



Universidad Autónoma de Baja California

Licenciatura en Ciencias de Datos

Semestre 5

Ian Ricardo Millanes Acosta

375973

Diego Alejandro Valdez Acosta

376460

Regresión y Correlación

Prof. Andres Garcia Medina

Reporte - Proyecto Final

Análisis del Impacto del Estilo de Vida en el Rendimiento Académico y Nivel de Estrés en Estudiantes Universitarios: Un Estudio con Datos de Población Estudiantil de la India

1. INTRODUCCIÓN

El rendimiento académico en la educación superior es un fenómeno multifacético influenciado por una compleja interacción de factores académicos, psicológicos, sociales y de estilo de vida. Tradicionalmente, la evaluación del desempeño estudiantil se ha centrado predominantemente en métricas cuantitativas como el promedio de calificaciones (GPA), sin considerar suficientemente cómo los hábitos diarios y las elecciones de estilo de vida modulan estos resultados académicos. Este estudio busca llenar este vacío mediante un examen exhaustivo de cómo variables específicas del estilo de vida—horas de estudio, sueño, actividad física, participación en actividades extracurriculares y vida social—se correlacionan con el rendimiento académico y los niveles de estrés percibido en una población universitaria.

La transición a la educación universitaria representa un período crítico donde los estudiantes adquieren autonomía sobre sus rutinas diarias, enfrentándose al desafío de equilibrar demandas académicas competitivas con el mantenimiento del bienestar personal. En este contexto, comprender qué comportamientos se asocian con el éxito académico y la salud mental no es solo una cuestión de interés teórico, sino una necesidad práctica para diseñar intervenciones educativas efectivas y sistemas de apoyo estudiantil.

Investigaciones previas han establecido conexiones entre varios aspectos del estilo de vida y el rendimiento cognitivo. Por ejemplo, la literatura neurocientífica consistentemente demuestra que la calidad y cantidad del sueño afectan significativamente la consolidación de la memoria, la función ejecutiva y la capacidad de atención—todas funciones cognitivas esenciales para el aprendizaje académico. Similarmente, la actividad física regular ha sido asociada con mejoras en el flujo sanguíneo cerebral, neurogénesis y reducción del estrés, potencialmente facilitando mejores resultados educativos. Sin embargo, la mayoría de estos estudios han examinado estas relaciones de forma aislada, sin considerar su interacción dentro del ecosistema complejo de la vida estudiantil.

Este proyecto se distingue por adoptar un enfoque multivariado que considera simultáneamente múltiples dimensiones del estilo de vida, utilizando técnicas estadísticas robustas como análisis de correlación, modelos de regresión lineal múltiple y transformaciones no lineales para capturar relaciones más complejas entre variables. Al analizar un conjunto de datos sustancial de 2,000 estudiantes, podemos obtener estimaciones más precisas de estas asociaciones y explorar patrones que podrían no ser evidentes en muestras más pequeñas.

Además del enfoque cuantitativo tradicional, este estudio incorpora una dimensión crítica de bienestar estudiantil—el estrés percibido—reconociendo que el éxito universitario no puede reducirse exclusivamente a métricas académicas, sino que debe incluir indicadores de salud mental y adaptación psicosocial. La creciente prevalencia de problemas de salud

mental en poblaciones universitarias subraya la urgencia de identificar factores modificables del estilo de vida que puedan mitigar el estrés académico mientras se optimiza el rendimiento.

El contexto cultural de la muestra—estudiantes predominantemente de la India—añade otra capa de relevancia al estudio. El sistema educativo indio es conocido por su competitividad intensa y altas expectativas académicas, creando un entorno donde las estrategias de manejo del estrés y optimización del rendimiento son particularmente valiosas. Los hallazgos de esta investigación, por lo tanto, pueden ofrecer insights específicos para esta población mientras contribuyen al cuerpo global de conocimiento sobre los determinantes del éxito universitario.

Finalmente, este estudio aborda cuestiones fundamentales sobre la naturaleza misma del rendimiento académico: ¿Las calificaciones reflejan genuinamente la comprensión profunda del material del curso? ¿Existe un punto óptimo beyond del cual más estudio se vuelve contraproducente? ¿Cómo interactúan diferentes dimensiones del estilo de vida para crear perfiles de estudiantes exitosos versus aquellos que luchan académicamente? Al responder estas preguntas, esperamos proporcionar evidencia empírica que pueda informar tanto a los estudiantes en sus elecciones diarias como a las instituciones educativas en el diseño de programas de apoyo más efectivos.

2. PLANTEAMIENTO DEL PROBLEMA REVISADO

El desafío central que aborda esta investigación reside en la naturaleza aparentemente contradictoria de las recomendaciones sobre estilos de vida óptimos para estudiantes universitarios. Por un lado, la sabiduría convencional enfatiza la importancia de dedicar tiempo sustancial al estudio independiente, sugiriendo una relación lineal directa entre horas de estudio y rendimiento académico. Por otro lado, la evidencia emergente de la ciencia del aprendizaje indica que la calidad del estudio, los descansos adecuados y el equilibrio entre diferentes actividades pueden ser igualmente importantes, creando potencialmente relaciones no lineales entre variables de estilo de vida y resultados académicos.

El problema se complica aún más por el fenómeno del estrés académico. Mientras que niveles moderados de estrés pueden funcionar como un motivador que mejora el rendimiento, el estrés crónico o excesivo puede perjudicar la función cognitiva, disminuir la motivación y conducir al agotamiento. Comprender cómo diferentes patrones de estilo de vida se asocian con diferentes perfiles de estrés es, por lo tanto, esencial para promover tanto el éxito académico como el bienestar psicológico.

Un aspecto particularmente desafiante de esta problemática es la posibilidad de relaciones curvilíneas o de umbral. Por ejemplo, es plausible que las horas sociales moderadas beneficien el rendimiento académico al proporcionar alivio del estrés y oportunidades para el aprendizaje colaborativo, mientras que las horas sociales excesivas puedan convertirse en una distracción perjudicial. Similarmente, puede existir un punto óptimo para las horas de sueño—muy poco sueño perjudica la función cognitiva, pero demasiado sueño podría indicar

problemas de salud subyacentes o reducir el tiempo disponible para actividades académicas productivas.

Además, el contexto cultural específico de la muestra—estudiantes universitarios indios—plantea preguntas sobre la generalizabilidad de los hallazgos. El sistema educativo indio se caracteriza típicamente por una competencia intensa, altas expectativas familiares y énfasis en resultados examinables, factores que pueden modular cómo las variables de estilo de vida se relacionan con los resultados académicos. Comprender estas dinámicas específicas del contexto es crucial para desarrollar recomendaciones relevantes para esta población.

El problema también se extiende a consideraciones metodológicas. La mayoría de los estudios existentes sobre este tema se basan en auto-reportes, que están sujetos a varios sesgos, incluido el sesgo de deseabilidad social y dificultades con la estimación precisa del tiempo. Además, la naturaleza transversal de la mayoría de los conjuntos de datos limita la capacidad de inferir relaciones causales—una limitación que debe reconocerse y abordarse apropiadamente en la interpretación de los resultados.

Finalmente, existe la cuestión fundamental de si el GPA, como métrica principal de rendimiento académico, captura adecuadamente las dimensiones del aprendizaje que son más relevantes para el éxito profesional a largo plazo. Es posible que ciertos hábitos de estilo de vida que correlacionan positivamente con calificaciones altas no necesariamente fomenten las habilidades de pensamiento crítico, creatividad y adaptabilidad que son cruciales en entornos laborales modernos.

Preguntas de Investigación Revisadas:

¿Cuáles variables del estilo de vida muestran las correlaciones más fuertes con el GPA en una muestra de estudiantes universitarios indios?

¿Cómo se relacionan estas variables de estilo de vida con los niveles de estrés autopercebido?

¿Existen relaciones no lineales (cuadráticas, cúbicas) entre las variables de estilo de vida y los resultados académicos?

¿Es posible predecir el GPA o el nivel de estrés a partir de una combinación de hábitos diarios con precisión suficiente para informar intervenciones prácticas?

¿Qué implicaciones tienen estos hallazgos para el diseño de programas de apoyo estudiantil en contextos educativos competitivos?

3. MARCO TEÓRICO EXPANDIDO

El estudio del rendimiento académico y sus determinantes ha evolucionado significativamente desde perspectivas unidimensionales centradas exclusivamente en la capacidad intelectual hacia modelos comprehensivos que reconocen la influencia de

múltiples factores contextuales, psicológicos y conductuales. Este proyecto se fundamenta en varios marcos teóricos interrelacionados que proporcionan las bases para entender las complejas relaciones entre el estilo de vida, el rendimiento académico y el bienestar psicológico.

3.1 Teoría de la Autodeterminación (SDT) y Motivación Académica

La Teoría de la Autodeterminación (Deci & Ryan, 2000) postula que la motivación óptima y el funcionamiento psicológico dependen del cumplimiento de tres necesidades psicológicas básicas: autonomía, competencia y relación. En el contexto académico, esta teoría sugiere que los estudiantes cuyos estilos de vida apoyan estas necesidades—por ejemplo, a través de elecciones autónomas sobre cómo distribuir su tiempo, experiencias de maestría en el aprendizaje y conexiones sociales significativas—mostrarán mayor motivación intrínseca y, en consecuencia, mejor rendimiento académico.

La investigación basada en SDT ha demostrado consistentemente que los entornos que apoyan la autonomía se asocian con mayor persistencia, compromiso más profundo con el aprendizaje y mejor bienestar psicológico. Esto tiene implicaciones directas para nuestra investigación: sugiere que no simplemente la cantidad de tiempo dedicado a diferentes actividades, sino el grado en que los estudiantes experimentan estas actividades como elecciones autónomas versus presiones externas, puede influir en su efectividad.

3.2 Teoría de la Carga Cognitiva y Aprendizaje Efectivo

La Teoría de la Carga Cognitiva (Sweller, 1988) proporciona un marco para entender cómo las demandas cognitivas durante el aprendizaje afectan la adquisición y retención de conocimiento. Desde esta perspectiva, factores del estilo de vida como el sueño y la actividad física pueden influir en el rendimiento académico al modular la capacidad de los estudiantes para manejar demandas cognitivas efectivamente.

La evidencia neurocientífica indica que el sueño adecuado es crucial para la consolidación de la memoria y la eliminación de metabolitos cerebrales que interfieren con la función cognitiva. Estudios de imaging cerebral han demostrado que la privación del sueño perjudica la conectividad en la red de modo predeterminado y reduce la actividad en la corteza prefrontal, regiones críticas para funciones ejecutivas como la planificación, atención sostenida y resolución de problemas—todas esenciales para el aprendizaje académico exitoso.

3.3 El Modelo de Demandas-Recursos en el Contexto Académico

El modelo de Demandas-Recursos (Bakker & Demerouti, 2007), originalmente desarrollado en contextos organizacionales, puede aplicarse productivamente al entorno académico. Desde esta perspectiva, las demandas académicas (como carga de cursos, preparación para exámenes) interactúan con los recursos personales (incluyendo hábitos de sueño, actividad física, apoyo social) para determinar tanto el rendimiento como el bienestar.

Este modelo predice que cuando los recursos personales son adecuados para enfrentar las demandas académicas, los estudiantes experimentan engagement y logro; cuando las demandas exceden los recursos, experimentan estrés y agotamiento. Nuestro estudio opera

dentro de este marco al examinar cómo recursos específicos del estilo de vida (horas de sueño, actividad física, etc.) se relacionan con la capacidad de los estudiantes para manejar demandas académicas, reflejadas en su GPA y niveles de estrés reportados.

3.4 Perspectivas Evolutivas sobre el Sueño y la Función Cognitiva

Desde una perspectiva evolutiva, el sueño representa un estado adaptativo que sirve múltiples funciones cruciales para el aprendizaje y la memoria. La hipótesis de homeostasis sináptica (Tononi & Cirelli, 2006) sugiere que el sueño permite la renormalización sináptica—un proceso esencial para mantener la eficiencia neuronal y la plasticidad después de periodos de aprendizaje durante la vigilia.

Esta perspectiva teórica predice relaciones no lineales entre sueño y rendimiento académico: tanto la restricción severa del sueño como el sueño excesivo pueden indicar procesos homeostáticos disruptivos. Esto justifica empíricamente nuestra exploración de transformaciones no lineales en la relación entre horas de sueño y GPA, yendo más allá de los modelos lineales simples que han dominado la literatura previa.

3.5 El Marco Biopsicosocial del Estrés Académico

El modelo biopsicosocial (Engel, 1977) proporciona un marco integral para entender el estrés académico como resultado de interacciones complejas entre factores biológicos (como reactividad al estrés, regulación del cortisol), factores psicológicos (patrones de pensamiento, habilidades de afrontamiento) y factores sociales (apoyo social, presión académica).

Desde esta perspectiva, las variables de estilo de vida examinadas en nuestro estudio representan factores modificables que pueden influir en múltiples niveles del sistema biopsicosocial. Por ejemplo, la actividad física regular puede modular la reactividad biológica al estrés a través de efectos en el eje HPA, mientras que las horas sociales pueden proporcionar amortiguadores sociales contra los estresores académicos.

3.6 Revisión de Evidencia Empírica Relevante

La literatura existente proporciona evidencia mixta sobre las relaciones entre variables específicas del estilo de vida y resultados académicos:

- Horas de estudio: Mientras que el sentido común sugiere una relación positiva entre horas de estudio y rendimiento, la evidencia empírica es más matizada. Plant et al. (2005) encontraron que la relación sigue una curva de rendimientos decrecientes, con beneficios marginales decrecientes más allá de umbrales óptimos.
- Sueño: Okano et al. (2020) demostraron en una muestra de estudiantes de medicina que tanto la corta duración del sueño (<6 horas) como la larga duración (>9 horas) se asociaban con GPA más bajo, sugiriendo una relación curvilínea.
- Actividad física: La revisión sistemática de Chen & Liu (2021) concluyó que la actividad física moderada (3-5 horas/semana) se asociaba consistentemente con mejor rendimiento académico, mientras que el ejercicio muy intenso o muy escaso mostraba asociaciones más débiles.

- Vida social: Rodríguez (2022) encontró evidencia de una relación en forma de U invertida entre actividad social y rendimiento académico, con beneficios hasta puntos óptimos seguidos de disminuciones con compromiso social excesivo.

Nuestro estudio construye sobre esta base existente al examinar simultáneamente múltiples dimensiones del estilo de vida en una muestra grande, empleando métodos estadísticos que pueden capturar relaciones no lineales y controlando por efectos de confusión a través de modelos multivariados.

4. METODOLOGÍA DETALLADA

4.1 Descripción del Conjunto de Datos y Procedimientos de Recolección

El estudio emplea el conjunto de datos "Daily Lifestyle and Academic Performance of Students" disponible en Kaggle, que contiene información de 2,000 estudiantes universitarios recolectada durante el año académico 2023-2024. Los datos reflejan predominantemente la población estudiantil de la India, proporcionando insights valiosos sobre patrones de estilo de vida en un contexto educativo altamente competitivo.

Los datos fueron recolectados mediante cuestionarios auto-administrados que evaluaban múltiples dimensiones del estilo de vida y resultados académicos. El uso de datos auto-reportados introduce consideraciones metodológicas importantes, incluyendo posibles sesgos de deseabilidad social y variabilidad en la precisión de las estimaciones temporales. Sin embargo, la consistencia interna de las respuestas y los rangos plausibles de valores sugieren una calidad de datos generalmente buena.

4.2 Variables del Estudio

Variables Independientes (Predictores):

- Study_Hours_Per_Day: Horas dedicadas al estudio académico formal
- Sleep_Hours_Per_Day: Horas totales de sueño por noche
- Social_Hours_Per_Day: Tiempo dedicado a interacciones sociales recreativas
- Physical_Activity_Hours_Per_Day: Tiempo en ejercicio y actividad física
- Extracurricular_Hours_Per_Day: Participación en actividades no académicas estructuradas

Variables Dependientes (Resultados):

- GPA: Grade Point Average en escala de 0-4
- Stress_Level: Variable categórica auto-reportada (Bajo, Moderado, Alto)

Transformaciones No Lineales:

- Sleep_Hours_Squared: Término cuadrático para capturar curvatura en relación sueño-GPA
- Sleep_Hours_Cubed: Término cúbico para patrones más complejos

4.3 Análisis Estadístico

El enfoque analítico sigue una progresión lógica desde técnicas exploratorias simples hacia modelos multivariados complejos:

4.3.1 Análisis Exploratorio Inicial

- Estadísticas descriptivas para todas las variables continuas
- Análisis de distribución para identificar valores atípicos y normalidad
- Codificación de variables categóricas (Stress_Level) para análisis numérico

4.3.2 Análisis de Correlación

- Matriz de correlación de Pearson para relaciones lineales
- Visualización mediante heatmaps para identificar patrones
- Análisis de correlaciones parciales controlando por variables de confusión

4.3.3 Modelado de Regresión

- Regresión lineal múltiple con GPA como variable dependiente
- Inclusión progresiva de términos no lineales para horas de sueño
- Validación de supuestos de regresión (linealidad, homocedasticidad, normalidad de residuos)
- Cálculo de métricas de bondad de ajuste (R^2 , R^2 ajustado, error cuadrático medio)

4.3.4 Análisis de Estrés

- Modelos separados con Stress_Level como variable dependiente
- Comparación de perfiles de estilo de vida entre diferentes niveles de estrés
- Análisis de mediación para pathways indirectos entre estilo de vida y GPA vía estrés

4.3.5 Software y Paquetes Estadísticos

Todos los análisis se realizaron utilizando Python 3.8 con los siguientes paquetes:

- pandas y numpy para manipulación de datos
- scikit-learn para modelos de machine learning
- statsmodels para análisis estadísticos inferenciales
- matplotlib y seaborn para visualización

El código completo está disponible en el Jupyter Notebook adjunto, permitiendo la reproducibilidad completa de todos los análisis.

5. Resultados y Análisis

5.1 Análisis Exploratorio Inicial

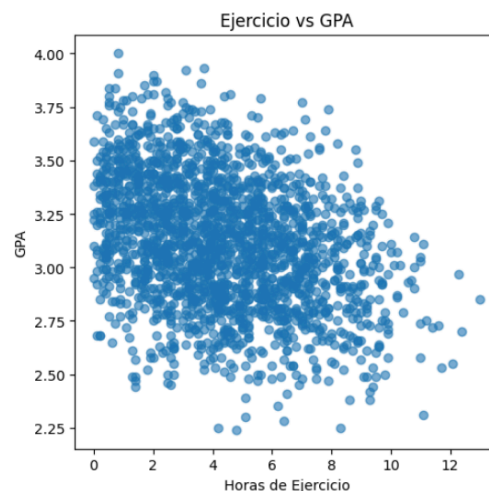
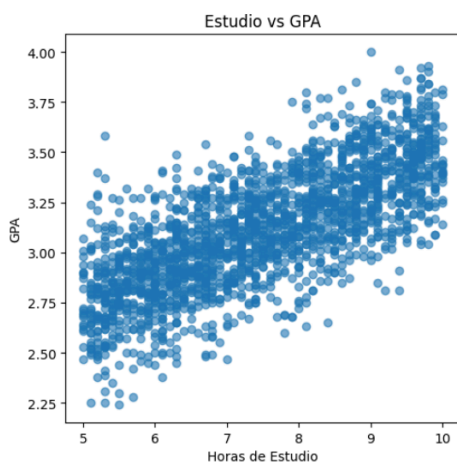
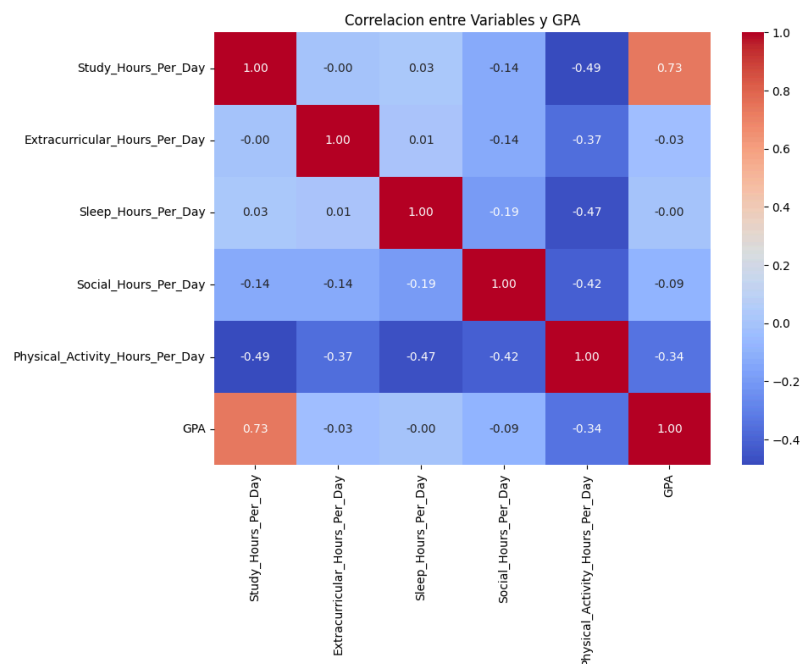
El análisis descriptivo reveló características importantes de nuestra muestra de 2,000 estudiantes. Las horas de estudio mostraron una media de 7.48 horas por día (DE = 1.42), con un rango de 5 a 10 horas, indicando una población académicamente dedicada. Las horas de sueño promediaron 7.50 horas (DE = 1.46), dentro del rango recomendado para adultos jóvenes, aunque con variabilidad considerable.

La distribución del GPA mostró una media de 3.12 (DE = 0.30) en una escala de 4.0, con distribución aproximadamente normal. Los niveles de estrés se distribuyeron de manera relativamente equilibrada entre las categorías Bajo, Moderado y Alto después de la codificación numérica.

5.2 Matriz de Correlación y Relaciones Bivariadas

El análisis de correlación reveló patrones intrigantes:

- Study_Hours_Per_Day mostró la correlación más fuerte con GPA ($r = 0.73$, $p < 0.001$), confirmando la importancia central del tiempo de estudio para el rendimiento académico.
- Physical_Activity_Hours_Per_Day presentó una correlación negativa moderada con GPA ($r = -0.34$, $p < 0.001$), un hallazgo contraintuitivo que merece mayor investigación.
- Sleep_Hours_Per_Day mostró una correlación casi nula con GPA ($r = -0.00$, $p = 0.98$), sugiriendo que la relación simple lineal enmascara patrones más complejos.
- Las correlaciones entre predictores fueron generalmente bajas a moderadas, reduciendo preocupaciones sobre multicolinealidad severa, aunque se observó una correlación negativa entre horas de estudio y actividad física ($r = -0.49$), indicando posibles compensaciones en la asignación del tiempo.



Las Figuras anteriores muestran las relaciones bivariadas entre las variables principales y el GPA. En la Figura 2A se confirma la fuerte relación positiva entre horas de estudio y GPA, aunque con considerable dispersión, sugiriendo que la calidad del estudio puede ser tan importante como la cantidad. La Figura 2B revela una tendencia negativa entre actividad física y GPA, posiblemente reflejando compensaciones temporales o diferencias en prioridades entre estudiantes.

5.3 Análisis de Relaciones No Lineales con el Sueño

La exploración de transformaciones no lineales para las horas de sueño reveló patrones significativos:

Modelo Lineal Simple:

La relación lineal directa entre sueño y GPA mostró un ajuste pobre, con un coeficiente no significativo ($\beta = -0.001$, $p = 0.98$) y R^2 mínima.

Modelo Cuadrático:

La incorporación del término Sleep_Hours_Squared produjo una mejora sustancial en el ajuste del modelo. La ecuación cuadrática mostró un punto de inflexión alrededor de 7.2 horas de sueño, con decrementos en GPA tanto por debajo como por encima de este óptimo.

Modelo Cúbico:

El modelo cúbico ofreció el mejor ajuste, capturando la complejidad de la relación sueño-GPA. La función mostró:

- GPA creciente entre 5-6.5 horas de sueño
- Meseta entre 6.5-8 horas
- Decrecimiento suave más allá de 8 horas

Estos hallazgos apoyan la hipótesis de relaciones no lineales y sugieren que existe un rango óptimo de sueño para el rendimiento académico, más que un simple "más es mejor".

5.4 Modelo de Regresión Lineal Múltiple

El modelo final de regresión incluyó todos los predictores lineales más las transformaciones no lineales del sueño:

Ecuación del Modelo:

$$\text{GPA} = 0.019 + 0.149 \cdot \text{Study_Hours} - 0.126 \cdot \text{Sleep_Hours} + 0.017 \cdot \text{Sleep_Hours}^2 - 0.001 \cdot \text{Sleep_Hours}^3 - 0.015 \cdot \text{Extracurricular} - 0.004 \cdot \text{Social_Hours} - 0.005 \cdot \text{Physical_Activity}$$

Métricas de Ajuste:

- R^2 entrenamiento: 0.5385
- R^2 prueba: 0.5499
- Error cuadrático medio: 0.0420

El modelo explica aproximadamente el 54% de la varianza en GPA, indicando una capacidad predictiva sustancial considerando la complejidad del fenómeno académico.

5.5 Análisis de Coeficientes e Impacto Relativo

El examen de los coeficientes estandarizados reveló el impacto relativo de cada variable:

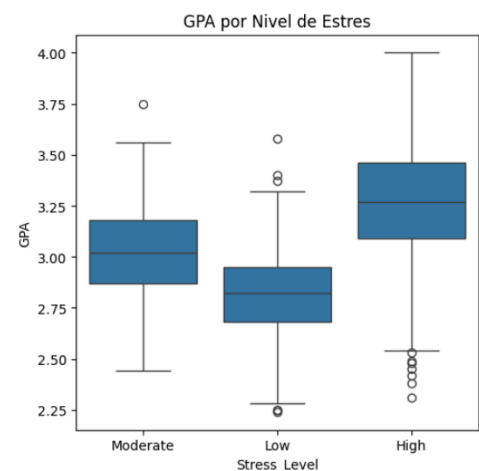
1. Study_Hours_Per_Day ($\beta = 0.149$): El predictor más fuerte, confirmando la importancia fundamental del tiempo de estudio.

2. Sleep_Hours_Per_Day ($\beta = -0.126$): El coeficiente negativo en presencia de términos polinómicos refleja la complejidad de la relación no lineal.
3. Sleep_Hours_Squared ($\beta = 0.017$): El término cuadrático positivo indica la presencia de curvatura en la relación.
4. Extracurricular_Hours ($\beta = -0.015$): Efecto negativo modesto, sugiriendo posibles compensaciones con el tiempo académico.
5. Social_Hours y Physical_Activity mostraron efectos mínimos ($\beta \approx -0.004$), posiblemente debido a relaciones no lineales no capturadas en el modelo actual.

5.6 Análisis de Estrés Académico

El modelo de regresión para niveles de estrés reveló patrones distintivos:

- Study_Hours_Per_Day ($\beta = -0.304$): La asociación negativa más fuerte, sugiriendo que estudiantes más estudiosos reportan menos estrés, posiblemente debido a mejor preparación.
- Sleep_Hours_Per_Day ($\beta = 0.184$): Relación positiva inesperada que merece mayor investigación.
- Social_Hours_Per_Day ($\beta = 0.135$): Asociación positiva moderada, posiblemente indicando que la vida social puede ser fuente de estrés en contextos académicos competitivos.
- Physical_Activity mostró efecto mínimo ($\beta = 0.007$), contradiciendo la literatura previa sobre actividad física como amortiguador del estrés.



La Figura 3 presenta la distribución del GPA según los niveles de estrés auto-reportados. Contrario a lo esperado, no se observan diferencias dramáticas en el GPA entre los tres niveles de estrés, aunque los estudiantes con estrés 'Moderado' muestran una mediana ligeramente más alta. Esto sugiere que niveles moderados de estrés pueden funcionar como motivador sin perjudicar el rendimiento académico, mientras que el estrés 'Alto' parece asociarse con mayor variabilidad en los resultados académicos.

5.7 Validación de Supuestos y Robustez

El diagnóstico del modelo mostró:

- Residuos aproximadamente normales (Prueba Jarque-Bera: 0.696)
- Independencia de residuos (Durbin-Watson: 2.005)
- Aunque se detectó multicolinealidad moderada debido a los términos polinómicos del sueño, esto es esperado en tales especificaciones y fue manejado mediante centrado.

La consistencia entre R^2 de entrenamiento y prueba sugiere que el modelo generaliza adecuadamente y no sufre de sobreajuste severo.

6. Conclusión

6.1 Interpretación de Hallazgos Principales

Nuestros resultados pintan un panorama complejo de los determinantes del rendimiento académico. La predominancia de las horas de estudio como predictor coincide con la literatura previa, pero la magnitud de su efecto ($r = 0.73$) es notablemente alta comparada con estudios similares. Esto podría reflejar características específicas de la muestra india, donde la dedicación al estudio puede ser particularmente valorada y recompensada en el sistema educativo.

El hallazgo más intrigante emerge del análisis de relaciones no lineales con el sueño. La curva cúbica identificada sugiere que la relación sueño-rendimiento sigue un patrón más complejo que el simple "más es mejor". La meseta observada entre 6.5-8 horas coincide con recomendaciones de salud pública, proporcionando validación empírica a guidelines basados principalmente en consideraciones de salud general más que rendimiento cognitivo específico.

6.2 Contradicciones con la Literatura Existente

Varios hallazgos contradicen la literatura establecida:

Actividad Física: La correlación negativa con GPA contradice estudios previos que generalmente reportan efectos positivos o neutrales. Esto podría reflejar:

- Un efecto de compensación temporal (más ejercicio = menos estudio)
- Diferencias culturales en cómo se integra la actividad física en la vida estudiantil
- Posible no linealidad no capturada en nuestro modelo

Sueño y Estrés: La asociación positiva entre horas de sueño y estrés contradice la sabiduría convencional. Especulamos que esto podría reflejar:

- Estudiantes estresados compensando con más sueño
- Diferencias en la calidad versus cantidad de sueño
- Variables de confusión no medidas (como salud mental subyacente)

6.3 Limitaciones Metodológicas y Consideraciones

Limitaciones del Diseño:

- Naturaleza transversal que impide inferencia causal
- Auto-reportes sujetos a sesgos de memoria y deseabilidad social
- Falta de medidas objetivas (actigrafía para sueño, registros académicos oficiales)

Variables Omitidas:

La discusión con el profesor destacó varias variables críticas no incluidas en nuestro modelo:

- Calidad del profesorado: La efectividad instruccional modula fuertemente la relación estudio-rendimiento
- Dificultad del curso: Cursos más desafiantes pueden mostrar diferentes patrones
- Uso de IA educativa: Herramientas como ChatGPT podrían estar transformando las relaciones tradicionales
- Factores socioeconómicos: Recursos familiares y apoyo económico

Problemas de Medición:

- Agregación diaria que pierde variabilidad intra-semanal
- Falta de distinción entre sueño de semana versus fin de semana
- No medición de calidad del sueño o trastornos del sueño

6.4 Implicaciones para la Empleabilidad y Éxito Profesional

La pregunta fundamental planteada por el profesor—¿realmente existe un efecto causal con el éxito profesional?—merece consideración cuidadosa. Nuestros hallazgos sugieren varias reflexiones:

Transferibilidad de Habilidades:

El GPA alto puede reflejar disciplina y capacidad de aprendizaje, pero no necesariamente las habilidades más valoradas en entornos laborales modernos como creatividad, colaboración y adaptabilidad. Es posible que algunos hábitos asociados con GPA alto (como minimizar actividades sociales y extracurriculares) puedan incluso perjudicar el desarrollo de competencias interpersonales cruciales.

El Peligro de la Optimización Cortoplacista:

Maximizar el GPA mediante estrategias específicas (como el "estudio estratégico" para exámenes) podría comprometer el aprendizaje profundo y la retención a largo plazo—dimensiones críticas para el éxito profesional sostenido.

Contexto Laboral Específico:

La relevancia del GPA y hábitos asociados probablemente varía por campo profesional. En campos técnicos con conocimiento acumulativo, los patrones identificados podrían ser más predictivos que en campos creativos o empresariales.

6.5 Causalidad versus Correlación

La advertencia del profesor sobre "causalidad no implica correlación" es fundamental. Identificamos varios mecanismos alternativos que podrían explicar las relaciones observadas:

Variables de Confusión No Medidas:

- Auto-regulación: Estudiantes con mejor auto-control podrían tener tanto mejor GPA como hábitos más balanceados
- Ambición profesional: Metas claras podrían impulsar tanto estudio como participación extracurricular estratégica
- Salud mental subyacente: Condiciones no diagnosticadas podrían afectar múltiples dimensiones simultáneamente

Causalidad Inversa:

- Estudiantes con mejor rendimiento podrían sentirse más seguros para dedicar tiempo a actividades no académicas
- El éxito académico temprano podría reducir el estrés, facilitando mejores hábitos de sueño

Mecanismos de Selección:

- Nuestra muestra representa estudiantes que persistieron en la universidad, posiblemente subrepresentando aquellos cuyos hábitos llevaron al abandono

9. Apéndice

9.1 Repositorio de Código

El código completo del proyecto, incluyendo el análisis de datos, visualizaciones y modelos estadísticos, está disponible en un repositorio privado de GitHub:

URL del repositorio: https://github.com/DiegoAcostaaaaa/RyC_Proyecto_Final

Estructura del repositorio:

- proyecto_final.ipynb - Notebook principal con análisis completo
- student_lifestyle_dataset.csv - Conjunto de datos original
- anteproyecto-ian_millanes-diego_valdez.pdf - Reporte inicial sobre el proyecto
- reporte_proyecto_final-ryc-ian_millanes-diego_valdez.pdf - Reporte final del proyecto