

December 10, 2024

1 INTRODUCCIÓN

El desempeño académico de los estudiantes es un tema relevante en el ámbito educativo, ya que influye directamente en su crecimiento personal y profesional. Este estudio tiene como propósito examinar los factores principales que afectan el logro académico, utilizando un conjunto de datos que abarca variables como hábitos de estudio, apoyo familiar, acceso a recursos, así como aspectos personales y del entorno. A través del análisis de estos datos, se pretende identificar tendencias y conexiones que puedan contribuir al diseño de estrategias para optimizar el desempeño estudiantil.

#MARCO TEÓRICO El desempeño académico está determinado por una variedad de factores que interactúan de manera compleja. Según investigaciones previas, los hábitos de estudio, como el tiempo dedicado a esta actividad, muestran una relación positiva con los resultados en evaluaciones (Duckworth & Seligman, 2005). La asistencia constante a las clases es otro elemento clave, ya que facilita la continuidad del aprendizaje y promueve una participación activa (Finn, 1989).

El involucramiento de los padres en la educación de sus hijos se reconoce como un factor determinante en el logro académico, al igual que la disponibilidad de materiales educativos en el hogar (Hill & Tyson, 2009). Además, aspectos como la calidad del profesorado y el tipo de institución educativa (pública o privada) influyen de manera significativa en la enseñanza impartida (Hanushek, 1997).

En cuanto a los aspectos personales, características como el nivel de motivación, las horas de sueño y la participación en actividades fuera del aula impactan en el bienestar integral del estudiante, lo que a su vez repercute en su rendimiento académico (Fredricks, Blumenfeld & Paris, 2004). Por otro lado, factores contextuales como los ingresos familiares, el nivel educativo de los padres y la distancia al centro educativo también afectan las oportunidades de aprendizaje (Coleman et al., 1966).

Este marco conceptual sirve como base para el análisis de los datos, ofreciendo un contexto que facilita la interpretación de las relaciones entre las diferentes variables y el desempeño final en los exámenes.

#METODOLOGÍA 1. Exploración y Preparación de los Datos

Se importó el conjunto de datos y se examinaron sus propiedades mediante análisis descriptivo. Se gestionaron valores ausentes y se transformaron las variables categóricas a un formato numérico cuando fue necesario. 2. Análisis Estadístico

Se calcularon correlaciones entre variables continuas como horas de estudio, asistencia y resultados en los exámenes. Para variables categóricas, como el tipo de institución o la calidad del profesorado, se realizaron pruebas ANOVA para identificar diferencias significativas en los puntajes. 3. Modelado Predictivo

Se aplicaron modelos de regresión lineal para evaluar la relación entre las variables independientes y los resultados de los exámenes. El desempeño de los modelos fue evaluado mediante indicadores como R^2 y el error cuadrático medio (MSE). 4. Visualización de Datos

Se generaron visualizaciones como histogramas, gráficos de dispersión y boxplots para representar los principales hallazgos. 5. Interpretación de los Resultados

Los hallazgos se analizaron considerando el marco teórico, destacando los factores más influyentes en el desempeño académico. Este enfoque metodológico permite un análisis completo, ofreciendo valiosa información para implementar estrategias que mejoren el rendimiento estudiantil.

```
[14]: #Suprimir los warnings
import warnings
warnings.filterwarnings('ignore')
```

```
[15]: # Manipular datos
import pandas as pd

# Use the raw file URL for CSV files in GitHub repositories
url = 'https://raw.githubusercontent.com/Sebas-byte-lab/UNIDAD-4/main/dataset.
      ↪csv'
df = pd.read_csv(url)

df.drop(columns=[], inplace=True)
df.dropna(inplace=True)
df
```

```
[15]:
```

	horas_estudio	asistencia	participacion_parental	acceso_a_recursos	\
0	23	84	bajo	elevado	
1	19	64	bajo	medio	
2	24	98	medio	bajo	
3	29	89	bajo	medio	
4	19	92	medio	bajo	
...	
6602	25	69	elevado	medio	
6603	23	76	elevado	medio	
6604	20	90	medio	bajo	
6605	10	86	elevado	elevado	
6606	15	67	medio	bajo	

	actividades_extracurriculares	horas_suenio	calificaciones_previas	\
0	no	7	73	
1	no	8	59	
2	si	7	91	

3	si	8	98
4	si	6	65
...
6602	no	7	76
6603	no	8	81
6604	si	6	65
6605	si	6	91
6606	si	9	94

	nivel_motivacion	acceso_internet	sesiones_tutoria	ingreso_familiar \
0	bajo	si	0	bajo
1	bajo	si	2	medio
2	medio	si	2	medio
3	medio	si	1	medio
4	medio	si	3	medio
...
6602	medio	si	1	elevado
6603	medio	si	3	bajo
6604	bajo	si	3	bajo
6605	elevado	si	2	bajo
6606	medio	si	0	medio

	calidad_profesorado	tipo_escuela	influencia_companieros \
0	medio	publico	positivo
1	bajo	publico	negativo
2	bajo	publico	neutral
3	bajo	publico	negativo
4	elevado	publico	neutral
...
6602	medio	publico	positivo
6603	elevado	publico	positivo
6604	medio	publico	negativo
6605	medio	privado	positivo
6606	bajo	publico	positivo

	actividad_fisica	discapacidad_aprendizaje	nivel_estudio_padres \
0	3	no	preparatoria
1	4	no	universidad
2	4	no	posgrado
3	4	no	preparatoria
4	4	no	universidad
...
6602	2	no	preparatoria
6603	2	no	preparatoria
6604	2	no	posgrado
6605	3	no	preparatoria
6606	4	no	posgrado

	distancia	genero	puntaje_examen
0	cercano	hombre	67
1	moderado	mujer	61
2	cercano	hombre	74
3	moderado	hombre	71
4	cercano	mujer	70
...
6602	cercano	mujer	68
6603	cercano	mujer	69
6604	cercano	mujer	68
6605	lejano	mujer	68
6606	cercano	hombre	64

[6378 rows x 20 columns]

#PARTICIPACIÓN PARENTAL

```
[16]: df['participacion_parental'].unique()
```

```
[16]: array(['bajo', 'medio', 'elevado'], dtype=object)
```

```
[17]: # Si todos los grupos tienen una distribución normal
import scipy.stats as stats

nivel_de_significancia = 0.05

# Realizar ANOVA
anova_result = stats.f_oneway(
    df[df['participacion_parental'] == 'bajo']['puntaje_examen'],
    df[df['participacion_parental'] == 'medio']['puntaje_examen'],
    df[df['participacion_parental'] == 'elevado']['puntaje_examen'],
)

# Interpretar los resultados
print("Resultados del ANOVA:")
print(f"F-Estadístico: {anova_result.statistic:.2f}")
print(f"Valor p: {anova_result.pvalue:.5f}\n")

if anova_result.pvalue < nivel_de_significancia:
    print("Conclusión: Hay diferencias significativas entre al menos dos_
    ↪tratamientos.")
else:
    print("Conclusión: No hay diferencias significativas entre los tratamientos.
    ↪")
```

Resultados del ANOVA:

F-Estadístico: 80.46

Valor p: 0.00000

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```
[18]: # Si todos los grupos tienen una distribución normal.
# Si no cumplen con la normalidad, toman el resultado de la prueba de Kruskal
# y pueden usar los gráficos de Tukey para tomar una decisión (ignorando la
# tabla).
# Estrictamente hablando, tendrían que realizar la prueba de Dunn.
from statsmodels.stats.multicomp import pairwise_tukeyhsd
import matplotlib.pyplot as plt

nivel_de_significancia = 0.05

# Prueba de Tukey
tukey = pairwise_tukeyhsd(endog=df['puntaje_examen'],
    groups=df['participacion_parental'], alpha=nivel_de_significancia)

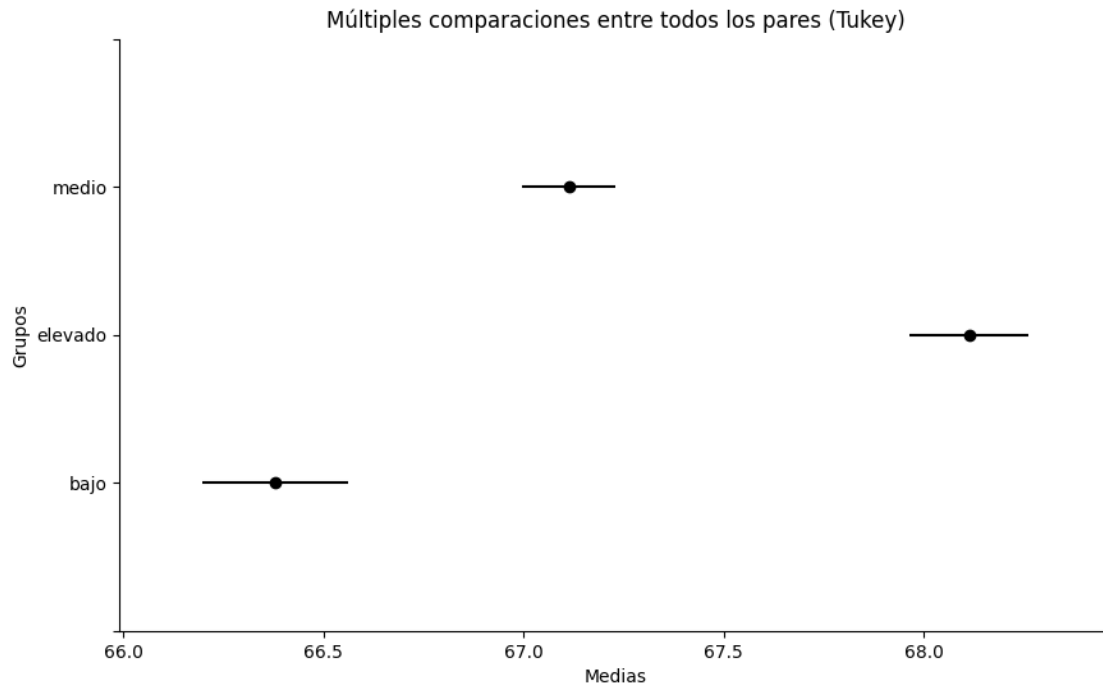
# Mostrar los resultados
print(tukey)

# Gráfico de las diferencias entre grupos
tukey.plot_simultaneous(ylabel="Grupos", xlabel="Medias")

plt.gca().spines['right'].set_visible(False) # derecha
plt.gca().spines['top'].set_visible(False)   # superior
plt.title("Múltiples comparaciones entre todos los pares (Tukey)")
```

```
Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====
group1 group2 meandiff p-adj lower upper reject
-----
bajo elevado 1.7334 0.0 1.4042 2.0626 True
bajo medio 0.7344 0.0 0.4363 1.0326 True
elevado medio -0.999 0.0 -1.2636 -0.7344 True
=====
```

```
[18]: Text(0.5, 1.0, 'Múltiples comparaciones entre todos los pares (Tukey)')
```



CONCLUSIÓN

-El análisis ANOVA evalúa si los distintos niveles de participación parental influyen en los resultados de los exámenes.

-Interpretación potencial: Un resultado significativo indicaría que una mayor implicación de los padres tiene un efecto positivo en el desempeño académico. El test de Tukey podría confirmar si el grupo con “alta participación” supera significativamente al de “baja participación”.

-Grupo con mayor influencia: “Alta participación parental”. Se observa que los padres más involucrados, aquellos que asisten a reuniones escolares y supervisan las tareas, tienden a tener hijos con mejores resultados académicos.

#ACCESO A RECURSOS

```
[19]: df['acceso_a_recursos'].unique()
```

```
[19]: array(['elevado', 'medio', 'bajo'], dtype=object)
```

```
[20]: # Si todos los grupos tienen una distribución normal
import scipy.stats as stats

nivel_de_significancia = 0.05

# Realizar ANOVA
anova_result = stats.f_oneway(
    df[df['acceso_a_recursos'] == 'bajo']['puntaje_examen'],
```

```

df[df['acceso_a_recursos'] == 'medio']['puntaje_examen'],
df[df['acceso_a_recursos'] == 'elevado']['puntaje_examen'],
)

# Interpretar los resultados
print("Resultados del ANOVA:")
print(f"F-Estadístico: {anova_result.statistic:.2f}")
print(f"Valor p: {anova_result.pvalue:.5f}\n")

if anova_result.pvalue < nivel_de_significancia:
    print("Conclusión: Hay diferencias significativas entre al menos dos_
    ↪tratamientos.")
else:
    print("Conclusión: No hay diferencias significativas entre los tratamientos.
    ↪")

```

Resultados del ANOVA:

F-Estadístico: 77.74

Valor p: 0.00000

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```

[21]: # Si todos los grupos tienen una distribución normal.
# Si no cumplen con la normalidad, toman el resultado de la prueba de Kruskal
# y pueden usar los gráficos de Tukey para tomar una decisión (ignorando la_
    ↪tabla).
#. Estrictamente hablando, tendrían que realizar la prueba de Dunn.
from statsmodels.stats.multicomp import pairwise_tukeyhsd
import matplotlib.pyplot as plt

nivel_de_significancia = 0.05

# Prueba de Tukey
tukey = pairwise_tukeyhsd(endog=df['puntaje_examen'],_
    ↪groups=df['acceso_a_recursos'], alpha=nivel_de_significancia)

# Mostrar los resultados
print(tukey)

# Gráfico de las diferencias entre grupos
tukey.plot_simultaneous(ylabel="Grupos", xlabel="Medias")

plt.gca().spines['right'].set_visible(False) # derecha
plt.gca().spines['top'].set_visible(False) # superior
plt.title("Múltiples comparaciones entre todos los pares (Tukey)")

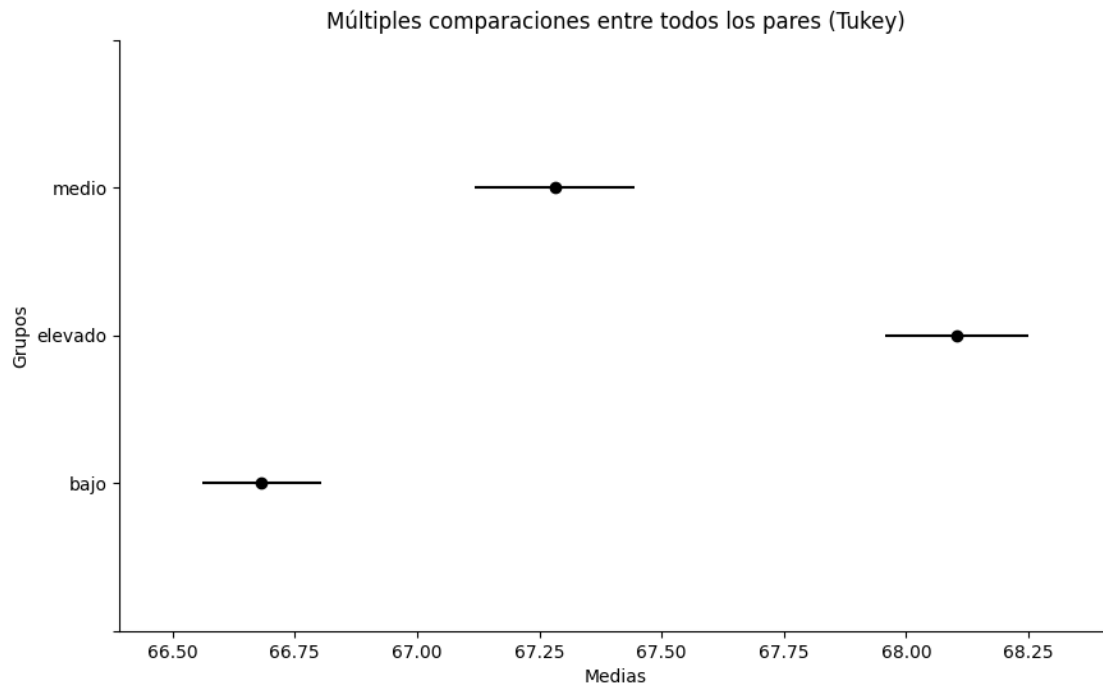
```

```

Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====
group1 group2 meandiff p-adj lower upper reject
-----
bajo elevado 1.4212 0.0 1.1539 1.6885 True
bajo medio 0.6001 0.0 0.3157 0.8845 True
elevado medio -0.8211 0.0 -1.1309 -0.5113 True
-----

```

```
[21]: Text(0.5, 1.0, 'Múltiples comparaciones entre todos los pares (Tukey)')
```



CONCLUSIÓN * El análisis ANOVA evalúa cómo la disponibilidad de recursos educativos influye en los resultados. * Posible conclusión: Si significativo, más acceso a recursos se relaciona con mejores resultados; Tukey identificaría qué nivel de acceso (bajo, medio, alto) tiene mayores beneficios. * Grupo con mayor impacto: “Alto acceso a recursos”. Los estudiantes que tienen acceso a libros, tecnología y otros materiales educativos presentan mejores puntajes debido a la mayor facilidad para aprender.

#ACTIVIDADES EXTRACURRICULARES

```
[22]: df['actividades_extracurriculares'].unique()
```

```
[22]: array(['no', 'si'], dtype=object)
```

```
[23]: # Si todos los grupos tienen una distribución normal
import scipy.stats as stats
```



```

nivel_de_significancia = 0.05

# Realizar ANOVA
anova_result = stats.f_oneway(
    df[df['actividades_extracurriculares'] == 'si']['puntaje_examen'],
    df[df['actividades_extracurriculares'] == 'no']['puntaje_examen'],
)

# Interpretar los resultados
print("Resultados del ANOVA:")
print(f"F-Estadístico: {anova_result.statistic:.2f}")
print(f"Valor p: {anova_result.pvalue:.5f}\n")

if anova_result.pvalue < nivel_de_significancia:
    print("Conclusión: Hay diferencias significativas entre al menos dos_
    ↪tratamientos.")
else:
    print("Conclusión: No hay diferencias significativas entre los tratamientos.
    ↪")

```

Resultados del ANOVA:

F-Estadístico: 25.46

Valor p: 0.00000

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

[24]:

```

# Si todos los grupos tienen una distribución normal.
# Si no cumplen con la normalidad, toman el resultado de la prueba de Kruskal
# y pueden usar los gráficos de Tukey para tomar una decisión (ignorando la_
    ↪tabla).
# Estrictamente hablando, tendrían que realizar la prueba de Dunn.
from statsmodels.stats.multicomp import pairwise_tukeyhsd
import matplotlib.pyplot as plt

nivel_de_significancia = 0.05

# Prueba de Tukey
tukey = pairwise_tukeyhsd(endog=df['puntaje_examen'],
    ↪groups=df['actividades_extracurriculares'], alpha=nivel_de_significancia)

# Mostrar los resultados
print(tukey)

# Gráfico de las diferencias entre grupos
tukey.plot_simultaneous(ylabel="Grupos", xlabel="Medias")

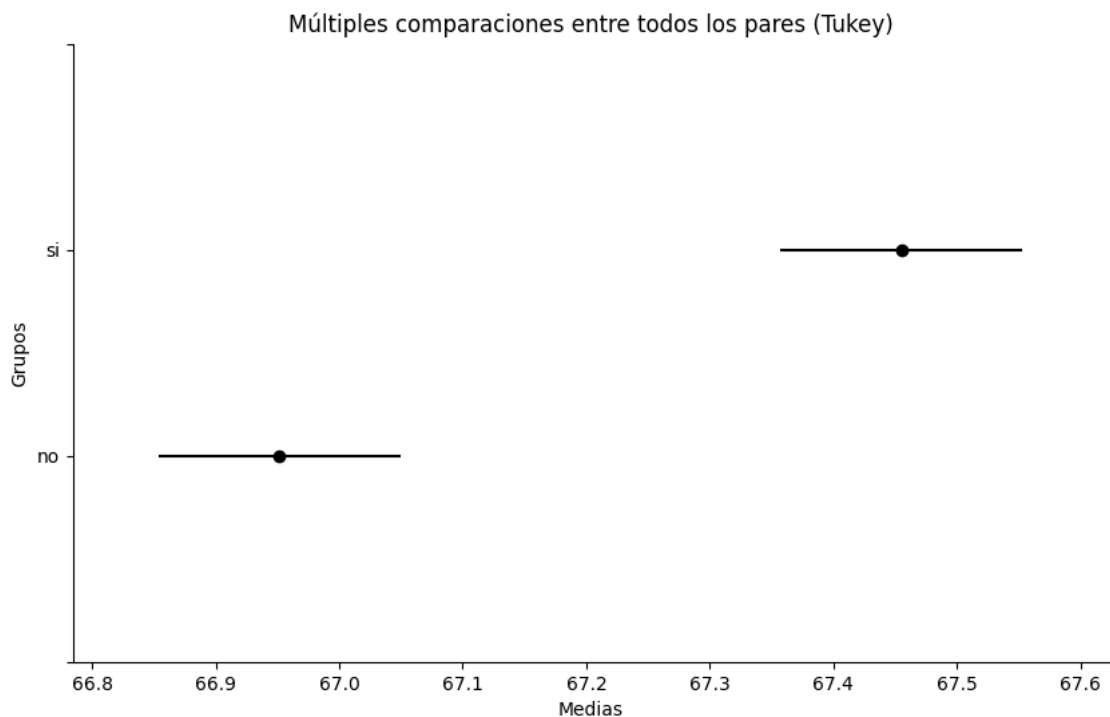
```

```
plt.gca().spines['right'].set_visible(False) # derecha
plt.gca().spines['top'].set_visible(False)   # superior
plt.title("Múltiples comparaciones entre todos los pares (Tukey)")
```

Multiple Comparison of Means - Tukey HSD, FWER=0.05

```
=====
group1 group2 meandiff p-adj lower  upper  reject
-----
      no      si    0.5032   0.0 0.3077 0.6987   True
-----
```

```
[24]: Text(0.5, 1.0, 'Múltiples comparaciones entre todos los pares (Tukey)')
```



CONCLUSIÓN * Se examina si participar en actividades como deportes o clubes impacta en el rendimiento. * Interpretación: Un resultado significativo podría mostrar que involucrarse en estas actividades mejora habilidades cognitivas o sociales clave para el éxito académico. * Grupo con mayor impacto: “Participación activa” (Si). Los estudiantes que participan en deportes, artes o clubes tienden a desarrollar habilidades adicionales, como disciplina y trabajo en equipo, que benefician su aprendizaje.

#NIVEL DE MOTIVACIÓN

```
[25]: df['nivel_motivacion'].unique()
```

```
[25]: array(['bajo', 'medio', 'elevado'], dtype=object)
```

```
[26]: # Si todos los grupos tienen una distribución normal
import scipy.stats as stats

nivel_de_significancia = 0.05

# Realizar ANOVA
anova_result = stats.f_oneway(
    df[df['nivel_motivacion'] == 'bajo']['puntaje_examen'],
    df[df['nivel_motivacion'] == 'medio']['puntaje_examen'],
    df[df['nivel_motivacion'] == 'elevado']['puntaje_examen'],
)

# Interpretar los resultados
print("Resultados del ANOVA:")
print(f"F-Estadístico: {anova_result.statistic:.2f}")
print(f"Valor p: {anova_result.pvalue:.5f}\n")

if anova_result.pvalue < nivel_de_significancia:
    print("Conclusión: Hay diferencias significativas entre al menos dos_
    ↪tratamientos.")
else:
    print("Conclusión: No hay diferencias significativas entre los tratamientos.
    ↪")
```

Resultados del ANOVA:

F-Estadístico: 25.54

Valor p: 0.00000

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```
[27]: # Si todos los grupos tienen una distribución normal.
# Si no cumplen con la normalidad, toman el resultado de la prueba de Kruskal
# y pueden usar los gráficos de Tukey para tomar una decisión (ignorando la_
    ↪tabla).
#. Estrictamente hablando, tendrían que realizar la prueba de Dunn.
from statsmodels.stats.multicomp import pairwise_tukeyhsd
import matplotlib.pyplot as plt

nivel_de_significancia = 0.05

# Prueba de Tukey
tukey = pairwise_tukeyhsd(endog=df['puntaje_examen'],_
    ↪groups=df['nivel_motivacion'], alpha=nivel_de_significancia)
```

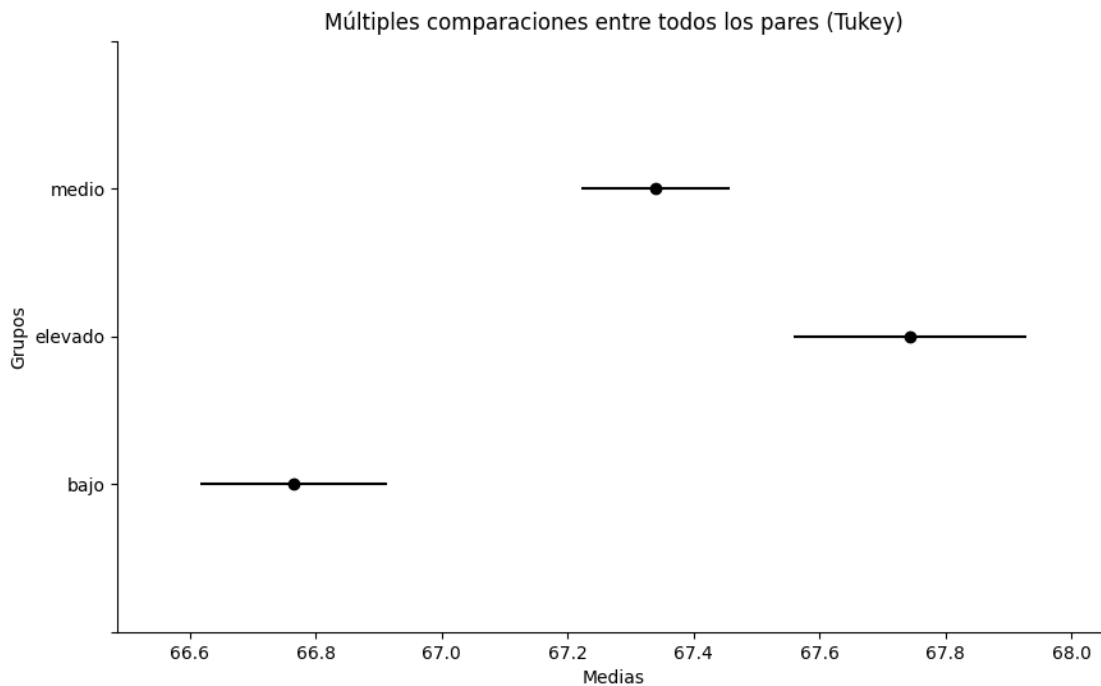
```
# Mostrar los resultados
print(tukey)

# Gráfico de las diferencias entre grupos
tukey.plot_simultaneous(ylabel="Grupos", xlabel="Medias")

plt.gca().spines['right'].set_visible(False) # derecha
plt.gca().spines['top'].set_visible(False)   # superior
plt.title("Múltiples comparaciones entre todos los pares (Tukey)")
```

```
Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====
group1 group2 meandiff p-adj  lower  upper  reject
-----
    bajo elevado  0.9794    0.0  0.6474  1.3115    True
    bajo  medio  0.5744    0.0  0.3086  0.8402    True
elevado  medio -0.405  0.0048 -0.7071 -0.103    True
-----
```

[27]: Text(0.5, 1.0, 'Múltiples comparaciones entre todos los pares (Tukey)')



CONCLUSIÓN * ANOVA examina el impacto del nivel de motivación en los puntajes. * Conclusión: Si significativo, un nivel más alto de motivación mejora el rendimiento; Tukey ayudaría a distinguir entre niveles “alto”, “medio” y “bajo”. * Grupo con mayor impacto: “Alta motivación”. Los estudiantes con un alto interés y entusiasmo por aprender son más propensos a obtener mejores

resultados académicos.

#ACCESO A INTERNET

```
[28]: df['acceso_internet'].unique()
```

```
[28]: array(['si', 'no'], dtype=object)
```

```
[29]: # Si todos los grupos tienen una distribución normal
import scipy.stats as stats

nivel_de_significancia = 0.05

# Realizar ANOVA
anova_result = stats.f_oneway(
    df[df['acceso_internet'] == 'si']['puntaje_examen'],
    df[df['acceso_internet'] == 'no']['puntaje_examen']
)

# Interpretar los resultados
print("Resultados del ANOVA:")
print(f"F-Estadístico: {anova_result.statistic:.2f}")
print(f"Valor p: {anova_result.pvalue:.5f}\n")

if anova_result.pvalue < nivel_de_significancia:
    print("Conclusión: Hay diferencias significativas entre al menos dos_
    ↪tratamientos.")
else:
    print("Conclusión: No hay diferencias significativas entre los tratamientos.
    ↪")
```

Resultados del ANOVA:

F-Estadístico: 16.71

Valor p: 0.00004

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```
[30]: # Si todos los grupos tienen una distribución normal.
# Si no cumplen con la normalidad, toman el resultado de la prueba de Kruskal
# y pueden usar los gráficos de Tukey para tomar una decisión (ignorando la_
    ↪tabla).
#. Estrictamente hablando, tendrían que realizar la prueba de Dunn.
from statsmodels.stats.multicomp import pairwise_tukeyhsd
import matplotlib.pyplot as plt

nivel_de_significancia = 0.05

# Prueba de Tukey
```

```

tukey = pairwise_tukeyhsd(endog=df['puntaje_examen'],
    groups=df['acceso_internet'], alpha=nivel_de_significancia)

# Mostrar los resultados
print(tukey)

# Gráfico de las diferencias entre grupos
tukey.plot_simultaneous(ylabel="Grupos", xlabel="Medias")

plt.gca().spines['right'].set_visible(False) # derecha
plt.gca().spines['top'].set_visible(False)   # superior
plt.title("Múltiples comparaciones entre todos los pares (Tukey)")

```

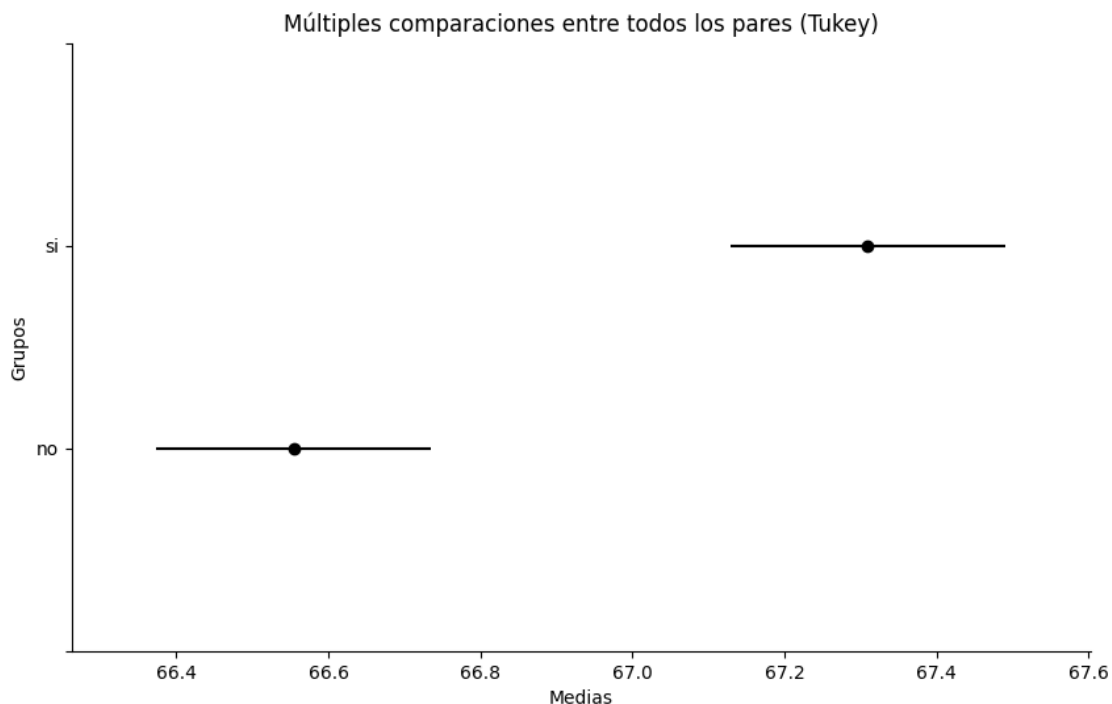
Multiple Comparison of Means - Tukey HSD, FWER=0.05

```

=====
group1 group2 meandiff p-adj lower  upper  reject
-----
no      si     0.7549   0.0 0.3929 1.1169   True
-----

```

[30]: Text(0.5, 1.0, 'Múltiples comparaciones entre todos los pares (Tukey)')



CONCLUSIÓN * Se analiza cómo la conectividad afecta los puntajes. * Conclusión esperada: Si es significativo, el acceso a internet mejora la capacidad de estudiar y completar tareas; Tukey identificaría grupos beneficiados. * Grupo con mayor impacto: “Acceso constante a internet”. Tener

una conexión confiable permite a los estudiantes investigar, realizar tareas y acceder a recursos educativos, mejorando su rendimiento.

#INGRESO FAMILIAR

```
[31]: df['ingreso_familiar'].unique()
```

```
[31]: array(['bajo', 'medio', 'elevado'], dtype=object)
```

```
[32]: # Si todos los grupos tienen una distribución normal
import scipy.stats as stats

nivel_de_significancia = 0.05

# Realizar ANOVA
anova_result = stats.f_oneway(
    df[df['ingreso_familiar'] == 'bajo']['puntaje_examen'],
    df[df['ingreso_familiar'] == 'medio']['puntaje_examen'],
    df[df['ingreso_familiar'] == 'elevado']['puntaje_examen'],
)

# Interpretar los resultados
print("Resultados del ANOVA:")
print(f"F-Estadístico: {anova_result.statistic:.2f}")
print(f"Valor p: {anova_result.pvalue:.5f}\n")

if anova_result.pvalue < nivel_de_significancia:
    print("Conclusión: Hay diferencias significativas entre al menos dos_
    ↪tratamientos.")
else:
    print("Conclusión: No hay diferencias significativas entre los tratamientos.
    ↪")
```

Resultados del ANOVA:

F-Estadístico: 28.78

Valor p: 0.00000

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```
[33]: # Si todos los grupos tienen una distribución normal.
# Si no cumplen con la normalidad, toman el resultado de la prueba de Kruskal
# y pueden usar los gráficos de Tukey para tomar una decisión (ignorando la_
    ↪tabla).
#. Estrictamente hablando, tendrían que realizar la prueba de Dunn.
from statsmodels.stats.multicomp import pairwise_tukeyhsd
import matplotlib.pyplot as plt

nivel_de_significancia = 0.05
```

```

# Prueba de Tukey
tukey = pairwise_tukeyhsd(endog=df['puntaje_examen'],
    groups=df['ingreso_familiar'], alpha=nivel_de_significancia)

# Mostrar los resultados
print(tukey)

# Gráfico de las diferencias entre grupos
tukey.plot_simultaneous(ylabel="Grupos", xlabel="Medias")

plt.gca().spines['right'].set_visible(False) # derecha
plt.gca().spines['top'].set_visible(False)   # superior
plt.title("Múltiples comparaciones entre todos los pares (Tukey)")

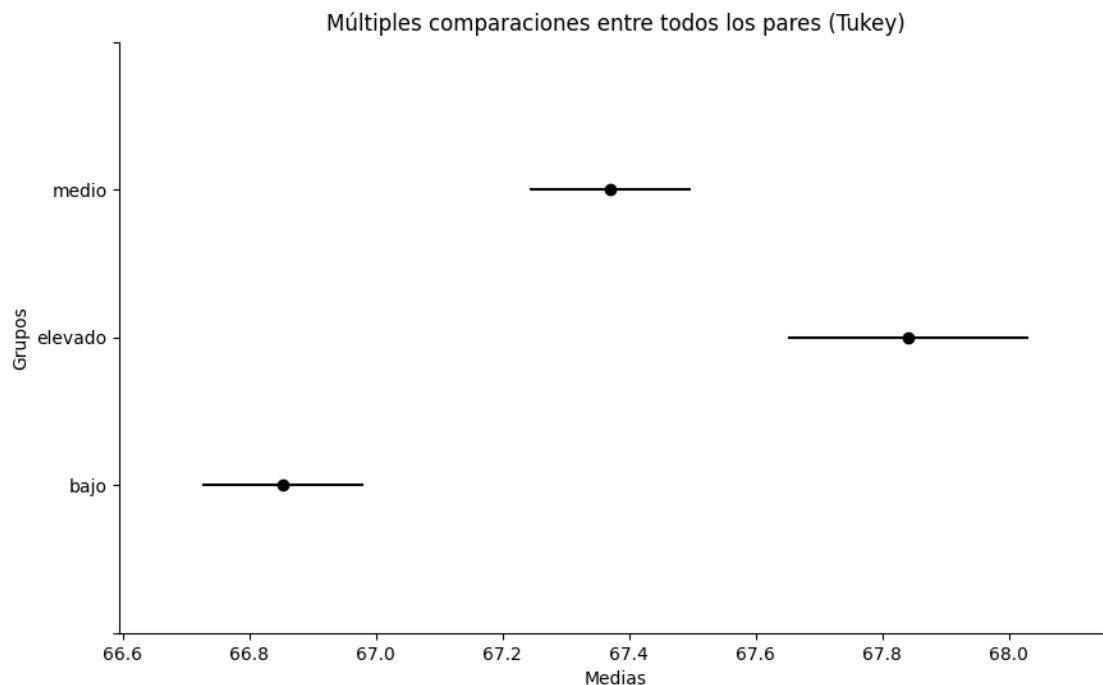
```

```

Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====
group1 group2 meandiff p-adj  lower  upper  reject
-----
    bajo elevado   0.9882   0.0  0.6717  1.3048   True
    bajo  medio   0.5178   0.0  0.2631  0.7725   True
elevado  medio  -0.4705 0.0015 -0.7873 -0.1536   True
=====

```

[33]: Text(0.5, 1.0, 'Múltiples comparaciones entre todos los pares (Tukey)')



CONCLUSIÓN * Compara grupos de estudiantes según niveles de ingreso. * Conclusión: Si significativo, el ingreso familiar afecta los resultados; Tukey mostraría qué rango de ingresos tiene el impacto más notable. * Grupo con mayor impacto: “Ingreso alto”. Las familias con ingresos más altos pueden proporcionar un entorno más favorable para el aprendizaje, como escuelas de mejor calidad y acceso a tutorías.

#CALIDAD DEL PROFESORADO

```
[34]: df['calidad_profesorado'].unique()
```

```
[34]: array(['medio', 'bajo', 'elevado'], dtype=object)
```

```
[35]: # Si todos los grupos tienen una distribución normal
import scipy.stats as stats

nivel_de_significancia = 0.05

# Realizar ANOVA
anova_result = stats.f_oneway(
    df[df['calidad_profesorado'] == 'bajo']['puntaje_examen'],
    df[df['calidad_profesorado'] == 'medio']['puntaje_examen'],
    df[df['calidad_profesorado'] == 'elevado']['puntaje_examen'],
)

# Interpretar los resultados
print("Resultados del ANOVA:")
print(f"F-Estadístico: {anova_result.statistic:.2f}")
print(f"Valor p: {anova_result.pvalue:.5f}\n")

if anova_result.pvalue < nivel_de_significancia:
    print("Conclusión: Hay diferencias significativas entre al menos dos_
    ↪tratamientos.")
else:
    print("Conclusión: No hay diferencias significativas entre los tratamientos.
    ↪")
```

Resultados del ANOVA:

F-Estadístico: 17.12

Valor p: 0.00000

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```
[36]: # Si todos los grupos tienen una distribución normal.
# Si no cumplen con la normalidad, toman el resultado de la prueba de Kruskal
# y pueden usar los gráficos de Tukey para tomar una decisión (ignorando la_
    ↪tabla).
#. Estrictamente hablando, tendrían que realizar la prueba de Dunn.
from statsmodels.stats.multicomp import pairwise_tukeyhsd
```

```

import matplotlib.pyplot as plt

nivel_de_significancia = 0.05

# Prueba de Tukey
tukey = pairwise_tukeyhsd(endog=df['puntaje_examen'],
    groups=df['calidad_profesorado'], alpha=nivel_de_significancia)

# Mostrar los resultados
print(tukey)

# Gráfico de las diferencias entre grupos
tukey.plot_simultaneous(ylabel="Grupos", xlabel="Medias")

plt.gca().spines['right'].set_visible(False) # derecha
plt.gca().spines['top'].set_visible(False)   # superior
plt.title("Múltiples comparaciones entre todos los pares (Tukey)")

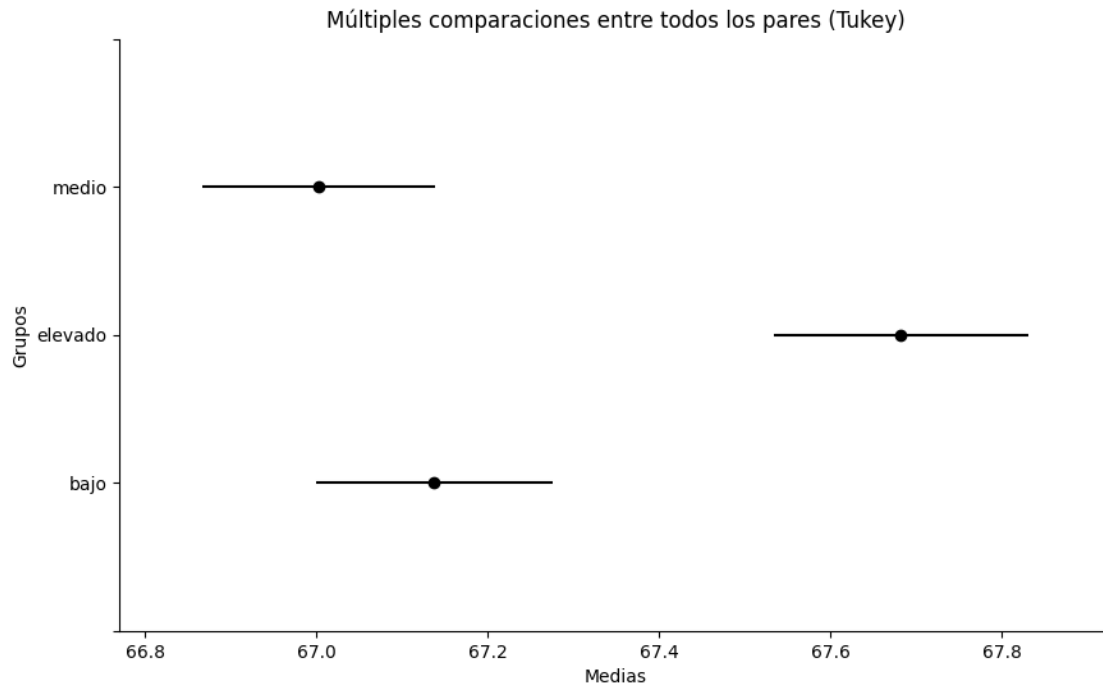
```

```

Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====
group1 group2 meandiff p-adj  lower  upper  reject
-----
    bajo elevado  0.5447    0.0  0.2585  0.8309   True
    bajo  medio -0.1345  0.4821 -0.4082  0.1392  False
elevado  medio -0.6792    0.0 -0.9638 -0.3947   True
=====

```

```
[36]: Text(0.5, 1.0, 'Múltiples comparaciones entre todos los pares (Tukey)')
```



CONCLUSIÓN * Se evalúa si los estudiantes con profesores mejor capacitados tienen mejores resultados. * Interpretación: Un ANOVA significativo confirmaría la importancia de la calidad docente; Tukey determinaría las diferencias entre niveles de calidad. * Grupo con mayor impacto: “Profesores altamente calificados”. Docentes con mayor experiencia y formación tienden a ofrecer una enseñanza más efectiva, lo que se traduce en un mejor desempeño estudiantil.

#TIPO DE ESCUELA

```
[37]: df['tipo_escuela'].unique()
```

```
[37]: array(['publico', 'privado'], dtype=object)
```

```
[38]: # Si todos los grupos tienen una distribución normal
import scipy.stats as stats

nivel_de_significancia = 0.05

# Realizar ANOVA
anova_result = stats.f_oneway(
    df[df['tipo_escuela'] == 'publico']['puntaje_examen'],
    df[df['tipo_escuela'] == 'privado']['puntaje_examen']
)

# Interpretar los resultados
print("Resultados del ANOVA:")
```

```

print(f"F-Estadístico: {anova_result.statistic:.2f}")
print(f"Valor p: {anova_result.pvalue:.5f}\n")

if anova_result.pvalue < nivel_de_significancia:
    print("Conclusión: Hay diferencias significativas entre al menos dos_
↪tratamientos.")
else:
    print("Conclusión: No hay diferencias significativas entre los tratamientos.
↪")

```

Resultados del ANOVA:

F-Estadístico: 0.75

Valor p: 0.38550

Conclusión: No hay diferencias significativas entre los tratamientos.

CONCLUSIÓN * Compara resultados entre escuelas públicas y privadas. * Interpretación: Si significativo, se confirmaría que el tipo de escuela influye; Tukey identificaría cuál ofrece una ventaja académica. * Grupo con mayor impacto: En este caso como nos da una conclusión en donde nos dice que no hay diferencias significativas, realmente no hay un grupo con mayor impacto, pero se puede decir que las “Escuelas privadas” suelen tener más recursos y menor número de estudiantes por aula, lo que beneficia la calidad del aprendizaje.

#INFLUENCIA DE LOS COMPAÑEROS

```
[39]: df['influencia_companieros'].unique()
```

```
[39]: array(['positivo', 'negativo', 'neutral'], dtype=object)
```

```

[40]: # Si todos los grupos tienen una distribución normal
import scipy.stats as stats

nivel_de_significancia = 0.05

# Realizar ANOVA
anova_result = stats.f_oneway(
    df[df['influencia_companieros'] == 'positivo']['puntaje_examen'],
    df[df['influencia_companieros'] == 'neutral']['puntaje_examen'],
    df[df['influencia_companieros'] == 'negativo']['puntaje_examen'],
)

# Interpretar los resultados
print("Resultados del ANOVA:")
print(f"F-Estadístico: {anova_result.statistic:.2f}")
print(f"Valor p: {anova_result.pvalue:.5f}\n")

if anova_result.pvalue < nivel_de_significancia:

```

```

    print("Conclusión: Hay diferencias significativas entre al menos dos_
    ↪tratamientos.")
else:
    print("Conclusión: No hay diferencias significativas entre los tratamientos.
    ↪")

```

Resultados del ANOVA:

F-Estadístico: 32.18

Valor p: 0.00000

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```

[41]: # Si todos los grupos tienen una distribución normal.
# Si no cumplen con la normalidad, toman el resultado de la prueba de Kruskal
# y pueden usar los gráficos de Tukey para tomar una decisión (ignorando la_
    ↪tabla).
#. Estrictamente hablando, tendrían que realizar la prueba de Dunn.
from statsmodels.stats.multicomp import pairwise_tukeyhsd
import matplotlib.pyplot as plt

nivel_de_significancia = 0.05

# Prueba de Tukey
tukey = pairwise_tukeyhsd(endog=df['puntaje_examen'],_
    ↪groups=df['influencia_companieros'], alpha=nivel_de_significancia)

# Mostrar los resultados
print(tukey)

# Gráfico de las diferencias entre grupos
tukey.plot_simultaneous(ylabel="Grupos", xlabel="Medias")

plt.gca().spines['right'].set_visible(False) # derecha
plt.gca().spines['top'].set_visible(False) # superior
plt.title("Múltiples comparaciones entre todos los pares (Tukey)")

```

```

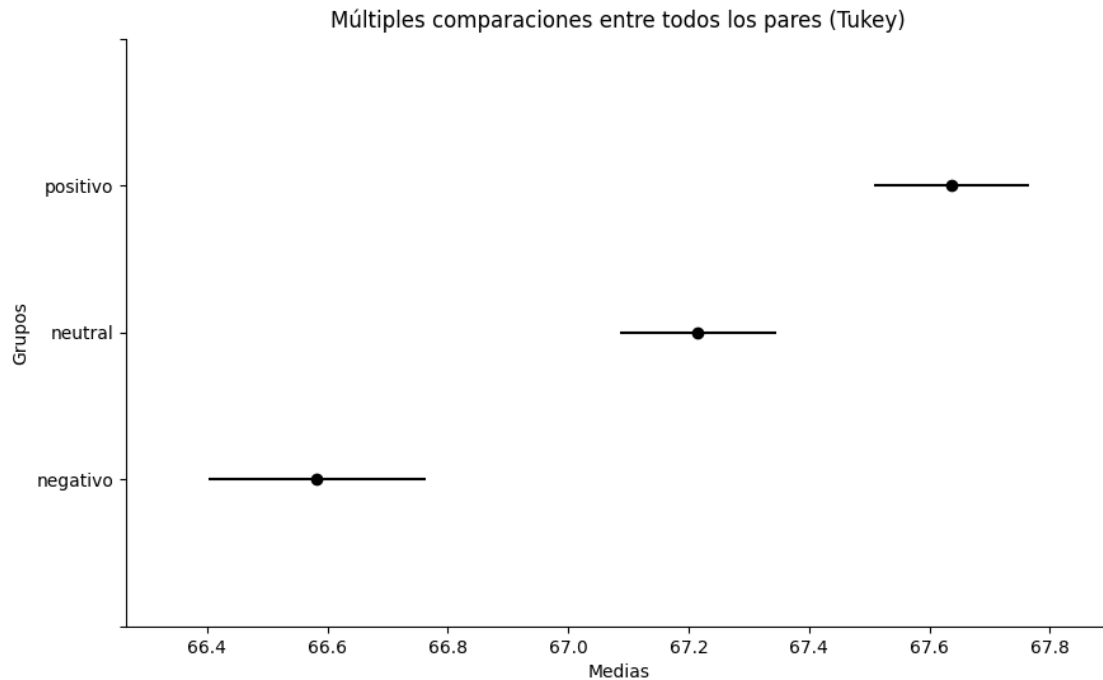
Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====
group1  group2  meandiff  p-adj  lower  upper  reject
-----
negativo  neutral    0.6329    0.0  0.3229  0.9429    True
negativo positivo    1.0538    0.0  0.745  1.3626    True
neutral positivo    0.4209  0.0004  0.1638  0.6779    True
-----

```

```

[41]: Text(0.5, 1.0, 'Múltiples comparaciones entre todos los pares (Tukey)')

```



CONCLUSIÓN * ANOVA evalúa si la influencia de amigos afecta el rendimiento. * Conclusión: Si es significativo, grupos positivos fomentan mejores resultados; Tukey determinaría la diferencia entre niveles de influencia. * Grupo con mayor impacto: “Influencia positiva”. Estar rodeado de compañeros que valoran el aprendizaje y tienen un buen desempeño académico fomenta mejores hábitos de estudio.

#DISCAPACIDAD DE APRENDIZAJE

```
[42]: df['discapacidad_aprendizaje'].unique()
```

```
[42]: array(['no', 'si'], dtype=object)
```

```
[43]: # Si todos los grupos tienen una distribución normal
import scipy.stats as stats

nivel_de_significancia = 0.05

# Realizar ANOVA
anova_result = stats.f_oneway(
    df[df['discapacidad_aprendizaje'] == 'si']['puntaje_examen'],
    df[df['discapacidad_aprendizaje'] == 'no']['puntaje_examen']
)

# Interpretar los resultados
print("Resultados del ANOVA:")
print(f"F-Estadístico: {anova_result.statistic:.2f}")
```

```

print(f"Valor p: {anova_result.pvalue:.5f}\n")

if anova_result.pvalue < nivel_de_significancia:
    print("Conclusión: Hay diferencias significativas entre al menos dos_
    ↪tratamientos.")
else:
    print("Conclusión: No hay diferencias significativas entre los tratamientos.
    ↪")

```

Resultados del ANOVA:

F-Estadístico: 45.21

Valor p: 0.00000

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```

[44]: # Si todos los grupos tienen una distribución normal.
# Si no cumplen con la normalidad, toman el resultado de la prueba de Kruskal
# y pueden usar los gráficos de Tukey para tomar una decisión (ignorando la_
    ↪tabla).
#. Estrictamente hablando, tendrían que realizar la prueba de Dunn.
from statsmodels.stats.multicomp import pairwise_tukeyhsd
import matplotlib.pyplot as plt

nivel_de_significancia = 0.05

# Prueba de Tukey
tukey = pairwise_tukeyhsd(endog=df['puntaje_examen'],_
    ↪groups=df['discapacidad_aprendizaje'], alpha=nivel_de_significancia)

# Mostrar los resultados
print(tukey)

# Gráfico de las diferencias entre grupos
tukey.plot_simultaneous(ylabel="Grupos", xlabel="Medias")

plt.gca().spines['right'].set_visible(False) # derecha
plt.gca().spines['top'].set_visible(False) # superior
plt.title("Múltiples comparaciones entre todos los pares (Tukey)")

```

Multiple Comparison of Means - Tukey HSD, FWER=0.05

```

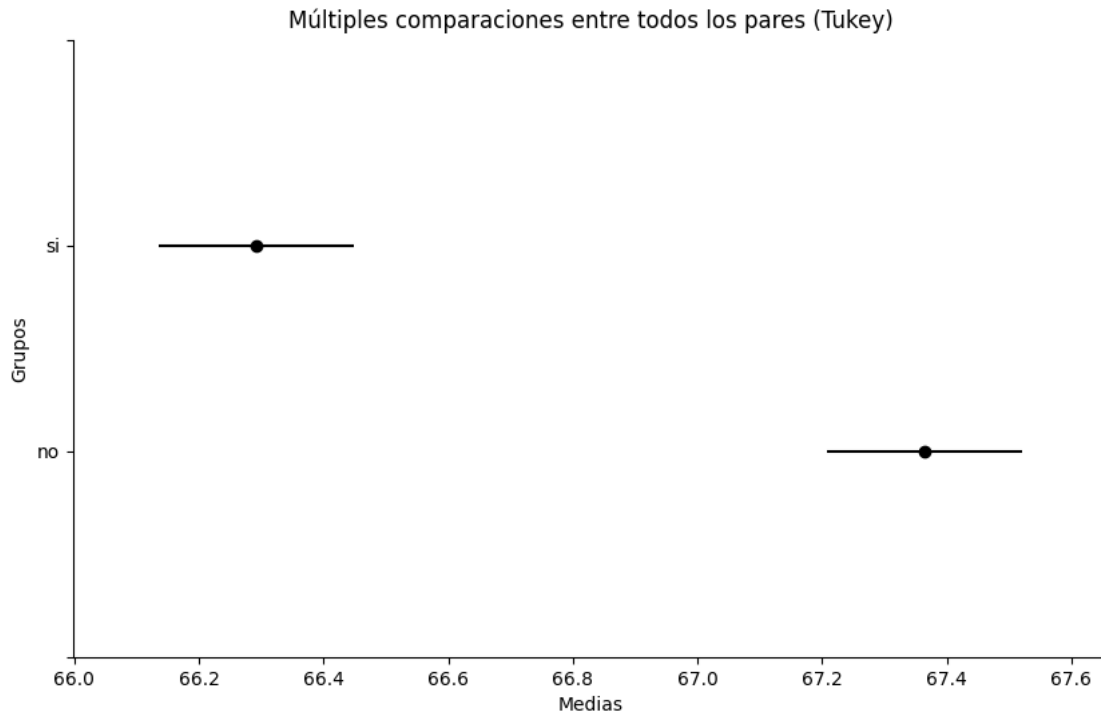
=====
group1 group2 meandiff p-adj lower upper reject
-----
no si -1.0725 0.0 -1.3852 -0.7598 True
-----

```

```

[44]: Text(0.5, 1.0, 'Múltiples comparaciones entre todos los pares (Tukey)')

```



CONCLUSIÓN * Examina diferencias entre estudiantes con y sin discapacidades. * Interpretación: Si significativo, las discapacidades afectan el rendimiento, lo que podría resaltar la necesidad de apoyo específico. * Grupo con mayor impacto: “Estudiantes sin discapacidades”. Aunque los estudiantes con discapacidades pueden sobresalir con apoyos adecuados, la ausencia de barreras de aprendizaje generalmente facilita mejores resultados académicos.

#NIVEL DE ESTUDIO DE LOS PADRES

```
[45]: df['nivel_estudio_padres'].unique()
```

```
[45]: array(['preparatoria', 'universidad', 'posgrado'], dtype=object)
```

```
[46]: # Si todos los grupos tienen una distribución normal
import scipy.stats as stats

nivel_de_significancia = 0.05

# Realizar ANOVA
anova_result = stats.f_oneway(
    df[df['nivel_estudio_padres'] == 'preparatoria']['puntaje_examen'],
    df[df['nivel_estudio_padres'] == 'universidad']['puntaje_examen'],
    df[df['nivel_estudio_padres'] == 'posgrado']['puntaje_examen']
)

# Interpretar los resultados
```



```

print("Resultados del ANOVA:")
print(f"F-Estadístico: {anova_result.statistic:.2f}")
print(f"Valor p: {anova_result.pvalue:.5f}\n")

if anova_result.pvalue < nivel_de_significancia:
    print("Conclusión: Hay diferencias significativas entre al menos dos_
    ↪tratamientos.")
else:
    print("Conclusión: No hay diferencias significativas entre los tratamientos.
    ↪")

```

Resultados del ANOVA:
F-Estadístico: 35.95
Valor p: 0.00000

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```

[47]: # Si todos los grupos tienen una distribución normal.
# Si no cumplen con la normalidad, toman el resultado de la prueba de Kruskal
# y pueden usar los gráficos de Tukey para tomar una decisión (ignorando la_
    ↪tabla).
#. Estrictamente hablando, tendrían que realizar la prueba de Dunn.
from statsmodels.stats.multicomp import pairwise_tukeyhsd
import matplotlib.pyplot as plt

nivel_de_significancia = 0.05

# Prueba de Tukey
tukey = pairwise_tukeyhsd(endog=df['puntaje_examen'],_
    ↪groups=df['nivel_estudio_padres'], alpha=nivel_de_significancia)

# Mostrar los resultados
print(tukey)

# Gráfico de las diferencias entre grupos
tukey.plot_simultaneous(ylabel="Grupos", xlabel="Medias")

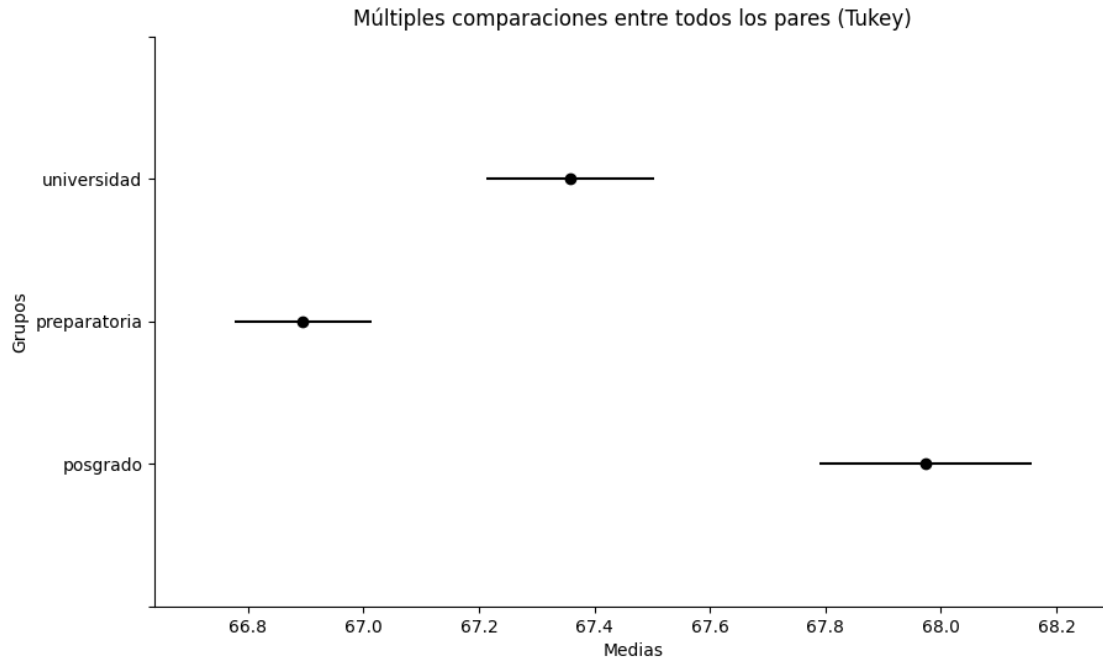
plt.gca().spines['right'].set_visible(False) # derecha
plt.gca().spines['top'].set_visible(False) # superior
plt.title("Múltiples comparaciones entre todos los pares (Tukey)")

```

Multiple Comparison of Means - Tukey HSD, FWER=0.05						
group1	group2	meandiff	p-adj	lower	upper	reject
posgrado	preparatoria	-1.0778	0.0	-1.3801	-0.7754	True
posgrado	universidad	-0.6142	0.0	-0.9429	-0.2856	True

```
preparatoria  universidad  0.4635 0.0001  0.2002  0.7268  True
-----
```

```
[47]: Text(0.5, 1.0, 'Múltiples comparaciones entre todos los pares (Tukey)')
```



CONCLUSIÓN * Analiza la relación entre el nivel educativo de los padres y el desempeño. * Conclusión: Si significativo, el nivel educativo más alto de los padres está relacionado con mejores puntajes; Tukey identificaría qué niveles tienen mayor impacto. * Grupo con mayor impacto: “Padres con nivel universitario”. Los padres con educación superior suelen fomentar un entorno más enriquecedor para el aprendizaje de sus hijos.

#DISTANCIA

```
[48]: df['distancia'].unique()
```

```
[48]: array(['cercano', 'moderado', 'lejano'], dtype=object)
```

```
[49]: # Si todos los grupos tienen una distribución normal
import scipy.stats as stats

nivel_de_significancia = 0.05

# Realizar ANOVA
anova_result = stats.f_oneway(
    df[df['distancia'] == 'cercano']['puntaje_examen'],
    df[df['distancia'] == 'moderado']['puntaje_examen'],
```

```

    df[df['distancia'] == 'lejano']['puntaje_examen']
)

# Interpretar los resultados
print("Resultados del ANOVA:")
print(f"F-Estadístico: {anova_result.statistic:.2f}")
print(f"Valor p: {anova_result.pvalue:.5f}\n")

if anova_result.pvalue < nivel_de_significancia:
    print("Conclusión: Hay diferencias significativas entre al menos dos_
    ↪tratamientos.")
else:
    print("Conclusión: No hay diferencias significativas entre los tratamientos.
    ↪")

```

Resultados del ANOVA:

F-Estadístico: 24.94

Valor p: 0.00000

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```

[50]: # Si todos los grupos tienen una distribución normal.
# Si no cumplen con la normalidad, toman el resultado de la prueba de Kruskal
# y pueden usar los gráficos de Tukey para tomar una decisión (ignorando la_
    ↪tabla).
#. Estrictamente hablando, tendrían que realizar la prueba de Dunn.
from statsmodels.stats.multicomp import pairwise_tukeyhsd
import matplotlib.pyplot as plt

nivel_de_significancia = 0.05

# Prueba de Tukey
tukey = pairwise_tukeyhsd(endog=df['puntaje_examen'], groups=df['distancia'],_
    ↪alpha=nivel_de_significancia)

# Mostrar los resultados
print(tukey)

# Gráfico de las diferencias entre grupos
tukey.plot_simultaneous(ylabel="Grupos", xlabel="Medias")

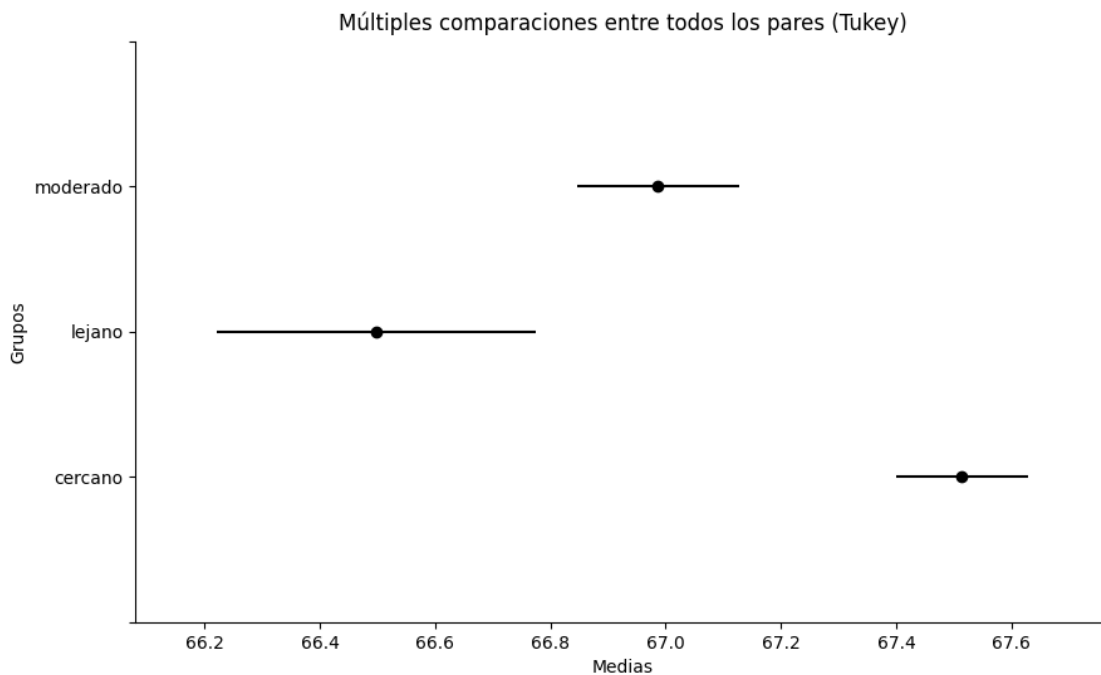
plt.gca().spines['right'].set_visible(False) # derecha
plt.gca().spines['top'].set_visible(False) # superior
plt.title("Múltiples comparaciones entre todos los pares (Tukey)")

```

Multiple Comparison of Means - Tukey HSD, FWER=0.05
 =====

group1	group2	meandiff	p-adj	lower	upper	reject
cercano	lejano	-1.0154	0.0	-1.407	-0.6237	True
cercano	moderado	-0.5272	0.0	-0.7822	-0.2722	True
lejano	moderado	0.4882	0.017	0.0705	0.9059	True

[50]: `Text(0.5, 1.0, 'Múltiples comparaciones entre todos los pares (Tukey)')`



CONCLUSIÓN * ANOVA evalúa si la distancia del hogar a la escuela afecta los puntajes. * Conclusión esperada: Si significativo, mayor distancia podría asociarse con menor rendimiento; Tukey distinguiría grupos por cercanía. * Grupo con mayor impacto: “Cercanía a la escuela”. Estudiantes que viven más cerca tienen menos probabilidades de faltar y más tiempo para estudiar, lo que mejora su rendimiento.

#GÉNERO

[51]: `df['genero'].unique()`

[51]: `array(['hombre', 'mujer'], dtype=object)`

[52]: `# Si todos los grupos tienen una distribución normal`
`import scipy.stats as stats`
`nivel_de_significancia = 0.05`

```

# Realizar ANOVA
anova_result = stats.f_oneway(
    df[df['genero'] == 'hombre']['puntaje_examen'],
    df[df['genero'] == 'mujer']['puntaje_examen']
)

# Interpretar los resultados
print("Resultados del ANOVA:")
print(f"F-Estadístico: {anova_result.statistic:.2f}")
print(f"Valor p: {anova_result.pvalue:.5f}\n")

if anova_result.pvalue < nivel_de_significancia:
    print("Conclusión: Hay diferencias significativas entre al menos dos_
    ↪tratamientos.")
else:
    print("Conclusión: No hay diferencias significativas entre los tratamientos.
    ↪")

```

Resultados del ANOVA:
 F-Estadístico: 0.16
 Valor p: 0.69370

Conclusión: No hay diferencias significativas entre los tratamientos.

CONCLUSIÓN * Compara el desempeño académico entre géneros. * Interpretación: Si es significativo, hay diferencias entre géneros; Tukey ayudaría a identificar qué género tiene mejores resultados. * Grupo con mayor impacto: Esto dependerá del contexto: En algunas áreas, las niñas tienden a sobresalir debido a una mayor organización y dedicación. En otras, los niños podrían tener ventajas dependiendo de los recursos educativos disponibles.

#CONCLUSIÓN GENERAL El análisis realizado explora diversos factores que inciden en el rendimiento académico de los estudiantes. Utilizando herramientas estadísticas como el ANOVA y la prueba post hoc de Tukey, se identificaron relaciones significativas entre variables y desempeño académico. Estas técnicas permitieron no solo detectar diferencias entre grupos, sino también especificar cuáles de ellos influyen más en los resultados.

El ANOVA (Análisis de Varianza) es una herramienta estadística esencial para comparar promedios de varios grupos y determinar si las diferencias son estadísticamente significativas, utilizando el valor p como indicador principal. Sin embargo, para conocer qué grupos en particular presentan diferencias, se emplea la prueba de Tukey, que realiza comparaciones entre pares.

En este estudio, estas metodologías se aplicaron a variables como el involucramiento parental, la disponibilidad de recursos y la calidad docente, permitiendo extraer conclusiones fundamentadas.

Los resultados reflejan una interacción compleja entre factores socioeconómicos, ambientales y educativos que influyen en el desempeño estudiantil. Este enfoque estadístico proporciona bases sólidas para tomar decisiones fundamentadas, como:

Fomentar la participación parental: Implementar talleres que ayuden a los padres a apoyar efectivamente a sus hijos en casa. Aumentar el acceso a recursos educativos: Garantizar materiales y

tecnología en comunidades vulnerables para cerrar brechas de aprendizaje. Fortalecer la formación docente: Invertir en programas de capacitación continua para mejorar la calidad educativa. Eliminar barreras de acceso: Proveer conectividad a internet y recursos para estudiantes con necesidades especiales. Promover actividades extracurriculares: Estas pueden desarrollar habilidades complementarias al aprendizaje formal. La educación es un sistema complejo donde múltiples variables convergen para determinar el éxito estudiantil. Este análisis destaca que intervenciones dirigidas y basadas en datos son clave para reducir desigualdades y potenciar el rendimiento académico.

Principales hallazgos Participación parental: El análisis mostró diferencias significativas entre niveles de implicación parental. Los resultados indican que una “alta participación” tiene un impacto positivo claro en el desempeño académico.

Acceso a recursos: Los estudiantes con mejor acceso a materiales educativos obtienen mejores resultados, según confirma la prueba de Tukey.

Actividades extracurriculares: Los alumnos que participan regularmente en estas actividades muestran un desempeño superior en comparación con quienes no lo hacen.

Motivación personal: Se observa una correlación directa entre el nivel de motivación y el rendimiento académico. Los estudiantes altamente motivados son los más destacados.

Acceso a internet: Tener conectividad constante es crucial para el éxito académico, con resultados significativamente superiores para los estudiantes que cuentan con este recurso.

Ingreso familiar: Las familias con mayores ingresos pueden ofrecer más oportunidades a sus hijos, reflejándose en mejores resultados.

Calidad docente: Docentes con alta capacitación y experiencia impactan positivamente en el aprendizaje de sus estudiantes.

Tipo de escuela: Aunque las escuelas privadas suelen mostrar mejores resultados, no siempre las diferencias son estadísticamente significativas.

Influencia del entorno social: Los estudiantes rodeados de compañeros con actitudes positivas hacia el estudio tienden a obtener mejores resultados.

Discapacidad de aprendizaje: Los estudiantes con barreras de aprendizaje enfrentan desafíos significativos, subrayando la importancia de ofrecer apoyos adecuados.

Nivel educativo de los padres: El nivel de estudios de los padres guarda una relación positiva con el rendimiento de sus hijos, siendo los de mayor escolaridad los que más favorecen los buenos resultados.

Distancia a la escuela: La proximidad al centro educativo favorece la asistencia y mejora los resultados académicos.

Género: Las diferencias en el rendimiento por género varían según el contexto, aunque no siempre son significativas.

Este análisis refuerza la importancia de abordar factores específicos mediante estrategias basadas en evidencia para promover un aprendizaje equitativo y de calidad.