significativo tanto en el estudio de Uzunçakmak, T (2022), que analiza a los estudiantes de enfermería. Por otro lado, nuestro estudio analiza a estudiantes de ingeniería, económica, derecho, administración, contabilidad, marketing, negocios internacionales y finanzas, por lo que la discrepancia se puede deber a los diferentes perfiles que existen entre los estudiantes de las diferentes carreras.

Implicancias

Una de las variables que mostró significancia en la regresión lineal múltiple es la edad. El modelo mostró una relación negativa entre la edad y la cantidad promedio diaria de horas de sueño de los estudiantes. Esta relación puede estar explicada en primer lugar por factores netamente biológicos que hacen que los humanos duermen menos horas mientras aumenta su edad. Por otro lado, esta relación inversa también puede estar explicada por el aumento de responsabilidades que conlleva la edad en los alumnos. Ante esta segunda razón si se podría aplicar una solución que consistirá en la promoción de talleres que enseñen a los alumnos herramientas para poder manejar los tiempos y responsabilidades de una mejor manera. Esta solución se podría implementar en forma de un EFE (Experiencia Formativa Extraacadémica)

disponible para los alumnos.

En el caso de la variable “Tiempo de viaje de la universidad a su hogar”, la universidad suele hacer charlas con el área de Bienestar y Formación Estudiantil (BFE). En estas charlas se suelen tocar temas relacionados a los alumnos y su bienestar y como menciona Chen et al. (2014) el sueño tiene un papel fundamental en la salud de las personas y nuestro trabajo estudia los factores que más afectan a la cantidad de sueño de los estudiantes. Es decir, el BFE con esta información podría realizar charlas sobre los beneficios de dormir una cantidad de horas adecuada y promover entre los alumnos el equilibrio entre la vida académica y otras actividades igualmente importantes como el tiempo libre y las relaciones interpersonales. Sumado a esto, la universidad debería proveer a los estudiantes de espacios en la universidad para que puedan mantener este balance en su día a día complementando los cubículos de estudio con zonas como parques y ambientes de descanso. El BFE también podría promover la investigación en temas de sueño y salud mental así como sugerir la inclusión de un EFE que trate temas relacionados con el bienestar de los alumnos. Esta sugerencia resulta especialmente importante para las carreras de la Facultad de Derecho según la información revelada por el modelo.

Una medida aún mejor sería discriminar a los estudiantes según cuanto se demoran en llegar a la universidad desde sus casas, ya que es la única variable que tiene un impacto significativo tanto en el modelo de regresión lineal múltiple como en el modelo de regresión logística. Otra medida para poder contrarrestar los efectos de esta variable en la cantidad de horas de sueño sería proporcionar un servicio de transporte para los estudiantes en especial los que incurren en tiempos altos para que puedan reducir su tiempo de viaje y aumentar sus horas de sueño promedio al día.

CONCLUSIONES

El sueño tiene un papel muy importante en el desarrollo tanto mental como académico en los estudiantes universitarios. Sin embargo, muchos alumnos universitarios no cumplen con este estándar, y además reportan tener problemas de sueño. Por lo tanto, es necesario poder evaluar cuáles son los factores más importantes que influyen en que los estudiantes duerman las horas de sueño necesarias para un correcto desempeño, considerando que son de 7 a 9 horas según el estudio de Hirshkowitz et al. (2015).

En el presente estudio de investigación, se llevó a cabo una minuciosa evaluación del modelo de regresión logística con el propósito de clasificar de forma precisa a la población de estudiantes de la Universidad del Pacífico. Los resultados obtenidos revelaron un rendimiento sumamente satisfactorio tanto en términos de precisión (accuracy) como en el área bajo la curva ROC. El accuracy del modelo, que cuantifica la proporción de predicciones correctas en relación al total de predicciones realizadas, exhibió valores notablemente elevados. Estos resultados indican que el modelo ha logrado tomar decisiones correctas en la gran mayoría de los casos al llevar a cabo la clasificación de los estudiantes. Asimismo, el área bajo la curva

ROC, una métrica de vital importancia para evaluar el desempeño de los modelos de clasificación, también arrojó un valor significativo. Esta métrica provee información acerca de la capacidad del modelo para distinguir entre las clases positivas y negativas a medida que se ajusta el umbral de clasificación. El elevado valor del área bajo la curva ROC obtenido respalda la efectividad del modelo en la tarea de clasificación de la población estudiantil de la Universidad del Pacífico. Estos resultados prometedores sugieren que el modelo de regresión logística tiene el potencial de ser empleado de manera segura y confiable en la clasificación de los estudiantes de la Universidad del Pacífico. No obstante, es importante destacar que estas conclusiones se basan en el conjunto de datos y las variables utilizadas en este estudio en particular. Por ende, se recomienda realizar validaciones adicionales y considerar posibles sesgos o limitaciones antes de implementar el modelo en un entorno práctico y real.

El análisis mediante el modelo de regresión lineal múltiple revela un coeficiente de determinación (R^2) de 47.6%. Esto indica que gran parte de la variabilidad observada en la variable dependiente no puede ser explicada por el modelo. No obstante, estos resultados proporcionan una visión preliminar sobre las variables relevantes que podrían ser consideradas en investigaciones futuras en el contexto de los alumnos de la Universidad del Pacífico. En cuanto a la evaluación del modelo, se observa un error promedio de aproximadamente ± 1.063066 horas de sueño. Este valor resulta considerablemente elevado para afirmar que el modelo es seguro y preciso para su uso en la predicción. Es importante tener en cuenta que la precisión y fiabilidad del modelo deben ser rigurosamente validadas antes de su implementación práctica. Por lo tanto, se sugiere realizar análisis adicionales y ajustes al modelo, así como considerar la incorporación de otras variables relevantes, con el fin de mejorar su capacidad predictiva y reducir el error en la estimación del tiempo de sueño. Esto permitirá garantizar la confiabilidad y utilidad del modelo para futuras aplicaciones en el ámbito de la Universidad del Pacífico.

BIBLIOGRAFÍA

- American College Health Association. National College Health Assessment II: Reference Group Executive Summary Spring 2019. Disponible en: <https://www.acha.org/>
- Ban, D. J., & Lee, T. J. (2001). Sleep duration, subjective sleep disturbances and associated factors among university students in Korea. *Journal of Korean medical science*, 16(4), 475–480. <https://doi.org/10.3346/jkms.2001.16.4.475>
- Carrell, S. E., Maghakian, T., & West, J. E. (2011). A's from Zzzz's? The Causal Effect of School Start Time on the Academic Achievement of Adolescents (Vol. 3, Número 3).
- Carrillo-Mora P., Ramírez-Peris J., Magaña-Vázquez K. (2013) Neurobiología del sueño y su importancia: antología para el estudiante universitario. *Rev Fac Med UNAM*. 2013, 56(4), 5-15.
- Chen, X., Gelaye, B., & Williams, M. A. (2014). Sleep characteristics and health-related quality of life among a national sample of American young adults: assessment of possible health disparities. *Research*, 23(2), 615-627. <https://doi.org/10.1007/sl>
- Chiang, Y. C., Arendt, S., & Sapp, S. (2020). Academic performance, employment, and sleep health: A comparison between working and nonworking students. *International Journal of Higher Education*, 9(3), 202-213. <https://doi.org/10.5430/ijhe.v9n3p202>
- Desjardins, S., & Grandbois, M. (2022). Sleep parameters associated with university students' grade point average and dissatisfaction with academic performance. *Sleep Epidemiology*, 2, 100038. <https://doi.org/10.1016/j.sleepe.2022.100038>
- Dewald, J. F., Meijer, A. M., Oort, F. J., Kerkhof, G. A., & Bögels, S. M. (2010). The influence of sleep quality, sleep duration and sleepiness on school performance in children and adolescents: A meta-analytic review. *Sleep medicine reviews*, 14(3), 179-189.
- Granados-Carrasco Z, Bartra-Aquinaga A, Bendezú-Barnuevo D, Huamanchumo-Merino J, Hurtado-Noblecilla E, Jiménez-Flores J, et al. Calidad del sueño en una facultad de medicina de Lambayeque. *An Fac Med*. 2013; 74(4): 311-4. doi: 10.15381/anales.v74i4.2704
- Hirshkowitz M, et al, National Sleep Foundation's updated sleep duration recommendations: final report, *Sleep Health* (2015), <http://dx.doi.org/10.1016/j.sleh.2015.10.004>
- Izaguirre Corcuera, M. et al (2019). Calidad y su relación con la satisfacción y lealtad del paciente externo en un hospital público de Lima Metropolitana, 2019 (Tesis de maestría). Universidad Peruana Cayetano Heredia. Recuperado de https://repositorio.upch.edu.pe/bitstream/handle/20.500.12866/11963/Calidad_Izaguirr

eCorcuera_Maria.pdf?sequence=1&isAllowed=y

- Joo HJ, Kwon KA, Shin J, Park S, Jang SI. Association between sleep quality and depressive symptoms. *J Affect Disord*. 2022 Aug 1;310:258-265. doi: 10.1016/j.jad.2022.05.004. Epub 2022 May 8. PMID: 35545156.
- Li, L., Wang, L., & Wang, X. (2022). Effect of smartphone use before bedtime on smartphone addiction behaviors among Chinese college students. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.1023245>
- Lu, L., Dong, M., Jian, S. Y., Gao, J., Ye, L. Z., Chen, H. R., Zhang, T. T., Liu, Y. Y., Shen, H. Y., Gai, X. Y., & Liu, S. (2021). Sex differences in the factors associated with sleep duration in university students: A cross-sectional study. *Journal of Affective Disorders*, 290, 345-352. <https://doi.org/10.1016/j.jad.2021.04.025>
- Memoria Participación Universitaria 2020 (2020) Casa UP. Bienestar y Formación Estudiantil (BFE)
- Monterrosa A, Ulloque L, Carriazo S. Calidad del dormir, insomnio y rendimiento académico en estudiantes de medicina. *Duazary*. 2014;11(2):85-97. doi: 10.21676/2389783x.814
- Owens, J. A. (2014). Insufficient sleep in adolescents and young adults: an update on causes and consequences. *Pediatrics*, 134(3), e921-e932
- Quevedo-Blasco, V. J., & Quevedo-Blasco, R. (2011). Influencia del grado de somnolencia, cantidad y calidad de sueño sobre el rendimiento académico en adolescentes. *International Journal of Clinical and Health Psychology*, 11(1), 49-65.
- Rocco, L., Oliari N. (2007). La encuesta mediante internet como alternativa metodológica. VII Jornadas de Sociología. Facultad de Ciencias Sociales, Universidad de Buenos Aires, Buenos Aires.
- Rosales, Edmundo, Egoavil, Martha, La Cruz, Claudia, & Rey de Castro, Jorge. (2007). Somnolencia y calidad del sueño en estudiantes de medicina de una universidad peruana. *Anales de la Facultad de Medicina*, 68(2), 150-158. Recuperado en 15 de mayo de 2023, de http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S1025-55832007000200007&lng=es&tlng=es.
- Universidad del Pacífico (2022). *Alumnos matriculados y egresados pregrado 2022*. https://www.up.edu.pe/Documents/Alumnos%20matriculados%20y%20egresado_Pregrado_2022.pdf

- Universidad del Pacífico. (2023). *Reglamento de Estudios de Pregrado 2023* [Archivo PDF]. Recuperado de <https://sar.up.edu.pe/wp-content/uploads/Reglamento-de-Estudios-de-Pregrado-2023.pdf>
- Uzunçakmak, T., Ayaz-Alkaya, S., & Akca, A. (2022). Prevalence and predisposing factors of smartphone addiction, sleep quality and daytime sleepiness of nursing students: A cross-sectional design. *Nurse Education in Practice*, 65. <https://doi.org/10.1016/j.nepr.2022.103478>
- Valenzuela B, Alfonso. (2010). EL CAFÉ Y SUS EFECTOS EN LA SALUD CARDIOVASCULAR Y EN LA SALUD MATERNA. *Revista chilena de nutrición*, 37(4), 514-523. <https://dx.doi.org/10.4067/S0717-75182010000400013>
- Vázquez-Chacón, Mariclaudia, Cabrejos-Llontop, Sarai, Yrigoin-Pérez, Yamily, Robles-Alfaro, Ricardo, & Toro-Huamanchumo, Carlos J.. (2019). Adicción a internet y calidad de sueño en estudiantes de medicina de una Universidad peruana, 2016. *Revista Habanera de Ciencias Médicas*, 18(5), 817-830. Epub 06 de octubre de 2019. Recuperado en 15 de mayo de 2023, de http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1729-519X2019000500817&lng=es&tlng=es.
- Wang, F., & Bíró, É. (2021). Determinants of sleep quality in college students: A literature review. *EXPLORE*, 17(2), 170-177. <https://doi.org/10.1016/j.explore.2020.11.003>
- Zheng, W., Chen, Q., Yao, L., Zhuang, J., Huang, J., Hu, Y., Yu, S., Chen, T., Wei, L., Wang, H., & Fan Sr, C. (2023). Prediction models for sleep quality among college students during the COVID-19 outbreak: a cross-sectional study based on Internet New Media (Preprint). *Journal of Medical Internet Research*. <https://doi.org/10.2196/45721>

Apéndice 1:

A continuación se presenta el código en Python que representa el procesamiento de los datos. En la sección "Parada 1", se realiza la recodificación de las variables nominales "Si" y "No" a valores binarios de uno y cero. En la sección "Parada 2", se selecciona inicialmente la columna del dataframe que contiene la variable para la cual se requieren variables dummy. A continuación, se elige la categoría de referencia y se crean las tres variables dummy con el formato siguiente: "f 'Facultad_de_{referencia_categoria}' ". Finalmente, se elimina la columna que contiene la variable original. En la sección "Parada 3", se crea la variable "Logística", la cual será utilizada para realizar la regresión logística.

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

# Leer el archivo CSV
df = pd.read_csv('D:/csv/Sueño.csv', encoding='latin1', delimiter=',')
df_Analisis = df

# Parada 1
df['¿Cuál es tu sexo?'] = df['¿Cuál es tu sexo?'].map({'Hombre': 0, 'Mujer': 1})
df['¿Consumes cigarrillos?'] = df['¿Consumes cigarrillos?'].map({'Si': 0, 'No': 1, 'Sí': 0, 'No': 1})
df['¿Tienes empleo?'] = df['¿Tienes empleo?'].map({'Si': 0, 'No': 1, 'Sí': 0, 'No': 1})
df['¿Eres miembro activo de alguna organización estudiantil?'] = df['¿Eres miembro activo de alguna
organización estudiantil?'].map({'Sí': 0, 'No': 1, np.nan: 0})

# Parada 2
columna7_dummy = pd.get_dummies(df['¿A qué facultad perteneces?'], prefix='Facultad_de')
referencia_categoria = 'Derecho'
columna7_dummy.drop(f'Facultad_de_{referencia_categoria}', axis=1, inplace=True)
columna7_dummy = columna7_dummy.astype(int)
df = pd.concat([df, columna7_dummy], axis=1)
df.drop('¿A qué facultad perteneces?', axis=1, inplace=True)

# Parada 3
df['logistica'] = np.where(df.iloc[:, 0] >= 7, 1, 0)

# Guardar en CSV
ruta_del_archivo = 'D:/csv/Sueno_preprocesado.csv'
df.dropna().to_csv(ruta_del_archivo, index=False, encoding='utf-8-sig')
```

Apéndice 2

A continuación se muestra el código en Python para crear el mapa de correlaciones, que se presentó anteriormente. En la sección "Parada 1", se define la variable corr_matrix. Luego, en la sección "Parada 2", se recorre esta matriz para establecer los valores y, finalmente, se grafica el mapa de correlaciones utilizando la función plot.

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
# Configurar estilo
plt.style.use('seaborn-white')
# Leer el archivo CSV
```

```

df = pd.read_csv('D:/csv/Sueño.csv', encoding='latin1', delimiter=',')
# Seleccionar las columnas de interés
cols = [1, 3, 6, 9, 11]
df_sel = df.iloc[:, cols]

#Parada 1
corr_matrix = df_sel.corr()
print(corr_matrix)
# Crear un mapa de correlaciones con valores
fig, ax = plt.subplots(figsize=(8, 6))
im = ax.imshow(corr_matrix, cmap='coolwarm', interpolation='nearest')
fig.colorbar(im, fraction=0.046, pad=0.04)
# Parada 2
for i in range(len(cols)):
    for j in range(len(cols)):
        text = ax.text(j, i, f'{corr_matrix.iloc[i, j]:.2f}', ha='center', va='center', color='w')
# Establecer los ticks y etiquetas de los ejes
var_labels = ['Edad', 'Créditos en el ciclo', 'Consumo de café', 'Horas diarias de uso de celular',
              'Tiempo promedio de viaje a la universidad']
ax.set_xticks(np.arange(len(cols)))
ax.set_yticks(np.arange(len(cols)))
ax.set_xticklabels(var_labels, rotation=45, ha='right')
ax.set_yticklabels(var_labels)
# Ajustar la posición de los textos
plt.tight_layout()
# Ajustar la calidad del gráfico
plt.savefig('mapa_correlaciones.png', dpi=300)
plt.show()

```

Apéndice 3

En el siguiente código de R, lo primero que se hace es establecer las categorías, segunda línea del código. En la parada 1 se va a iterar dentro del csv de los datos originales y se van a extraer índices de manera aleatoria siempre y cuando se cumpla con la condición de num_obs, esta condición se basa en la tabla número 10. En la parada 2 se lee el csv Sueno_preprocesado.csv, este csv ya tiene las variables dummies codificadas, a diferencia del archivo “Sueño.csv”. En la parada 3 se extraen del archivo “Sueno_preprocesado.csv” las observaciones definidas para el training y testing para luego crear 2 archivos diferentes, uno con la data de training y otro con la data de testing.

```

datos_sueno = read.csv("D:/csv/Sueño.csv", sep = ";")
categorias = c("Ciencias Empresariales", "Economía y Finanzas", "Ingeniería", "Derecho")
num_obs = c(23, 20, 11, 4)
train_indices = c()
test_indices = c()

#Parada 1
for (i in 1:length(categorias)) {
  categoria = categorias[i]
  n = num_obs[i]

  indices = which(grepl(categoria, datos_sueno[, 9]))

  if (length(indices) >= n) {
    train_indices_cat = sample(indices, n)

    train_indices = c(train_indices, train_indices_cat)
  }
}

```

```

test_indices_cat = setdiff(indices, train_indices_cat)

test_indices = c(test_indices, test_indices_cat)
} else {
  cat("No hay suficientes observaciones para la categoría", categoria, "\n")
}
}
#Parada 2
datos_preprocesados = read.csv("D:/csv/Sueno_preprocesado.csv")
#Parada 3
train_data = datos_preprocesados[train_indices, ]
test_data = datos_preprocesados[test_indices, ]

colnames(train_data) = c("Horas_Sueno", "Edad", "Sexo", "Creditos_Ciclo", "Consumo_Cafe", "Consumo_Cigarro",
  "Tiempo_Celular", "Empleo", "Tiempo_Casa", "Orga_Est", "Actividad_Fisica", "Facu_Ciencias",
  "Facu_Eco", "Facu_Ing", "Log")
colnames(test_data) = c("Horas_Sueno", "Edad", "Sexo", "Creditos_Ciclo", "Consumo_Cafe", "Consumo_Cigarro",
  "Tiempo_Celular", "Empleo", "Tiempo_Casa", "Orga_Est", "Actividad_Fisica", "Facu_Ciencias",
  "Facu_Eco", "Facu_Ing", "Log")

write.csv(train_data, file = "D:/csv/train_data.csv", row.names = FALSE, fileEncoding = "UTF-8")
write.csv(test_data, file = "D:/csv/test_data.csv", row.names = FALSE, fileEncoding = "UTF-8")

```

Apéndice 4

El siguiente código de R hace el testeo de la regresión lineal múltiple y calcula el RMSE, para ello primero creó el modelo de regresión lineal múltiple de entrenamiento, los valores concuerdan con el SPSS.

```
train_data = read.csv("D:/csv/train_data.csv", fileEncoding = "UTF-8")
```

```
Y = train_data[, 1]
```

```
df = train_data[, -c(1, ncol(train_data))]
```

```
lm_model = lm(Y ~ ., data = df)
```

```
print(summary(lm_model))
```

```
test_data = read.csv("D:/csv/test_data.csv", fileEncoding = "UTF-8")
```

```
df_test = test_data[, -c(1, ncol(test_data))]
```

```
y_predicho = predict(lm_model, newdata = df_test)
```

```
comparison_matrix = data.frame(y_predicho, y_real = test_data[, 1])
```

```
print(comparison_matrix)
```

```
rmse = sqrt(mean((y_predicho - test_data[, 1])^2))
```

```
print(rmse)
```

Apéndice 5

El siguiente código de R testea el modelo de regresión logística, al igual que para la regresión lineal múltiple, los resultados del entrenamiento fueron corroborados con los resultados obtenidos del SPSS. Luego de crear el modelo con la data de entrenamiento, va a utilizar los coeficientes del modelo con la data de testing y con esta data crear la matriz de confusión y graficar la curva ROC de testing, se encuentra en el Anexo 12.

```
library(pROC)
```

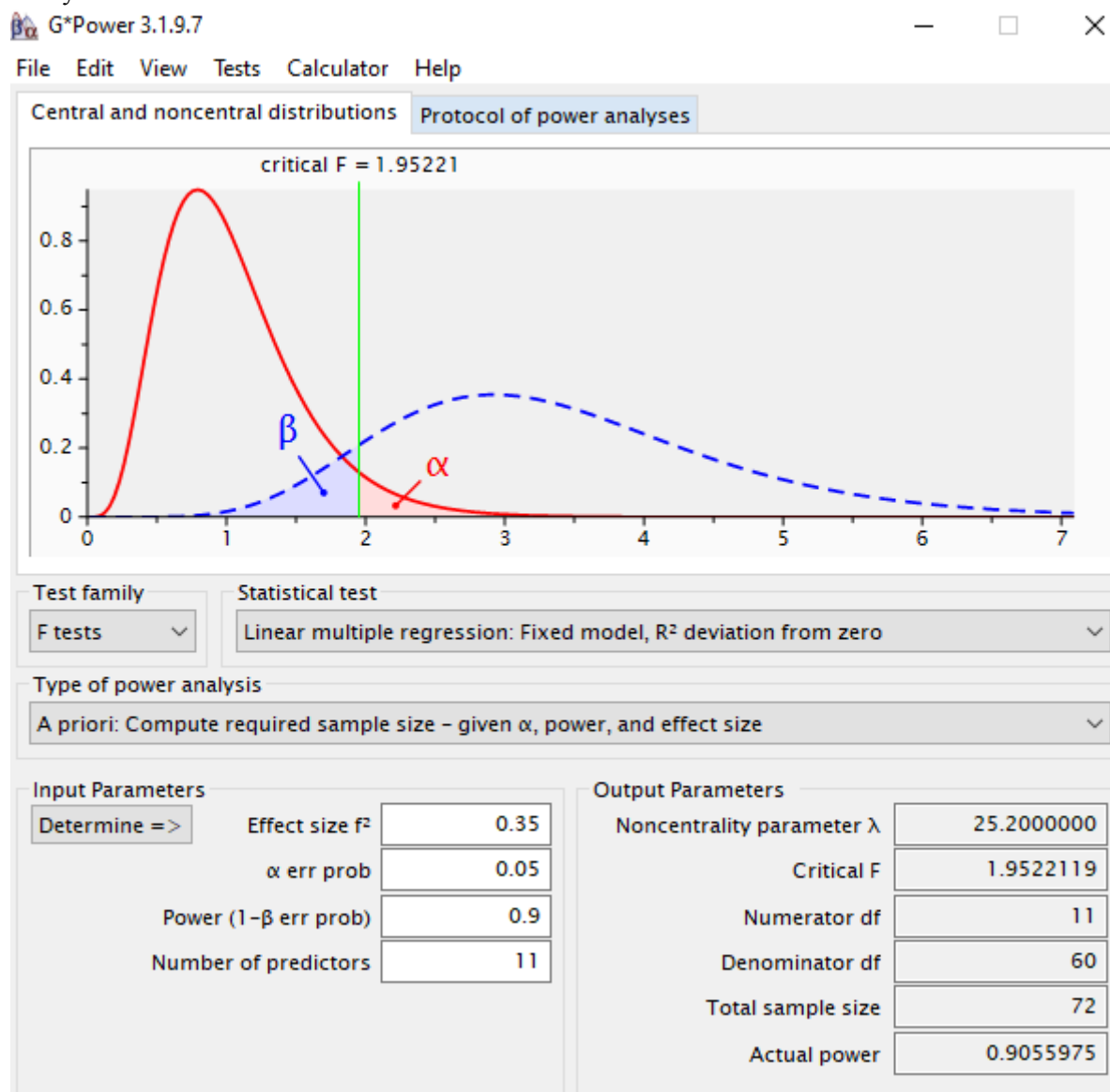
```

train_data <- read.csv("D:/csv/train_data.csv", fileEncoding = "UTF-8")
train_data <- train_data[, -1]
Y <- train_data[, ncol(train_data)]
X <- train_data[, -ncol(train_data)]
logit_model <- glm(Y ~ ., data = X, family = binomial)
train_prob <- predict(logit_model, type = "response")
roc_obj <- roc(Y, train_prob)
roc_auc <- auc(roc_obj)
print(roc_auc)
plot(roc_obj, col = "blue", lwd = 2)
test_data <- read.csv("D:/csv/test_data.csv", fileEncoding = "UTF-8")
test_data <- test_data[, -1]
test_prob <- predict(logit_model, newdata = test_data, type = "response")
test_predictions <- ifelse(test_prob > 0.5, 1, 0)
confusion_matrix <- table(test_data[, ncol(test_data)], test_predictions)
print(confusion_matrix)
roc_obj_test <- roc(test_data[, ncol(test_data)], test_prob)
print(roc_obj_test$auc)
plot(roc_obj_test, col = "blue", lwd = 2)
abline(a = 1, b = -1, col = "red", lwd = 2)

```


ANEXO 1:

Imagen del interfaz del software G-Power mediante la cual establecimos el tamaño de muestra mínimo que debíamos incluir en el trabajo. Aquí se puede apreciar los parámetros que introducimos y los parámetros de salida, los cuales incluyen al tamaño de muestra.



ANEXO 2:

Resultado del SPSS del resumen del modelo de regresión lineal múltiple donde se puede ver el R^2 y el R^2 - ajustado.

Resumen del modelo^b

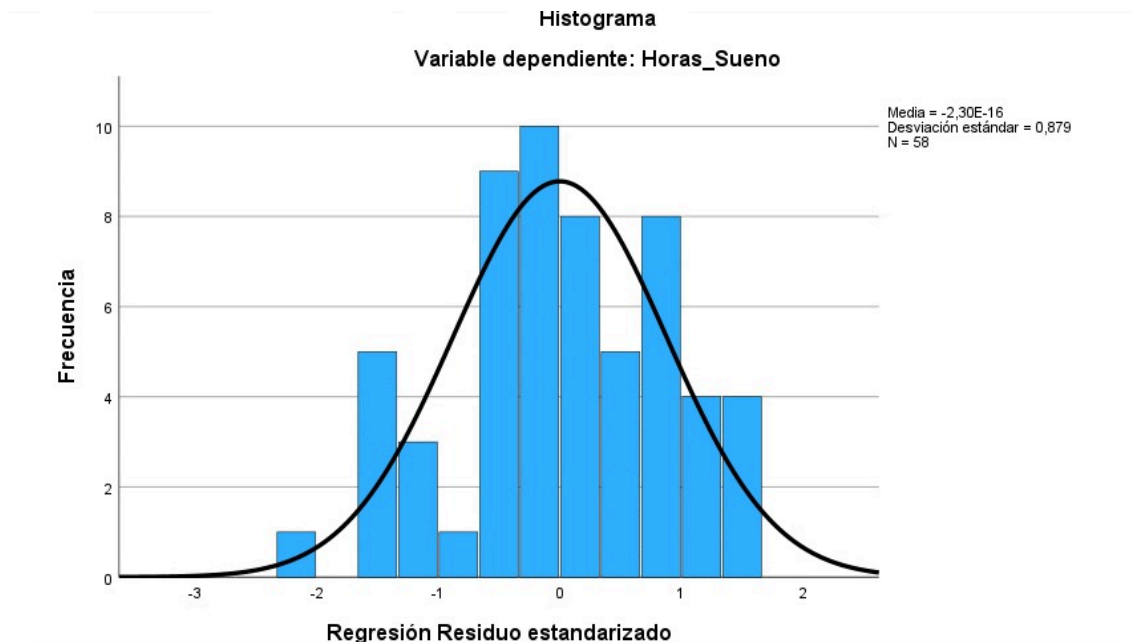
Modelo	R	R cuadrado	R cuadrado ajustado	Error estándar de la estimación	Durbin-Watson
1	,690 ^a	,476	,321	,8733	2,357

a. Predictores: (Constante), Facu_Ing, Creditos_Ciclo, Empleo, Tiempo_Celular, Orga_Est, Actividad_Fisica, Consumo_Cigarro, Sexo, Tiempo_Casa, Edad, Consumo_Cafe, Facu_Ciencias, Facu_Eco

b. Variable dependiente: Horas_Sueno

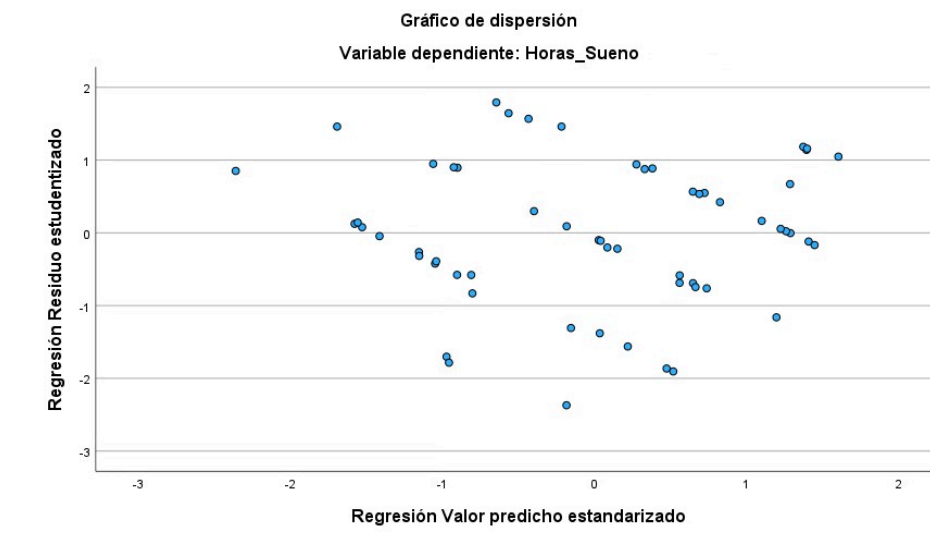
ANEXO 3:

Histograma de la prueba de normalidad de los errores.



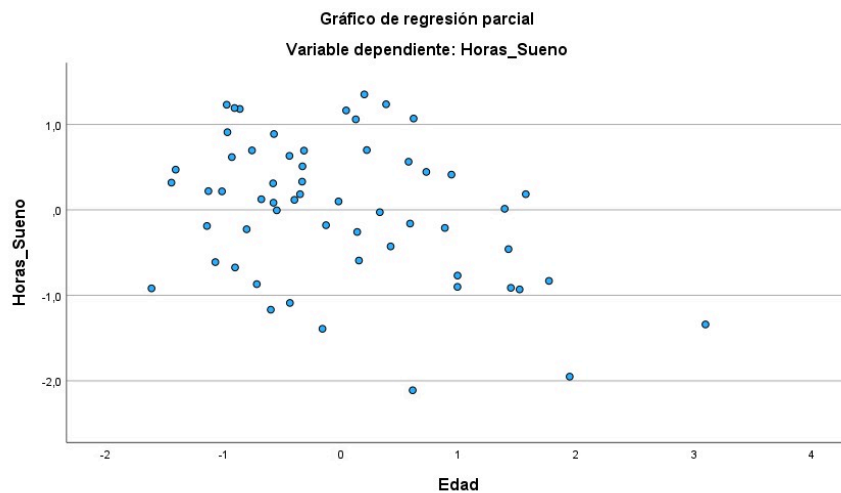
ANEXO 4:

Gráfico para la prueba de homocedasticidad, la cual demuestra si la varianza de los errores es constante a lo largo del tiempo



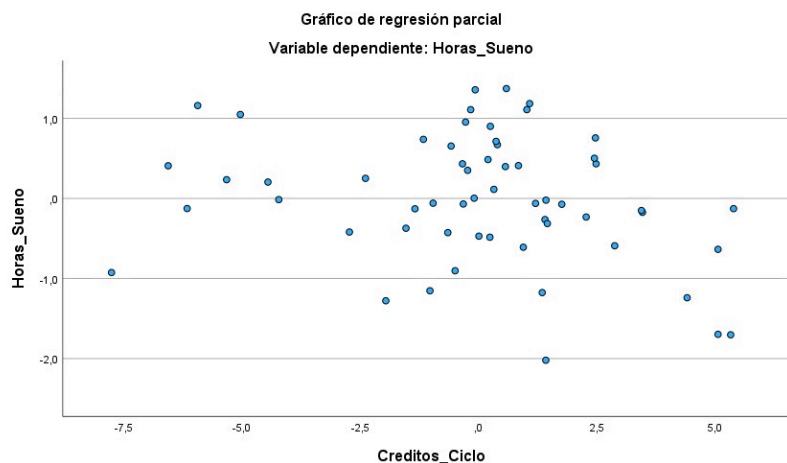
ANEXO 5:

Gráfico del supuesto de linealidad de la variable Edad con respecto a las Horas de sueño.



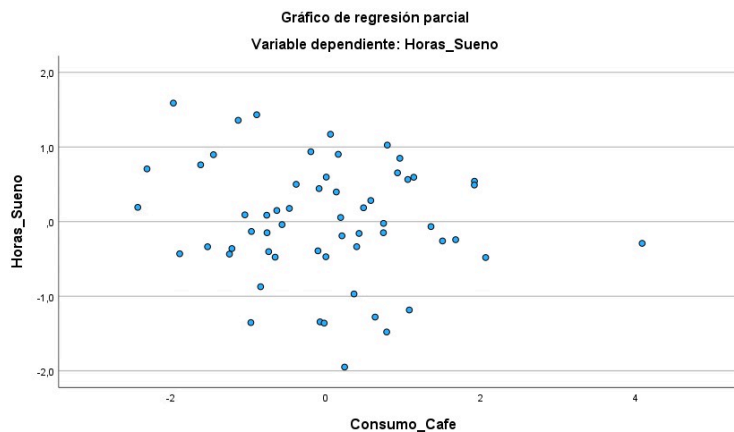
ANEXO 6:

Gráfico del supuesto de linealidad de la variable Créditos en el ciclo con respecto a las Horas de sueño.



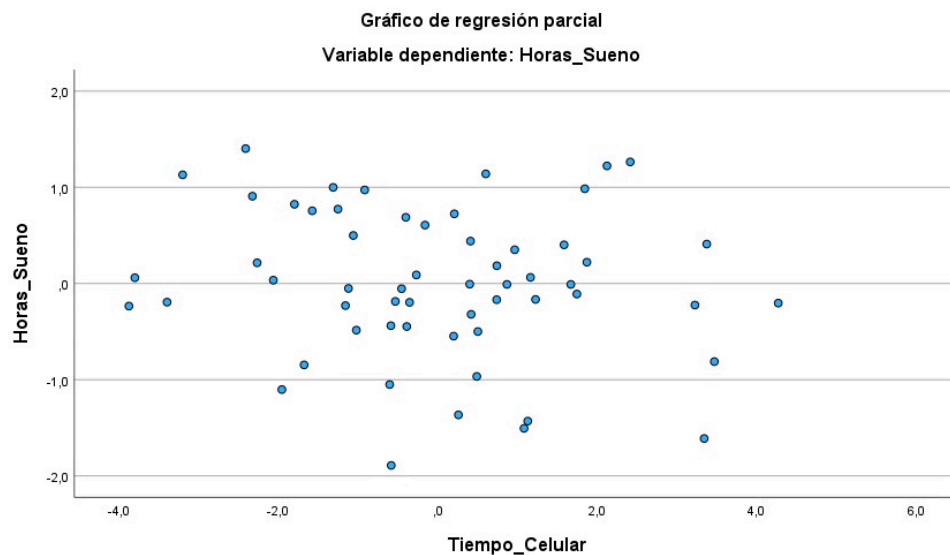
ANEXO 7:

Gráfico del supuesto de linealidad de la variable Consumo de café con respecto a las Horas de sueño.



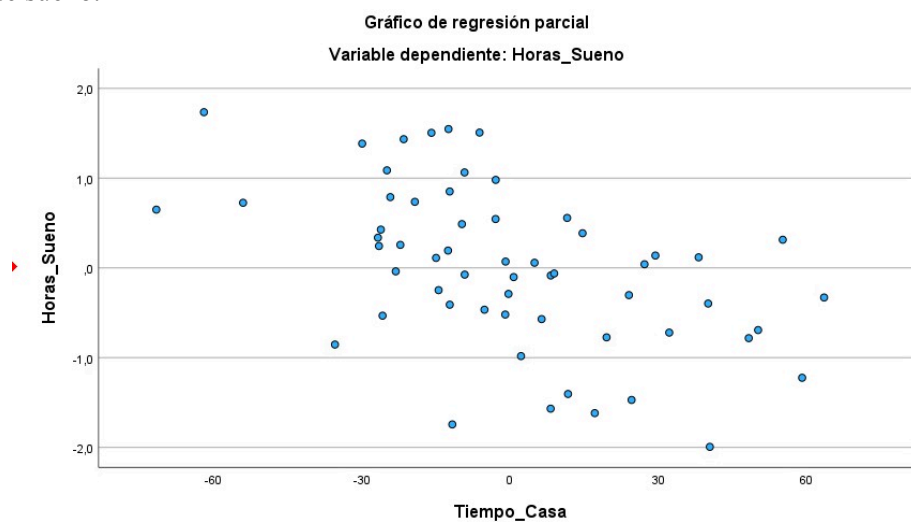
ANEXO 8:

Gráfico del supuesto de linealidad de la variable Horas diarias de uso del celular con respecto a las Horas de sueño.



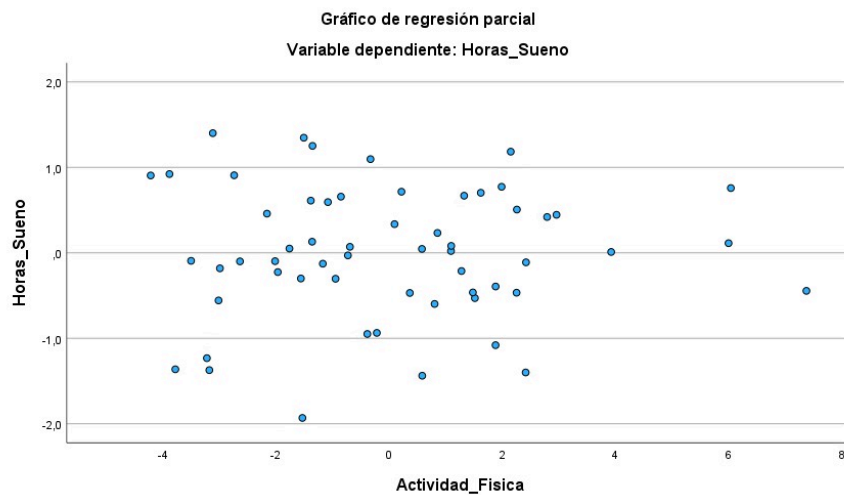
ANEXO 9:

Gráfico del supuesto de linealidad de la variable Tiempo promedio de viaje hasta la universidad con respecto a las Horas de sueño.



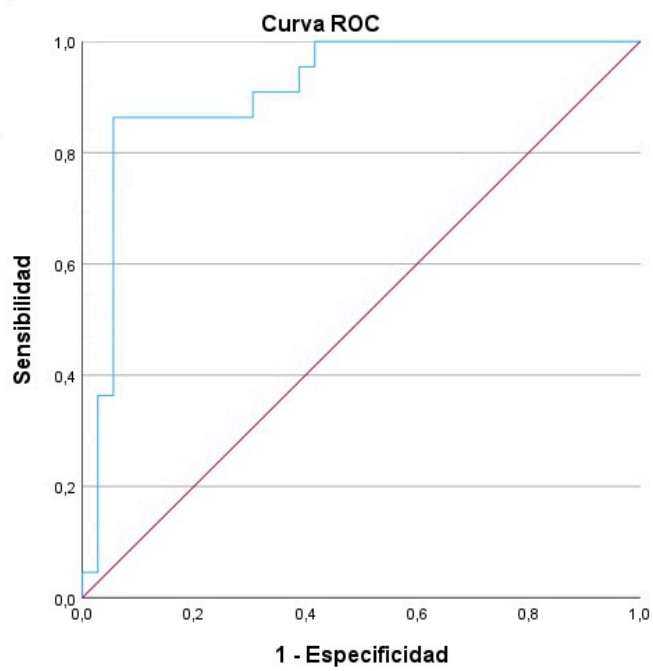
ANEXO 10:

Gráfico del supuesto de linealidad de la variable Actividad física con respecto a las Horas de sueño.



ANEXO 11:

Gráfico de la curva ROC del training del modelo de regresión logística.



ANEXO 12:

Gráfico de la curva ROC del testing del modelo de regresión logística.

