

Trabajo Matemáticas IV: 1

Introducción

Dataset: ***Screen Time vs Mental Wellness Survey - 2025***

El Dataset consta de una investigación de 400 participantes que relaciona el uso diario de pantallas con su bienestar mental. Esto surge por el uso de dispositivos digitales diariamente, se intenta comprender el vínculo que se desarrolla entre la calidad de sueño, el estrés y la productividad. Esto lo hace un estudio crucial para la ciencia de datos, la psicología y la salud.

Cada columna representa a un participante e incluye:

- Demografía del participante, tanto como su edad, género, ocupación, y si es estudiante o trabajador (incluso ambas).
- Total del promedio del tiempo diario en pantalla el cual se divide en tiempo de trabajo y libre de la persona (incluye TV, Computadoras, Teléfonos, etc...)
- Calidad de sueño (a percepción del participante)
- Nivel de estrés (en una escala del 1 al 10)
- Nivel de productividad (a percepción del participante)
- Estados de bienestar (De humor, nivel de energía, capacidad de concentración)
- Relación entre tiempo en pantalla (podemos resumir esta variable como Screen Time ó ST) y nivel de estrés (también como Stress Level ó SL)
- ¿un mayor ST afecta a la calidad del sueño?
- Patrones de uso de pantallas entre estudiantes y profesionales
- Predicción del puntaje de bienestar mental en base a las variables de estilo de vida del participante.

Regresión lineal simple (incisos a, b y c)

En esta parte del trabajo analizaremos la relación entre la variable(s) predictora y la variable respuesta, además justificaremos la dependencia entre estas variables.

Variable de respuesta (Y)

Como variable de respuesta (la variable denominada Y) usaremos el Nivel de estrés, este Nivel de estrés tendrá una escala del 1 al 10 siendo el número 10 el nivel más alto posible y 1 el menor. Esta variable es ideal porque es uno de los mejores indicadores que reflejan el bienestar del participante haciendo posible comprender los factores que impactan a dicho bienestar.

Variables predictoras (X)

Como variables predictoras (las variables denominadas x) usaremos:

1. Productividad (x_1): medido según la percepción del participante (del 1 al 10).
2. Calidad de sueño (x_2): medido según la percepción del participante (del 1 al 10).
3. Horas de sueño (x_3): medido en horas

Consideramos estas 3 variables como herramienta para sostener la hipótesis de que la productividad durante el día, las horas de sueño durante la noche y la calidad de su descanso, afectan en el bienestar mental del participante. A continuación justificamos cada una de ellas:

1. **Productividad (x_1):** Un mayor nivel de productividad autopercibida en ocasiones se asocia con un sentido de logro, propósito y control sobre las propias actividades (ya sean laborales, domésticas, hobbies, etc...). Se teoriza que esta percepción positiva puede ser un factor protector contra el estrés, contribuyendo a un mejor estado de bienestar mental.
2. **Calidad de sueño (x_2):** Un sueño de alta calidad (reparador y sin interrupciones) es crucial para la recuperación mental y la regulación emocional. Por lo tanto, se espera que una mejor calidad de sueño se correlacione negativamente con los niveles de estrés.
3. **Horas de sueño (x_3):** Más allá de la calidad, la cantidad de horas de sueño es un factor esencial para el día a día. Un déficit de sueño puede afectar negativamente la función cognitiva y la capacidad para manejar presiones diarias, lo que podría llevar a un incremento en la percepción del estrés.

Y	$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$	$\hat{\sigma}^2$	R^2	r	IC(β_1)	IC(β_0)	ICM(Y)	IP(Y)
productivity (x_1)	$\hat{y}_1 = 14.8122 - 0.1227x_i$	0.996	0.774	-0.876	[-0.129, -0.116]	[14.444, 15.181]	[8.052, 8.248]	[6.186, 10.115]
sleep quality (x_2)	$\hat{y}_2 = 10.6585 - 1.7947x_i$	3.0253	0.312	-0.56	[-2.057, -1.532]	[10.254, 11.063]	[7.98, 8.321]	[4.727, 11.574]
sleep hours (x_3)	$\hat{y}_3 = 16.4429 - 1.1824x_i$	3.381	0.231	-0.481	[-1.395, -0.97]	[14.943, 17.943]	[7.97, 8.331]	[4.531, 11.770]

Nota: Los datos de la tabla han sido calculados mediante un algoritmo de Python en la plataforma Google Colab

Con los resultados a nuestro alcance, podemos deducir cual es la mejor variable predictora.

Para esto, comparamos los valores de R^2 obtenidos:

- **Productividad:** $R^2 = 0.774$ ó 77.4%
- **Calidad de sueño:** $R^2 = 0.312$ ó 31.2%
- **Horas de sueño:** $R^2 = 0.231$ ó 23.1%

Esto nos indica que la variable de productividad es una pieza fundamental y con gran peso para la predicción buscada. Siendo el porcentaje de estas variables el que explica la variabilidad del estrés, en este caso el 77.4% es explicado por el modelo. A continuación daremos algunos comentarios sobre algunos de los resultados:

- El modelo se define por la ecuación $\hat{y} = 14.8122 - 0.1227x_i$. La pendiente con valor $\hat{\beta}_1 = -0.1227$ nos indica que por cada punto que aumenta la productividad, se calcula una disminución de 0.12 puntos del nivel de estrés.
- La correlación toma un valor de $r = -0.876$ que se traduce en una correlación lineal negativa muy fuerte (cercana a -1) entre la productividad y el estrés, es decir, mientras mayor productividad, menor estrés.
- El intervalo de confianza para la pendiente $IC(\beta_1) = [-0.129, -0.116]$ no contiene el valor cero. Demostrando con un 95% de confianza que la relación entre la productividad y el estrés es real.

Regresión lineal múltiple (incisos d, e y f):

Usando el método de descenso de gradiente tenemos como ecuación:

$$\hat{y} = 15.4044 - 0.1206 * x_1 + 0.0685 * x_2 - 0.115 * x_3$$

Mientras que con el método de mínimos cuadrados tenemos como ecuación:

$$\hat{y} = 15.4985 - 0.1224 * x_1 + 0.1231 * x_2 - 0.1238 * x_3$$

Nota: Los datos de las ecuaciones (β) han sido obtenidos mediante un algoritmo de Python en Google Colab

Análisis de los resultados

Las diferencias observadas entre los coeficientes obtenidos por **mínimos cuadrados** y por **descenso de gradiente** se deben principalmente a la naturaleza de cada método. Mientras que el **método de mínimos cuadrados calcula los coeficientes de forma exacta** mediante una solución directa de la ecuación normal, **el descenso de gradiente aproxima los valores iterativamente**, actualizando los coeficientes paso a paso según el *learning_rate* y deteniéndose cuando la diferencia en la función de costo es menor a la tolerancia. Por esta razón, los coeficientes del gradiente pueden diferir ligeramente, especialmente si la tasa de aprendizaje no permite converger completamente al mínimo exacto.

El término independiente (15.4985) representa el nivel de estrés estimado cuando todas las variables predictoras son cero.

El coeficiente de **productividad** (-0.1224) indica que, al aumentar la productividad en una unidad, el nivel de estrés disminuye en promedio 0.12 unidades, manteniendo las demás variables constantes.

Por el contrario, el coeficiente de **calidad del sueño** (0.1231) muestra un leve efecto positivo, lo que podría interpretarse como una relación con las **horas de sueño**, dado que esta última variable presenta un coeficiente negativo (-0.1238) y la correlación entre ambas es de 0.614 y si junto a esto la información que se obtiene de esta variable es similar a la de la calidad del sueño entonces esto puede influir en la dirección y magnitud de los coeficientes estimados.

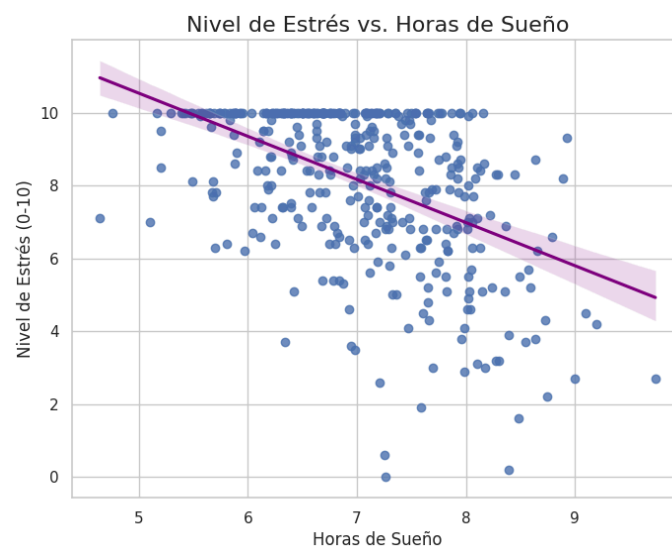
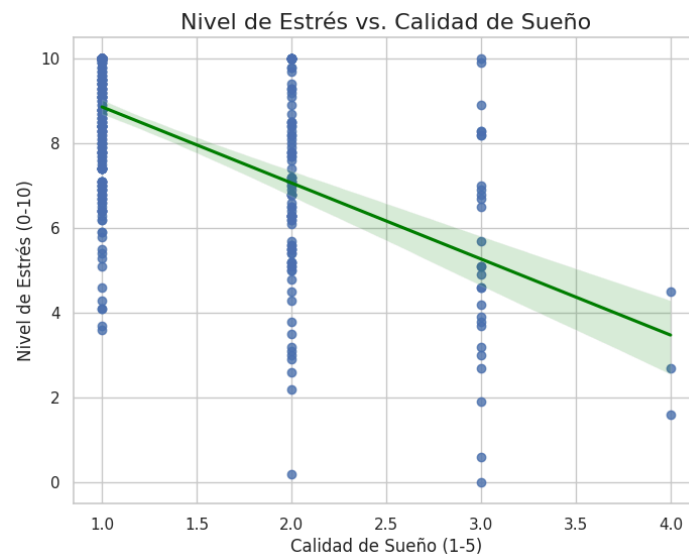
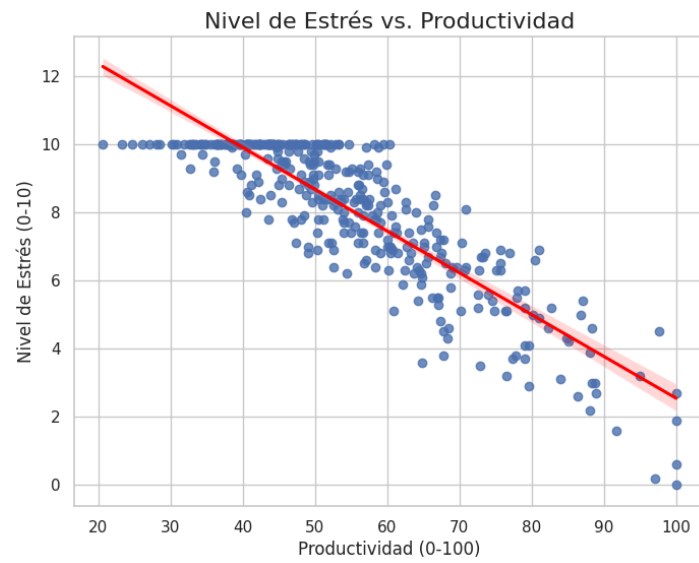
¿Más variables predictoras mejoraron el modelo?

Para evaluar si el modelo mejoró agregando más variables predictoras lo que tenemos que tener en cuenta es el coeficiente de determinación ajustado de R^2 ya que nos guiará a la hora de saber si estas variables aportan valor ó por el contrario no son necesarias (observar que entre la variable R^2 y la variable R^2 ajustada no hay una diferencia notable), en este caso $R^2_{adj} = 0.774$ y el de $R^2 = 0.7752$. Esto nos dice que las tres variables explican el 77.4% de la variabilidad del estrés.

Comparando estos dos valores obtenidos, se concluye que la adición de las variables del sueño no implican una mejora en la predicción del modelo. Esto **no significa** que las variables del sueño no afecten al estrés, si no que, afectan directamente a la productividad (y en consecuencia, al estrés).

Concluimos que el modelo más eficiente y sencillo es el de regresión simple que usa a la productividad como variable **calculada en el inciso C**.

Extra: Gráficos para poder ver las relaciones de una forma más amigable



Gentile Valentino, Roa Elías, Rodríguez Ricon Mauricio

Códigos usados para la regresión lineal:

https://colab.research.google.com/drive/1OmtBegmyRG1_JWAFazHsCuhB_C6pwvft?usp=sharing

Link del dataset usado:

[Screen Time vs Mental Wellness Survey - 2025](#)