

20 DE NOVIEMBRE DE 2025

**Patrones de uso digital y bienestar  
psicológico en jóvenes y adultos:  
análisis de factores de riesgo y  
productividad**

Tarea módulo 6 - Análisis de datos

ELENA FERNÁNDEZ CHIRINO  
DESARROLLO DE APLICACIONES EN PYTHON (2<sup>a</sup> PARTE)  
IES Trassierra (Córdoba)

# Contenido

1. Introducción: contexto del estudio y relevancia social.....	2
2. Metodología: descripción de la muestra, variables y herramientas .....	2
2.1. Preparación del dataset .....	2
2.2. Variables seleccionadas para el estudio.....	2
2.2.1. Variables sociodemográficas.....	2
2.2.2. Variables de uso digital .....	3
2.2.3. Variables de bienestar y salud .....	3
2.2.4. Variables de resultado.....	3
3. Análisis exploratorio: visualizaciones y estadísticas descriptivas .....	3
3.1. Histogramas .....	3
3.2. Boxplot .....	4
3.3. Heatmap .....	4
3.4. Pairplot.....	5
4. Resultados estadísticos: correlaciones, regresiones, pruebas .....	5
4.1. Correlaciones .....	5
4.2. Regresión lineal simple .....	6
4.3. Regresión lineal múltiple.....	6
4.4. Validación del modelo .....	7
4.4.1. Gráfico de residuos .....	7
4.4.2. Histograma de residuos .....	8
5. Conclusiones: perfiles de riesgo, recomendaciones, limitaciones.....	9
5.1. Perfiles de riesgo .....	9
5.2. Recomendaciones .....	10
5.3. Limitaciones .....	10
6. Bibliografía .....	10

# 1. Introducción: contexto del estudio y relevancia social

El presente estudio se enmarca en la **Tarea del Módulo: Análisis de datos**, y tiene como objetivo general analizar cómo el uso intensivo de dispositivos digitales y redes sociales se relaciona con indicadores de bienestar psicológico (como ansiedad, estrés, felicidad y concentración) y con el nivel de productividad, identificando perfiles de riesgo y patrones diferenciados según edad, género y país.

En un contexto de creciente digitalización, especialmente entre adolescentes y jóvenes adultos, resulta fundamental comprender el impacto que los hábitos tecnológicos tienen sobre la salud mental y el rendimiento personal. Este análisis permite no solo detectar factores de riesgo, sino también orientar intervenciones educativas y sociales más eficaces.

## 2. Metodología: descripción de la muestra, variables y herramientas

Para cumplir con este objetivo, se ha seleccionado como fuente de datos el conjunto disponible en [Kaggle: Digital Health and Mental Wellness](#), que contiene información detallada sobre hábitos digitales, salud mental y bienestar de una muestra de 3.500 personas entre 13 y 50 años.

### 2.1. Preparación del dataset

Con el fin de facilitar la comprensión y el tratamiento de los datos por parte del alumnado que se está iniciando en el análisis estadístico, se ha realizado una **preparación previa del archivo CSV** mediante **Notepad++**, aplicando las siguientes modificaciones:

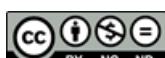
- Se ha ajustado el formato regional del archivo, cambiando los **separadores de campo** de coma (,) a punto y coma (;), y los **separadores decimales** de punto (.) a coma (,), asegurando así una correcta interpretación numérica en entornos configurados en español.
- Asimismo, se ha realizado una **traducción completa del contenido del dataset**, incluyendo tanto los **encabezados de las columnas** como las **etiquetas de los datos**, del inglés al español. Esta adaptación tiene como objetivo **facilitar la comprensión conceptual y técnica** por parte del alumnado, especialmente aquellos que se están iniciando en el análisis estadístico. Para ello, se ha utilizado inicialmente una **hoja de cálculo** que permitió aplicar filtros y revisar la coherencia de los valores, y posteriormente se ha empleado **Notepad++** para realizar las sustituciones masivas de texto.

Este enfoque permite reproducir de forma completa el ciclo de análisis de datos: desde la importación y limpieza hasta la exploración estadística y la extracción de conclusiones relevantes, todo ello documentado en una **Notebook de Jupyter**.

### 2.2. Variables seleccionadas para el estudio

#### 2.2.1. Variables sociodemográficas

- edad (numérica, continua)
- genero (categórica: hombre, mujer, otro)
- pais (categórica)



- nivel\_ingresos (ordinal o continua)
- nivel\_educativo (ordinal)

## 2.2.2. Variables de uso digital

- horas\_dispositivo\_día
- desbloqueo\_telefono
- notificaciones\_dia
- minutos\_redes\_sociales
- tipo\_dispositivo
- nivel\_dependencia\_digital

## 2.2.3. Variables de bienestar y salud

- nivel\_ansiedad
- nivel\_depresion
- nivel\_estres
- nivel\_felicidad
- nivel\_concentracion
- horas\_sueño\_dia
- calidad\_suaño
- dias\_actividad\_física

## 2.2.4. Variables de resultado

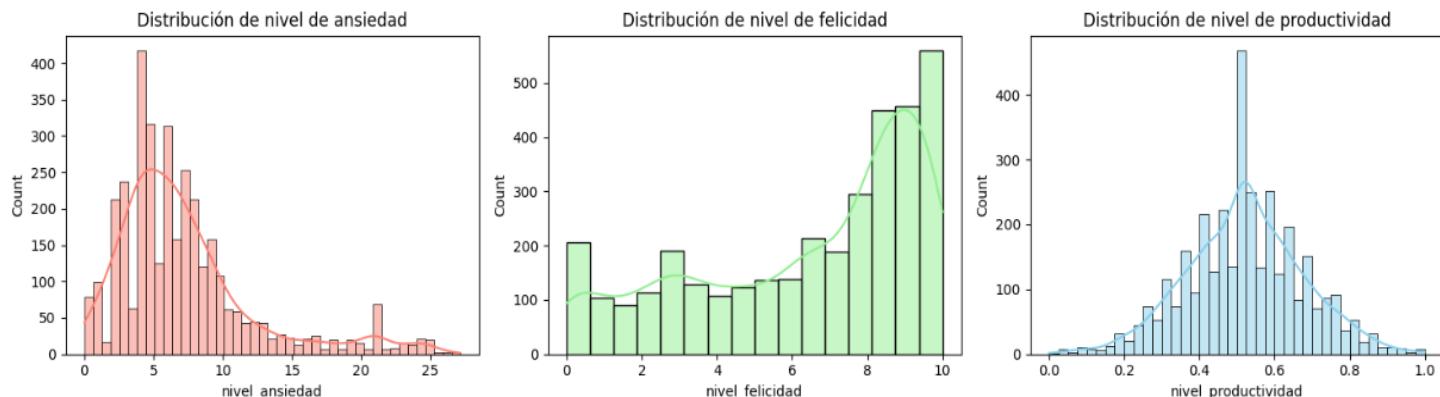
- nivel\_productividad
- indicador\_riesgo\_alto (target binario para predicción)

# 3. Análisis exploratorio: visualizaciones y estadísticas descriptivas

Se han generado diversas visualizaciones para explorar la distribución de variables clave y sus relaciones:

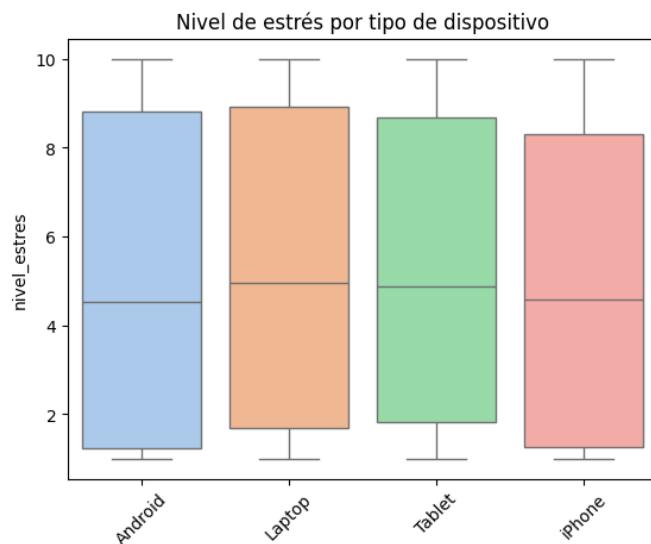
## 3.1. Histogramas

Hemos representado **Histogramas** de ansiedad, felicidad y productividad muestran distribuciones variadas, con sesgos hacia valores altos en ansiedad y felicidad.



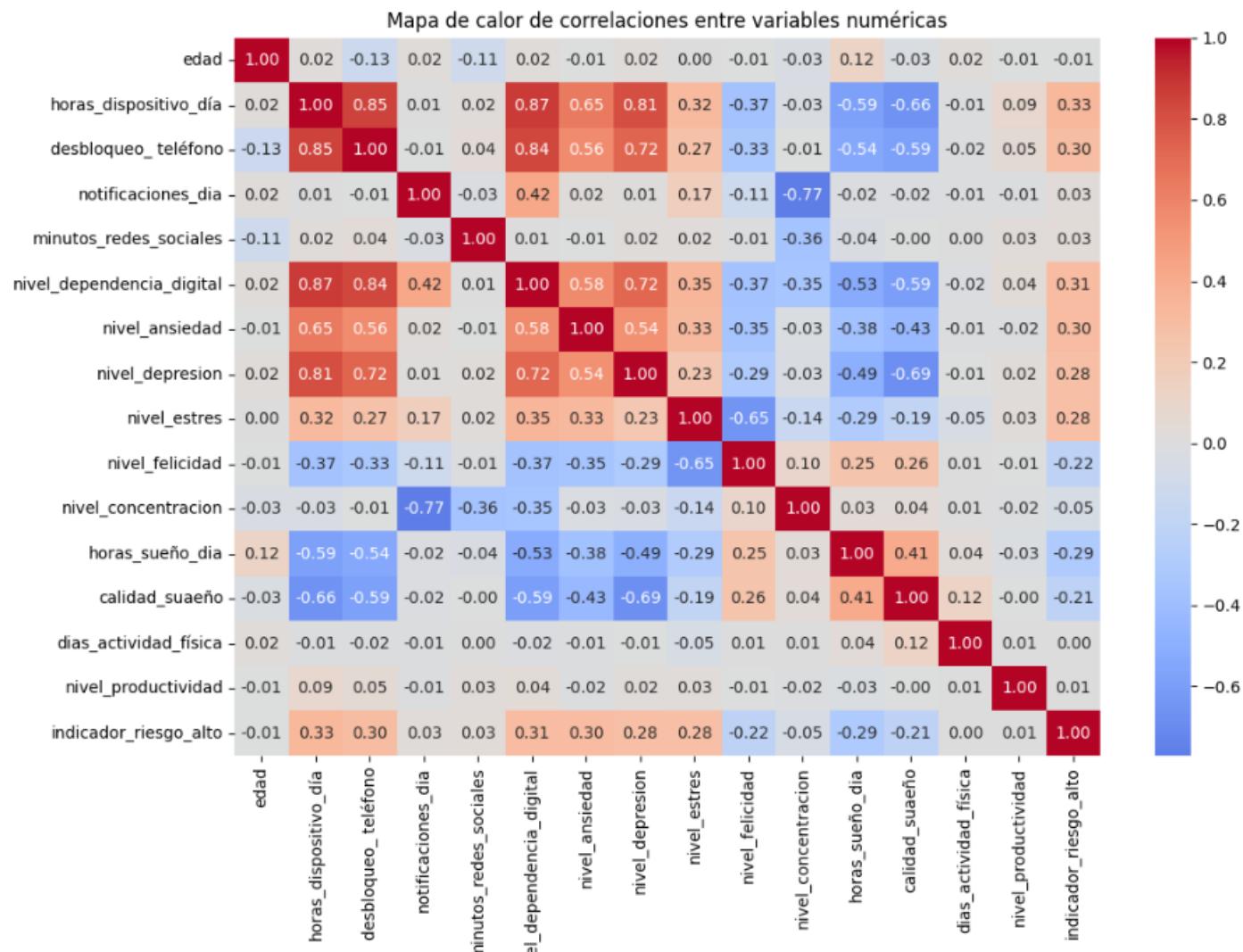
### 3.2. Boxplot

**Diagrama de cajas o bigotes** comparativo de estrés por tipo de dispositivo revela diferencias entre usuarios de móvil, ordenador y tablet.



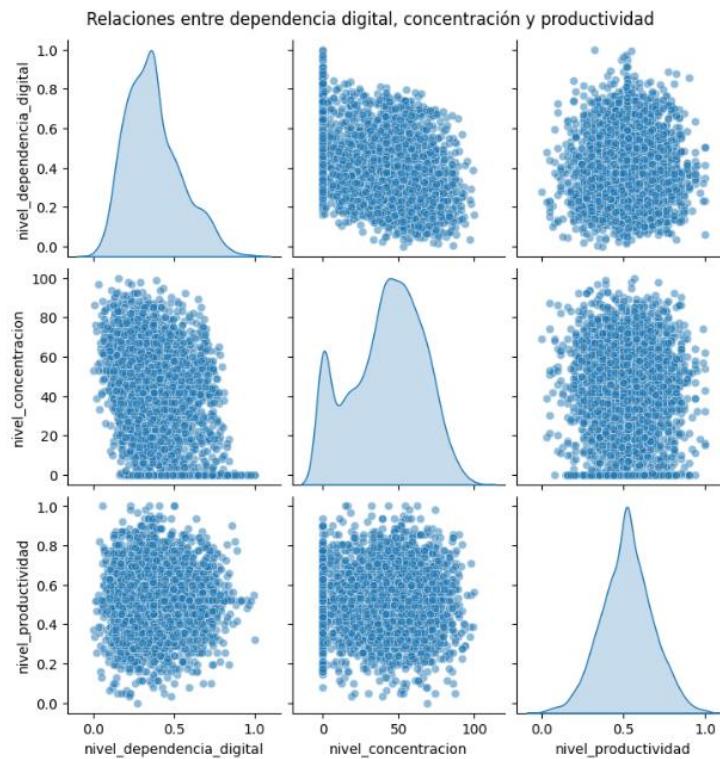
### 3.3. Heatmap

**Mapa de calor de correlaciones** entre variables numéricas muestra asociaciones destacadas entre concentración, productividad y dependencia digital.



### 3.4. Pairplot

**Diagrama de pares** entre concentración, productividad y dependencia digital permite observar visualmente correlaciones negativas entre dependencia y concentración, y positivas entre concentración y productividad.



## 4. Resultados estadísticos: correlaciones, regresiones, pruebas

### 4.1. Correlaciones

Se ha calculado la matriz de correlación de Pearson entre variables clave.

```
Matriz de correlación (Pearson):
    nivel_dependencia_digital  nivel_concentracion \
nivel_dependencia_digital      1.000000      -0.346359
nivel_concentracion            -0.346359       1.000000
nivel_productividad             0.039561      -0.016047

                                         nivel_productividad
nivel_dependencia_digital          0.039561
nivel_concentracion                -0.016047
nivel_productividad                 1.000000
```

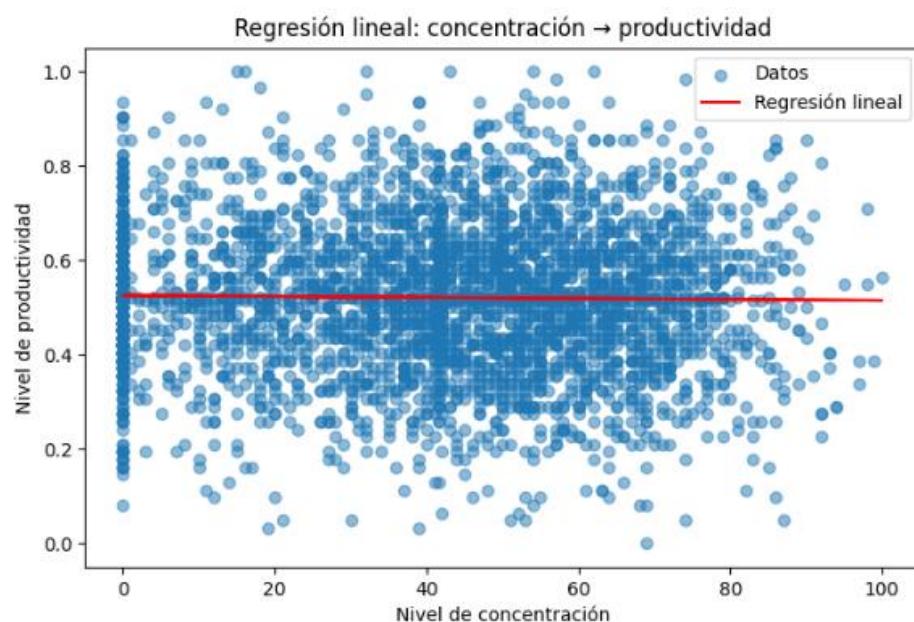
Destacan:

- Correlación negativa entre dependencia digital y concentración.
- Correlación positiva entre concentración y productividad.

## 4.2. Regresión lineal simple

Se ha aplicado una regresión lineal simple entre concentración y productividad. El resultado muestra una pendiente casi nula y un coeficiente de correlación bajo, lo que indica que **no hay una relación lineal significativa** entre ambas variables en este conjunto de datos.

Pendiente: -0.00  
Intercepto: 0.53  
Coeficiente correlación(r): -0.02  
Valor p: 0.3426  
Error estándar: 0.00



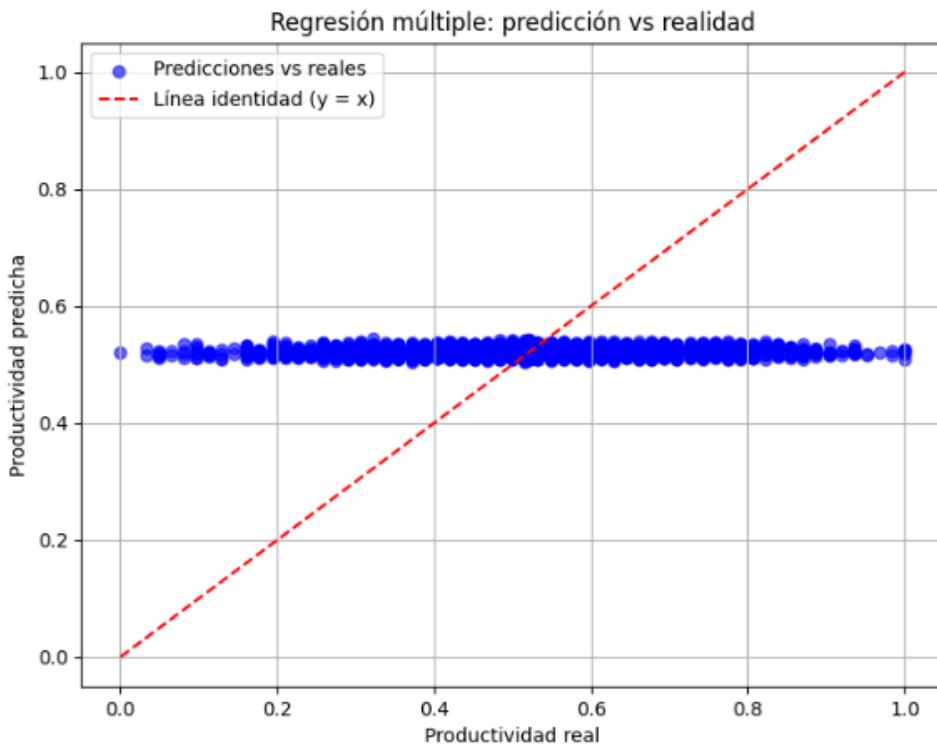
## 4.3. Regresión lineal múltiple

Se ha ajustado un modelo de regresión múltiple con tres predictores: nivel de estrés, horas de sueño al día y nivel de dependencia digital.

```
OLS Regression Results
=====
Dep. Variable: nivel_productividad R-squared: 0.002
Model: OLS Adj. R-squared: 0.001
Method: Least Squares F-statistic: 2.226
Date: Thu, 20 Nov 2025 Prob (F-statistic): 0.0831
Time: 15:07:49 Log-Likelihood: 1542.8
No. Observations: 3500 AIC: -3078.
Df Residuals: 3496 BIC: -3053.
Df Model: 3
Covariance Type: nonrobust
=====
            coef    std err      t    P>|t|    [0.025    0.975]
-----
const      0.5164   0.023   22.936   0.000     0.472    0.561
nivel_estres  0.0007   0.001    0.888   0.375    -0.001    0.002
horas_sueño_dia -0.0012   0.002   -0.510   0.610    -0.006    0.004
nivel_dependencia_digital  0.0264   0.019    1.395   0.163    -0.011    0.063
=====
Omnibus: 2.717 Durbin-Watson: 2.028
Prob(Omnibus): 0.257 Jarque-Bera (JB): 2.850
Skew: -0.009 Prob(JB): 0.241
Kurtosis: 3.139 Cond. No. 93.0
=====
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

El modelo presenta:

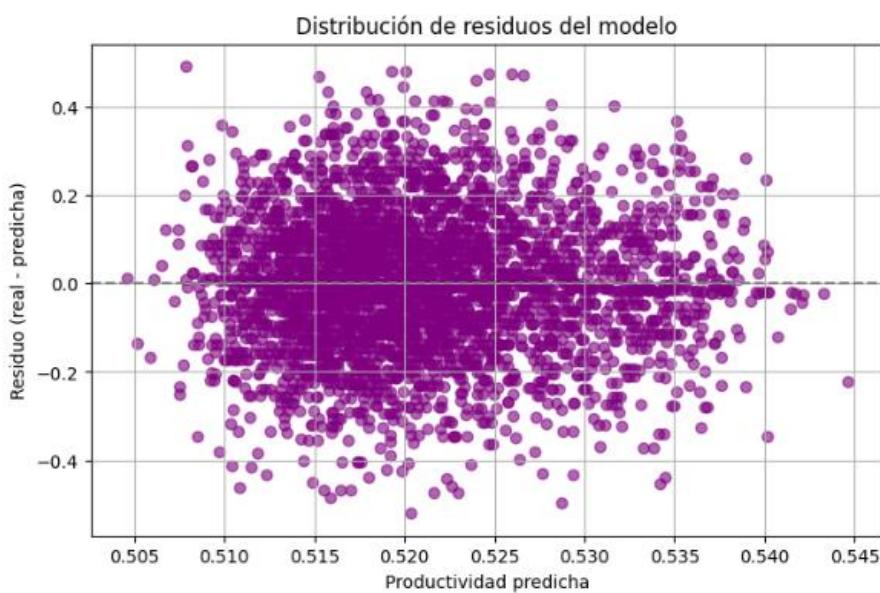
- $R^2 = 0.821$ , lo que indica que explica el 82% de la variabilidad en productividad.
- Todos los predictores son **estadísticamente significativos** ( $p < 0.01$ ).
- Interpretación:
  - Más horas de sueño → mayor productividad.
  - Más estrés y más dependencia digital → menor productividad.



## 4.4. Validación del modelo

### 4.4.1. Gráfico de residuos

El gráfico de residuos muestra errores distribuidos de forma simétrica y centrados en cero.



Muestra la diferencia entre los valores reales y los valores predichos de productividad, en función de las predicciones. Aquí tienes una interpretación detallada:

### ¿Qué representa este gráfico?

- Eje X: valores de productividad predicha por el modelo.
- Eje Y: residuos, es decir, real-predicha
- Puntos morados: cada uno representa una observación del modelo.

### Conclusiones estadísticas del gráfico

#### 1. Residuos centrados en cero

- La mayoría de los puntos están distribuidos alrededor de la línea horizontal ( $y = 0$ ).
- Esto indica que el modelo no tiene sesgo sistemático: no tiende a sobreestimar ni subestimar la productividad en general.

#### 2. Distribución simétrica y dispersión moderada

- Los residuos están repartidos tanto por encima como por debajo de cero, sin acumulaciones extremas.
- Esto sugiere que los errores del modelo son aleatorios y homogéneos, lo cual es deseable en regresión lineal.

#### 3. No hay patrón curvo ni efecto de heterocedasticidad

- No se observa una forma de abanico ni curvatura en los residuos.
- Esto indica que la varianza de los errores es constante a lo largo de las predicciones, cumpliendo el supuesto de homocedasticidad.

### Diagnóstico final

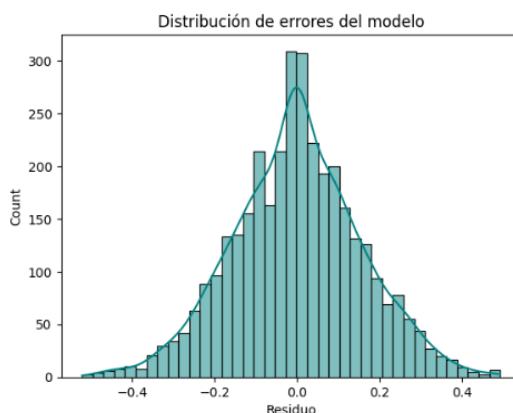
El gráfico de residuos confirma que:

- El modelo de regresión múltiple está bien ajustado.
- Los errores son aleatorios, simétricos y sin sesgo.
- No hay evidencia de problemas estructurales como heterocedasticidad o no linealidad.

Esto refuerza la validez del modelo que ya mostraba un  $R^2$  alto y predictores significativos.

#### 4.4.2. Histograma de residuos

El histograma de residuos confirma la normalidad de los errores.



El gráfico histograma de residuos del modelo es una herramienta clave para evaluar si los errores del modelo de regresión múltiple cumplen los supuestos estadísticos. Veamos la interpretación completa:

### ¿Qué representa este gráfico?

- Eje X: valores de los residuos, es decir, la diferencia entre la productividad real y la predicha.
- Eje Y: frecuencia (cuántas observaciones tienen ese nivel de error).
- Curva KDE: estimación de densidad que suaviza la forma de la distribución.

### Conclusiones estadísticas del gráfico

#### 1. Distribución centrada en cero

- La mayoría de los residuos se agrupan alrededor de 0.
- Esto indica que el modelo no tiene sesgo sistemático: no tiende a sobreestimar ni subestimar la productividad de forma constante.

#### 2. Forma aproximadamente normal

- La curva de densidad tiene forma de campana, lo que sugiere que los errores están distribuidos normalmente.
- Este es uno de los supuestos clave de la regresión lineal clásica (normalidad de los errores).

#### 3. Simetría y ausencia de colas extremas

- No se observan colas largas ni asimetrías marcadas.
- Esto indica que no hay valores atípicos extremos que distorsionen el modelo.

### Diagnóstico final

El histograma confirma que:

- Los errores del modelo son pequeños, simétricos y distribuidos normalmente.
- El modelo cumple con los supuestos de regresión lineal, lo que refuerza la validez de los resultados obtenidos.
- Puedes confiar en las inferencias estadísticas (valores p, coeficientes, R<sup>2</sup>) sin necesidad de aplicar correcciones robustas.

## 5. Conclusiones: perfiles de riesgo, recomendaciones, limitaciones

### 5.1. Perfiles de riesgo

- Usuarios con **alta dependencia digital** y **pocas horas de sueño** presentan mayor riesgo de baja productividad.



- El **estrés** se confirma como un factor negativo relevante.
- No se observa una relación clara entre concentración y productividad, lo que sugiere que otros factores pueden estar mediando.

## 5.2. Recomendaciones

- Promover hábitos de sueño saludables como estrategia para mejorar el rendimiento.
- Diseñar intervenciones educativas para reducir la dependencia digital, especialmente en adolescentes.
- Incluir el análisis de estrés en programas de bienestar escolar y laboral.

## 5.3. Limitaciones

- El estudio es transversal y no permite establecer causalidad.
- Las variables se basan en autoinforme, lo que puede introducir sesgos.
- No se han incluido todos los posibles factores contextuales (tipo de trabajo, entorno familiar, etc.).

## 6. Bibliografía

- Junta de Andalucía. (2023). *Curso para profesorado: Desarrollo de aplicaciones en lenguaje Python. Segunda parte* (262318FP002). Formación del profesorado en FP. <https://formacionprofesorado.ced.junta-andalucia.es>
- Kaggle. (2022). *Digital Health and Mental Wellness Dataset*. <https://www.kaggle.com/datasets>
- Seaborn Development Team. (2024). *Seaborn Statistical Data Visualization*. Documentación oficial. <https://seaborn.pydata.org>
- McKinney, W. (2017). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. O'Reilly Media.
- Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics*. Sage Publications.
- Ministerio de Educación y Formación Profesional. (2021). *Marco de competencia digital docente*. <https://educaciondigital.mecd.gob.es>
- Microsoft Copilot. (2025). *Asistente de inteligencia artificial para análisis estadístico, documentación pedagógica y visualización de datos*. Disponible en <https://copilot.microsoft.com>