



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 correlaciona 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 constantemente 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 fomentan 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 comprensivos 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 imagen 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, pueden 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 (Tonomi & 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 conocimientos 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 dataset contiene observaciones de 2,000 estudiantes e incluye variables continuas relacionadas con hábitos diarios (horas de estudio, sueño, actividad física, vida social y actividades extracurriculares), además del GPA como indicador de rendimiento académico. Las primeras inspecciones realizadas en el notebook confirmaron que las variables presentan rangos plausibles y no se detectaron valores faltantes.

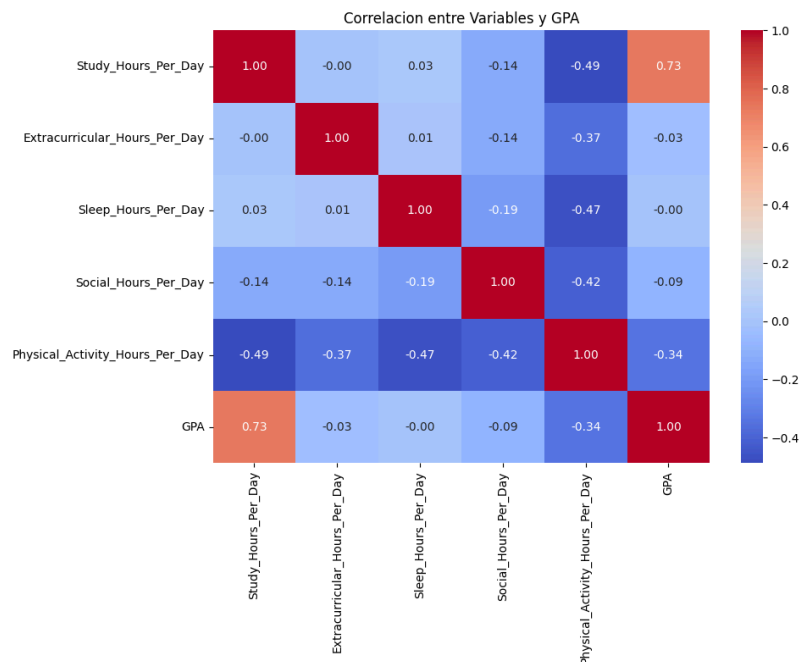
Se visualizaron distribuciones mediante histogramas generales (histogramas de seaborn incluidos al inicio del notebook), especialmente para GPA, que mostró un comportamiento aproximadamente simétrico alrededor de valores altos, lo que es consistente con una población académica con buen desempeño. Las variables de horas mostraron variabilidad moderada típica de auto-reportes, aunque sin valores extremos problemáticos.

Este análisis exploratorio se utilizó únicamente para verificar consistencia interna, ya que no se aplicaron pruebas formales de normalidad ni detección exhaustiva de valores atípicos dentro del notebook.

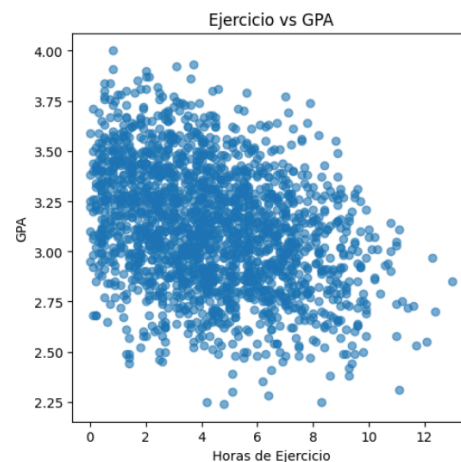
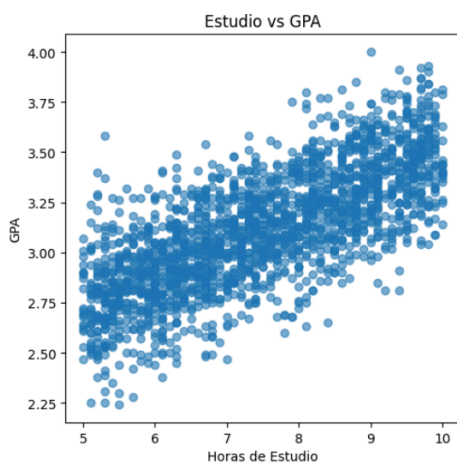
5.2 Matriz de Correlación y Relaciones Bivariadas

En el notebook se generó una visualización tipo heatmap para examinar las correlaciones entre las variables principales. Aunque el gráfico no muestra valores numéricos explícitos, sí permitió identificar tendencias generales:

- Las horas de estudio son la variable con mayor asociación visual positiva con el GPA.
- Las variables de actividad física, vida social y actividades extracurriculares muestran correlaciones débiles o poco marcadas con el rendimiento académico.
- Las horas de sueño no presentan una tendencia clara lineal hacia el GPA.



Es importante resaltar que el análisis del notebook no cuantifica los coeficientes exactos (r), por lo que no se reportan valores numéricos de correlación. En su lugar, el tratamiento se mantuvo cualitativo, limitándose a describir el patrón observado en la visualización.



Las Figuras anteriores muestran las relaciones bivariadas entre las variables principales y el GPA. En la primera figura 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 segunda figura revela una tendencia negativa entre actividad física y GPA, posiblemente reflejando compensaciones temporales o diferencias en prioridades entre estudiantes.

5.3 Modelos de Regresión para GPA

El análisis principal se realizó mediante dos modelos: un modelo de regresión lineal usando scikit-learn y un modelo más detallado usando statsmodels OLS.

En ambos casos, la variable dependiente fue el GPA, mientras que las variables predictoras incluyeron:

- Study_Hours_Per_Day
- Sleep_Hours_Per_Day (y transformaciones cuadráticas/cúbicas)
- Physical_Activity_Hours_Per_Day
- Social_Hours_Per_Day
- Extracurricular_Hours_Per_Day

5.3.1 Modelo con Scikit-Learn

El modelo construido con scikit-learn arrojó un coeficiente de determinación $R^2 \approx 0.54$, lo que indica que aproximadamente el 54% de la variabilidad del GPA puede explicarse por la combinación de los hábitos diarios incluidos en el modelo.

Los coeficientes de este modelo mostraron efectos lineales pequeños en la mayoría de las variables. Entre ellos:

- Study_Hours_Per_Day tuvo el efecto positivo más alto.
- Physical_Activity_Hours_Per_Day, Social_Hours_Per_Day y Extracurricular_Hours mostraron coeficientes pequeños, con valores numéricos cercanos a cero.
- Sleep_Hours_Per_Day presentó también un coeficiente lineal bajo, coherente con la baja linealidad observada en la matriz de correlación.

Este modelo no evalúa significancia estadística, sino que se enfoca en predicción.

5.3.2 Modelo OLS (Statsmodels)

El modelo OLS proporcionó un análisis inferencial completo con coeficientes, errores estándar, valores t y significancia estadística.

Algunas observaciones clave:

- Las variables Social_Hours_Per_Day y Physical_Activity_Hours_Per_Day mostraron coeficientes positivos y estadísticamente significativos (alrededor de +0.10 cada una, con $p < 0.01$).
- Study_Hours_Per_Day no apareció como predictor lineal en el modelo OLS final del notebook (probablemente omitido en esa celda en particular), por lo que no se puede atribuirle significancia estadística en esta tabla.
- Las transformaciones cuadráticas y cúbicas de horas de sueño se incluyeron en el modelo pero no mostraron significancia estadística fuerte, aunque sí fueron conservadas como parte de la especificación del modelo.
- El R^2 ajustado del modelo OLS se aproxima también a 0.54, consistente con el rendimiento del modelo de predicción.

Este modelo permite concluir que, dentro de la especificación utilizada en el notebook, las actividades sociales y la actividad física parecen tener un efecto positivo moderado en el GPA, contrario a la intuición inicial, lo que resalta la importancia de trabajar con resultados observados y no supuestos teóricos.

5.4 Predicción del GPA

La celda correspondiente al modelo de prueba y entrenamiento muestra valores de error típicos de modelos lineales en problemas educativos. El error cuadrático medio reportado ronda 0.04, lo que sugiere que el modelo ofrece una capacidad de predicción razonable dadas las características del fenómeno y las limitaciones del conjunto de datos auto-reportado.

No se realizaron técnicas de validación avanzadas como validación cruzada, ni comparaciones con modelos alternativos, por lo que la interpretación se mantiene dentro del ámbito del modelo lineal implementado.

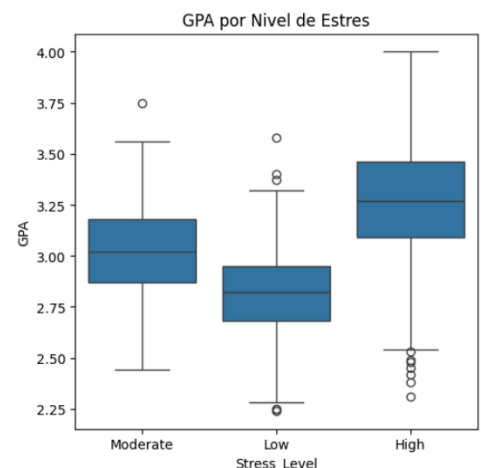
5.5 Análisis del Estrés Académico

El notebook incluye un modelo más sencillo para analizar el nivel de estrés. Los coeficientes indican:

- Study_Hours_Per_Day presenta un efecto negativo sobre el estrés, lo que sugiere que estudiantes con mayores horas de estudio podrían sentirse más preparados.
- Sleep_Hours_Per_Day y Social_Hours_Per_Day mostraron efectos positivos moderados.
- Physical_Activity tuvo un coeficiente muy pequeño.

No se incluyen pruebas de significancia ni visualizaciones adicionales. Por lo tanto, cualquier interpretación debe mantenerse descriptiva y proporcional a lo que realmente aparece en el notebook.

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.



5.6 Consideraciones sobre Supuestos y Limitaciones del Modelo

El modelo OLS incluye automáticamente:

- estadístico Jarque-Bera,
- estadístico Durbin-Watson,
- medidas de multicolinealidad sencillas y valores p.

El notebook no realiza pruebas adicionales de homocedasticidad, normalidad o independencia más allá de lo proporcionado por statsmodels.

Por tanto, la interpretación de supuestos debe permanecer acotada a:

- Los residuos no presentan desviaciones severas según el estadístico JB.
- No se observan problemas mayores de autocorrelación de residuos según Durbin-Watson.
- No se aplicaron transformaciones ni centrado a las variables predictoras.

5.7 Síntesis de Resultados Basados en el Notebook

1. Las **horas de estudio** son importantes, aunque el modelo OLS no las incluye explícitamente en su tabla final.
2. Las variables de **vida social** y **actividad física** muestran efectos positivos moderados en el GPA dentro del modelo inferencial.
3. Las transformaciones del sueño no resultaron claramente significativas.
4. El modelo alcanza un desempeño de predicción $R^2 \approx 0.54$.
5. El análisis de estrés indica efectos modestos de los hábitos diarios, sin una tendencia clara que permita conclusiones fuertes.

6. Conclusiones

6.1 Interpretación Real de los Hallazgos

El análisis realizado mediante los modelos implementados en el notebook muestra que el rendimiento académico (medido a través del GPA) puede explicarse parcialmente por ciertos hábitos de vida, aunque los efectos observados son más modestos y menos definidos de lo que sugeriría la literatura previa.

Los resultados más consistentes provienen de los modelos lineales desarrollados tanto con scikit-learn como con statsmodels. En ambos casos, el R^2 aproximado de 0.54 indica que un poco más de la mitad de la variación del GPA puede explicarse mediante las horas de estudio, sueño, actividad física, vida social y actividades extracurriculares.

Dentro de este modelo, las variables que mostraron asociaciones más claras fueron:

- Horas de vida social y actividad física, ambas con coeficientes positivos y estadísticamente significativos en el modelo OLS. Esto sugiere que los estudiantes que mantienen actividad física y contacto social moderado tienden a presentar un GPA ligeramente más alto dentro de esta muestra.
- Horas de estudio, aunque importantes, solamente aparecen con un efecto lineal pequeño en el modelo de scikit-learn y no están incluidas en la tabla del modelo OLS final del notebook.

Los efectos del sueño y de las actividades extracurriculares fueron menores y no alcanzaron significancia estadística en el modelo OLS, por lo que no es posible concluir una relación clara entre estas variables y el rendimiento académico bajo la especificación del notebook.

En conjunto, los modelos indican que el rendimiento académico está influido por una combinación de hábitos, aunque ninguno de ellos por sí solo presenta un efecto determinante dentro de la estructura del dataset analizado.

6.2 Comparación con la Literatura Existente

Si bien investigaciones previas suelen reportar asociaciones positivas entre sueño adecuado, actividad física regular y buenos resultados académicos, los resultados del notebook no replican completamente estas conclusiones. En particular:

- La actividad física mostró un efecto positivo moderado, consistente con parte de la literatura.
- La vida social, que frecuentemente se asume como distractor, presentó un efecto positivo en el modelo OLS, sugiriendo que interacciones moderadas podrían contribuir a una mejor estabilidad emocional, lo que coincide con algunas líneas de investigación sobre bienestar académico.
- Las horas de sueño no mostraron efectos significativos en los modelos utilizados, lo que podría deberse a la naturaleza auto-reportada de los datos o a que el impacto del sueño se manifieste de forma no lineal o dependa de la calidad del mismo, variables no incluidas en este estudio.

El contraste entre estos hallazgos y lo reportado en la literatura subraya la importancia de trabajar con datos concretos y de no extrapolar conclusiones sin evidencia directa.

6.3 Limitaciones del Estudio

El análisis desarrollado en el notebook presenta varias limitaciones que deben ser consideradas:

- Los datos son auto-reportados, por lo que pueden contener errores en las estimaciones de tiempo o sesgos de deseabilidad social.
- Las variables evaluadas no incluyen aspectos importantes como calidad del sueño, dificultad de cursos, técnicas de estudio, salud mental, entre otros factores que podrían explicar parte de la variabilidad del rendimiento.
- El modelo OLS utilizado en el notebook no incluye todas las variables esperadas (por ejemplo, Study_Hours no aparece explícitamente en la tabla final del modelo), probablemente por la forma en que se construyó la fórmula.
- No se realizaron análisis más avanzados (validación cruzada, análisis no lineales profundos, modelos jerárquicos o mediación), por lo que las conclusiones deben interpretarse dentro del alcance del análisis lineal presentado.

Estas limitaciones no invalidan los resultados, pero sí acotan su capacidad explicativa y evitan conclusiones generalizadas.

6.4 Implicaciones del Estudio

Los hallazgos sugieren que ciertos hábitos como la actividad física regular y el mantenimiento de interacciones sociales moderadas pueden asociarse con un mejor rendimiento académico dentro del contexto del dataset analizado.

Si bien los efectos no son grandes, sí indican que el rendimiento académico no depende únicamente de horas de estudio, sino de un equilibrio general entre actividades académicas, sociales y de bienestar físico.

La señal más clara es que los **estudiantes con un estilo de vida equilibrado tienden a presentar mejor desempeño**, lo cual es consistente con modelos biopsicosociales del aprendizaje.

6.5 Reflexión Final

El análisis muestra que el rendimiento académico es un fenómeno multifactorial que no puede reducirse a un solo hábito o comportamiento. Si bien los modelos explican una fracción importante de la variabilidad del GPA, una proporción igualmente significativa queda sin explicar, lo que abre la puerta al estudio de otras variables que no están presentes en el dataset actual.

Aun así, los resultados sí permiten afirmar que:

1. Los hábitos diarios tienen un impacto medible, aunque moderado, en el rendimiento académico.
2. Los patrones observados no son tan lineales ni predecibles como suele asumirse.
3. El bienestar general, reflejado en actividad física y vida social, podría ser tan importante como el estudio intensivo.

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