



EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO
NACIONAL DE MÉXICO



Instituto **Tecnológico**
de Aguascalientes



INGENIERÍA
EN GESTIÓN EMPRESARIAL

TECNOLÓGICO NACIONAL DE MÉXICO

INSTITUTO TECNOLÓGICO DE AGUASCALIENTES

Ingeniería en gestión empresarial

Estadística inferencial I

Proyecto-Examen

Docente

José Cruz Esparza Muñoz

Alumna

Pamela Rubí Meza Garza

Fecha

09/Diciembre/2024

December 9, 2024

#Introducción El desempeño académico de los estudiantes es influenciado por múltiples factores, tanto internos como externos, que van desde la calidad del entorno educativo hasta las características individuales. Este estudio utiliza datos recopilados sobre 6,378 estudiantes, incluyendo variables como la participación parental, la motivación, el acceso a recursos y la calidad del profesorado. El objetivo principal es identificar cuáles de estas variables tienen un impacto significativo en los puntajes de los exámenes y cómo estas diferencias pueden ser aprovechadas para mejorar la calidad educativa. Para lograrlo, se realizaron análisis estadísticos, incluyendo ANOVA y pruebas post hoc de Tukey, a fin de determinar las relaciones existentes entre las variables categóricas y los puntajes académicos.

#Marco teórico El rendimiento académico es uno de los indicadores fundamentales para evaluar la efectividad de los sistemas educativos. Según la teoría del capital social, la participación parental actúa como un mediador clave en el desarrollo de las habilidades de los estudiantes (Coleman, 1988). La disponibilidad de recursos materiales y tecnológicos, como el acceso a internet, también es crucial para permitir a los estudiantes completar tareas y acceder a información adicional.

La motivación del estudiante, según Deci y Ryan (1985), se vincula con la teoría de la autodeterminación, que sugiere que niveles altos de motivación intrínseca generan mejores resultados. Asimismo, la calidad del profesorado y la influencia de los compañeros han sido identificadas como factores determinantes en el desarrollo de habilidades cognitivas y sociales. Este estudio se fundamenta en estos conceptos teóricos para explorar cómo variables como la calidad del entorno educativo, las características familiares y el contexto personal afectan significativamente los resultados académicos.

#Metodología Datos:

Se utilizó un conjunto de datos de 6,378 estudiantes, en el que se analizaron 20 variables relacionadas con aspectos personales, familiares y educativos. Entre las variables clave se incluyen: participación parental, acceso a recursos, actividades extracurriculares, calidad del profesorado, nivel de motivación, y distancia a la escuela.

Análisis Estadístico:

ANOVA (Análisis de Varianza):

Se aplicó para evaluar si existen diferencias significativas en los puntajes de exámenes entre los grupos categóricos de cada variable.

Prueba de Tukey:

Se empleó para identificar pares específicos de grupos con diferencias significativas tras el análisis ANOVA. Herramientas Utilizadas: El análisis fue realizado con Python, utilizando las librerías pandas para el manejo de datos, scipy.stats para el ANOVA y statsmodels para las pruebas post hoc.

Nivel de Significancia:

El nivel de significancia se fijó en $\alpha = 0.05$. Una p-valor menor a este umbral indicó diferencias significativas entre los grupos analizados.

Variables Analizadas:

Variables con efecto significativo: Participación parental, acceso a recursos, calidad del profesorado, influencia de compañeros, entre otras. Variables sin efecto significativo: Actividad física y género.

```
[1]: import pandas as pd

df = pd.read_csv('https://raw.githubusercontent.com/pame45/Proyecto-U4/refs/heads/main/dataset.csv')
df.drop(columns=[], inplace=True)
df.dropna(inplace=True)
df
```

```
[1]:
```

	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

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

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

```
[3]: # 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.

Resultados

La participación parental tiene un impacto significativo en el desempeño académico de los estudiantes. Aquellos con mayor participación parental tienden a obtener mejores puntajes en los exámenes.

```
[4]: # 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)

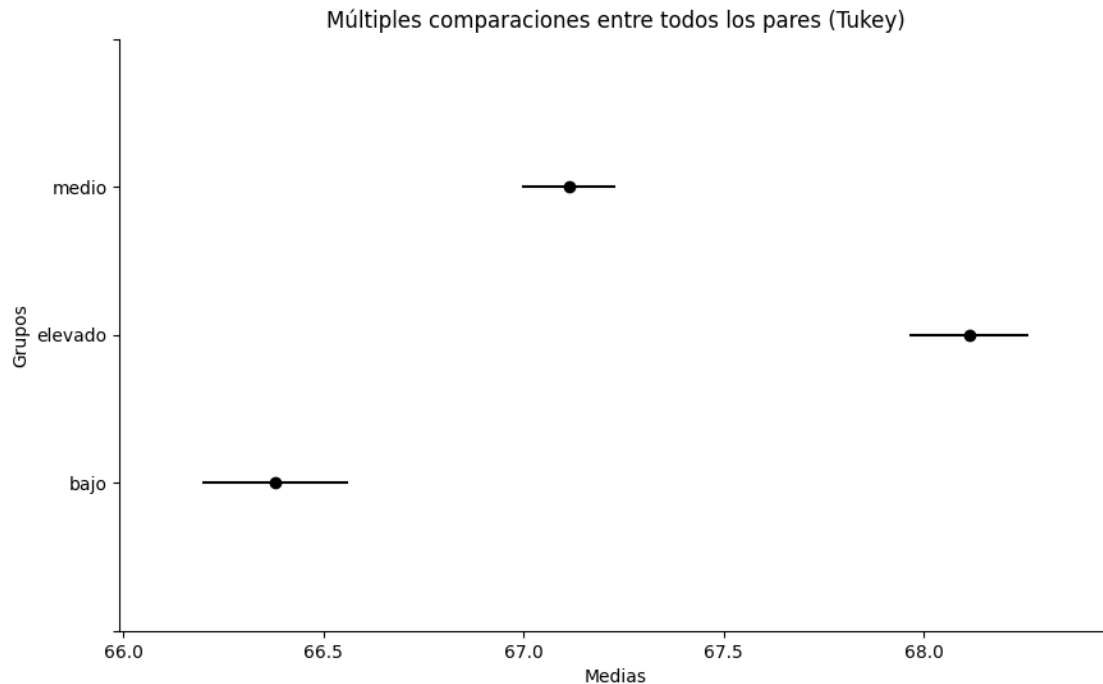
# Mostrar los resultados
print(tukey)

# 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
-----
```

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



Resultados

Los estudiantes con participación elevada obtuvieron puntajes significativamente mejores que los de participación baja o media. Además, los estudiantes con participación media superaron a los de participación baja.

#Acceso a recursos

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

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

```
[6]: # 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'] == 'elevado']['puntaje_examen'],
    df[df['acceso_a_recursos'] == 'medio']['puntaje_examen'],
    df[df['acceso_a_recursos'] == 'bajo']['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.

Resultados

Los recursos disponibles para los estudiantes afectan significativamente su rendimiento. Aquellos con acceso elevado a recursos obtienen mejores puntajes que los que tienen acceso medio o bajo.

```

[7]: # 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

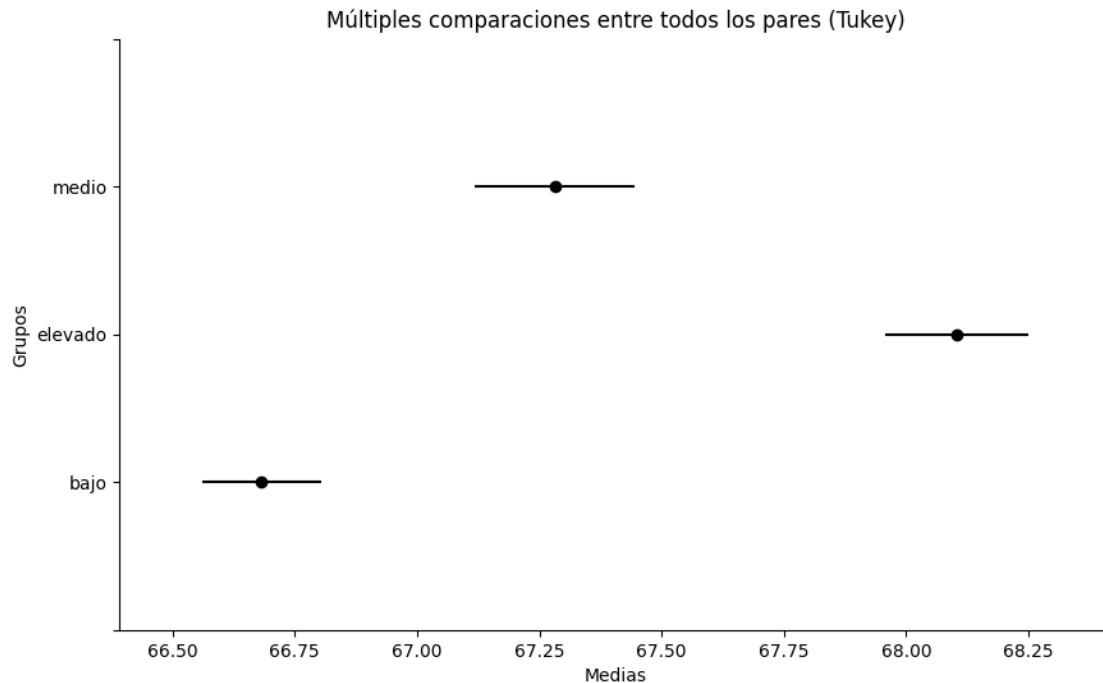
```

=====
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

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



Resultados

Diferencias significativas entre todos los niveles: bajo, medio y elevado. Los estudiantes con acceso elevado lograron los mejores puntajes.

#Actividades extracurriculares

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

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

```
[9]: # 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'] == 'no']['puntaje_examen'],
df[df['actividades_extracurriculares'] == 'si']['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.

Resultados

Participar en actividades extracurriculares tiene un efecto positivo en los puntajes. Los estudiantes que participan tienen un mejor desempeño académico en comparación con aquellos que no lo hacen.

```

[10]: # 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

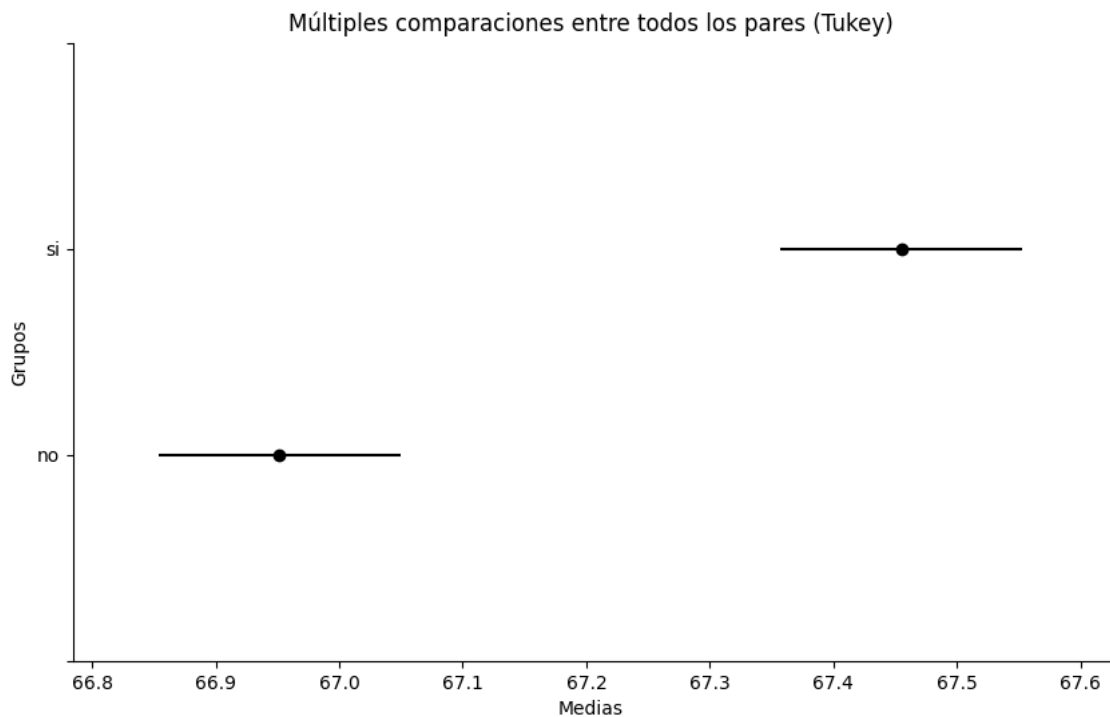
```

```
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
-----
```

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



Resultados

La diferencia entre los grupos (sí y no) es significativa, indicando que la participación contribuye positivamente al rendimiento.

#Nivel motivación

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

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

```
[12]: # 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.

Resultados

La motivación de los estudiantes influye directamente en sus puntajes. Aquellos con niveles más altos de motivación tienden a obtener mejores resultados.

```

[13]: # 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)")

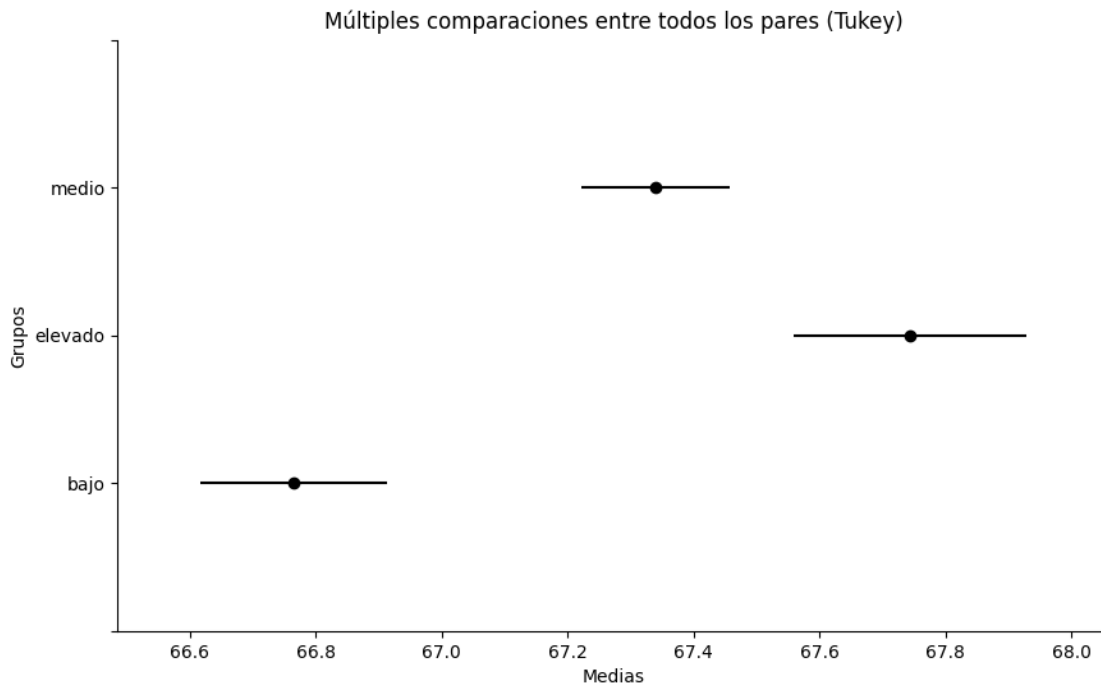
```

```

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
-----

```

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



Resultados

Diferencias significativas entre todos los niveles: bajo, medio y elevado. Los estudiantes con motivación elevada destacaron en rendimiento académico.

#Acceso internet

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

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

```
[15]: # 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.

Resultados

El acceso a internet afecta significativamente el desempeño de los estudiantes. Aquellos con acceso a internet obtienen mejores puntajes en promedio.

```
[16]: # 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)")

```

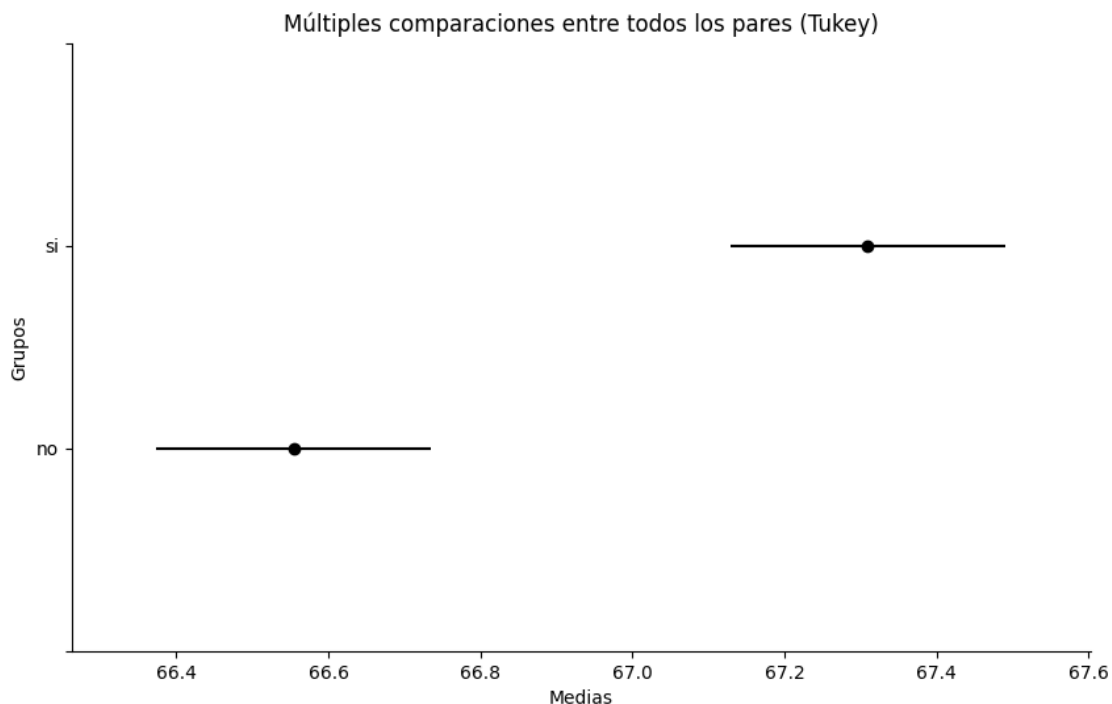
Multiple Comparison of Means - Tukey HSD, FWER=0.05

```

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

```

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



Resultados

Diferencia significativa entre estudiantes con acceso (sí) y sin acceso (no).

#Calidad profesorado

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

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

```
[18]: # 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'] == 'medio']['puntaje_examen'],
    df[df['calidad_profesorado'] == 'bajo']['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.

Resultados

La calidad del profesorado tiene un impacto considerable en el desempeño estudiantil. Profesores de calidad elevada se asocian con mejores puntajes.

```
[19]: # 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)")

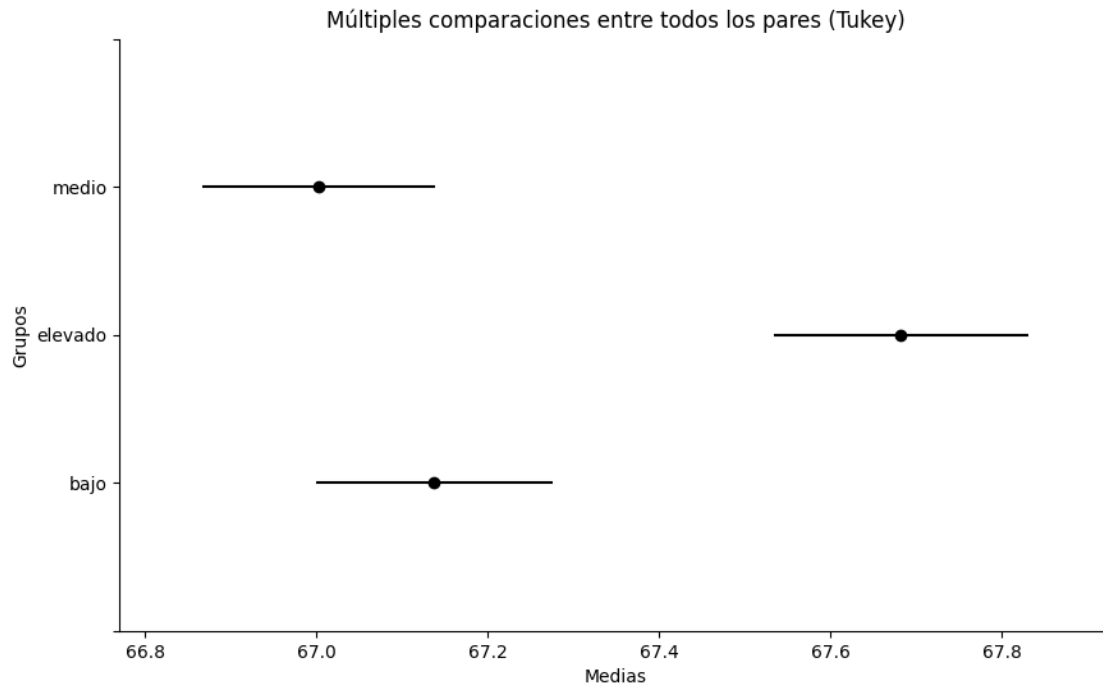
```

```

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
=====

```

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



Resultados

Diferencias significativas entre bajo y elevado, y entre medio y elevado. Profesores con calidad elevada muestran los mejores resultados.

#Tipo escuela

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

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

```
[21]: # 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.

Resultados

No se observaron diferencias significativas en los puntajes académicos entre estudiantes de escuelas públicas y privadas.

```

[22]: # 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['tipo_escuela'], 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)")

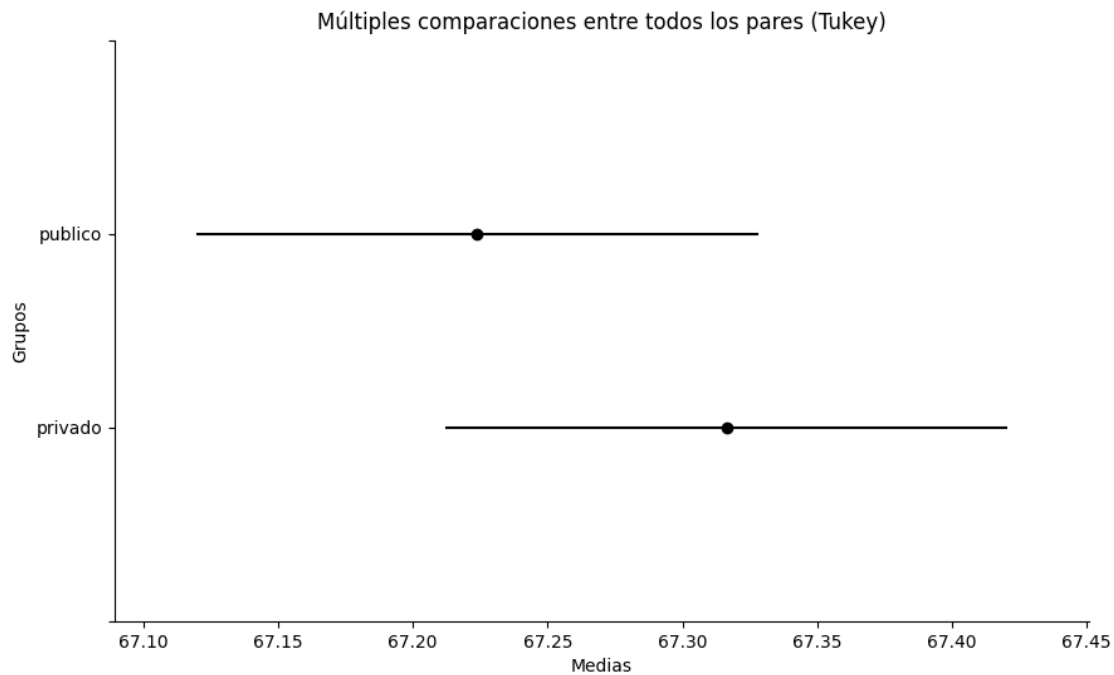
```

```

Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====
group1 group2 meandiff p-adj lower upper reject
-----
privado publico -0.0924 0.3855 -0.3011 0.1163 False

```

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



Resultados

No hay diferencias significativas en el desempeño académico entre estudiantes de escuelas públicas y privadas según la prueba de Tukey.

#Influencia compañeros

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

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

```
[24]: # 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'] == 'negativo']['puntaje_examen'],
    df[df['influencia_companieros'] == 'neutral']['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.

Resultados

La percepción de la influencia de los compañeros afecta significativamente los puntajes. Influencias positivas conducen a un mejor desempeño que las neutrales o negativas.

```

[25]: # 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")

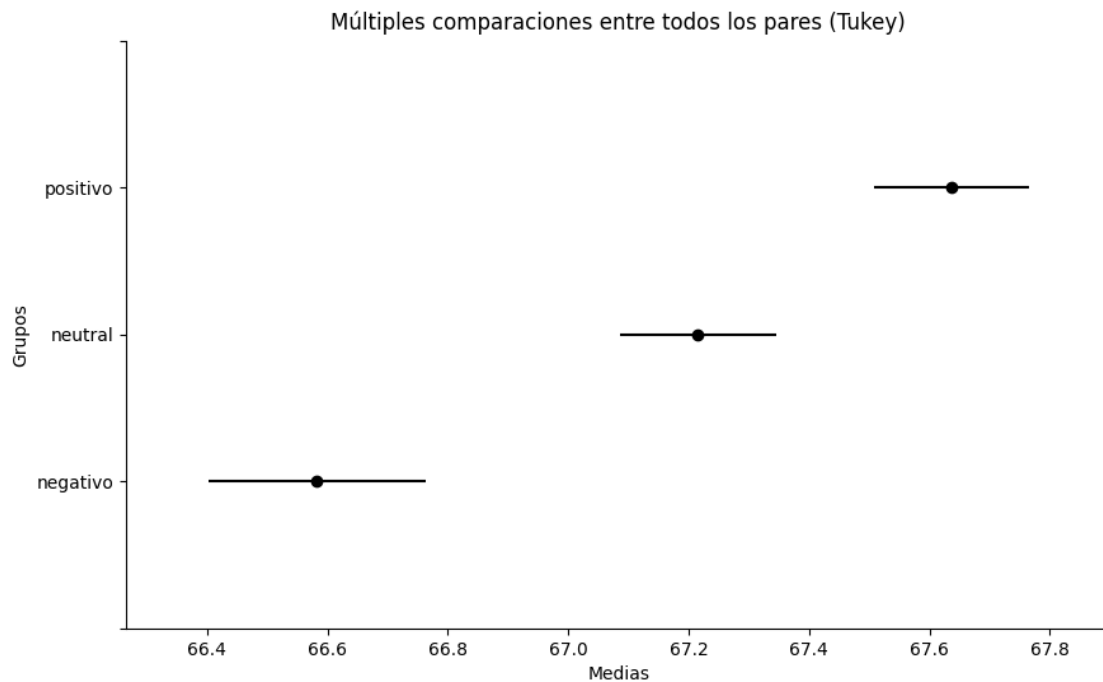
plt.gca().spines['right'].set_visible(False) # derecha
plt.gca().spines['top'].set_visible(False) # superior
plt.title("Múltiples comparaciones entre todos los pares (Tukey)")

```

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
-----
```

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



Resultados

Diferencias significativas entre positivo, neutral y negativo. La influencia positiva genera los mejores resultados.

#Actividad física

[26]: `df['actividad_fisica'].unique()`

[26]: `array([3, 4, 2, 1, 5, 0, 6])`

[27]: `# 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['actividad_fisica'] == 3]['puntaje_examen'],
    df[df['actividad_fisica'] == 4]['puntaje_examen'],
    df[df['actividad_fisica'] == 2]['puntaje_examen'],
    df[df['actividad_fisica'] == 1]['puntaje_examen'],
    df[df['actividad_fisica'] == 5]['puntaje_examen'],
    df[df['actividad_fisica'] == 0]['puntaje_examen'],
    df[df['actividad_fisica'] == 6]['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: 1.22

Valor p: 0.29195

Conclusión: No hay diferencias significativas entre los tratamientos.

Resultados

El nivel de actividad física no muestra un impacto significativo en los puntajes académicos. No se observaron diferencias claras entre los grupos analizados.

[28]:

```

# 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['actividad_fisica'], 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)")

```

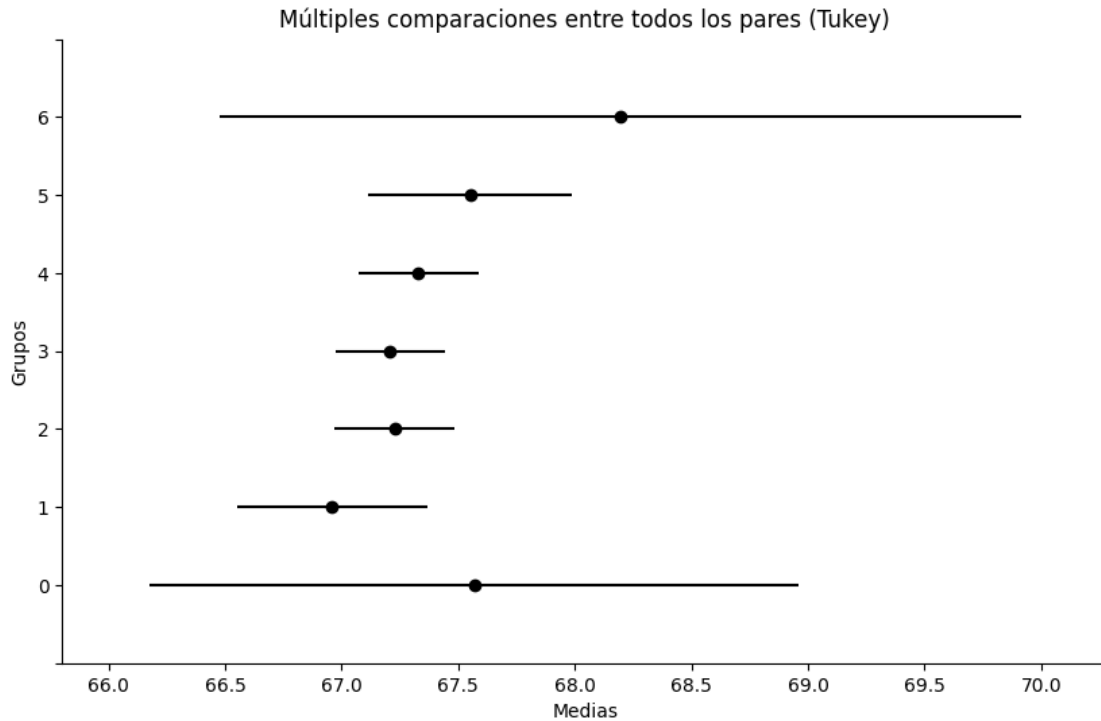
Multiple Comparison of Means - Tukey HSD, FWER=0.05

```

=====
group1 group2 meandiff p-adj  lower  upper  reject
-----
0      1  -0.6082 0.9587 -2.4415  1.2252  False
0      2  -0.3415 0.9976 -2.106   1.4229  False
0      3  -0.3605 0.9967 -2.1161  1.3952  False
0      4  -0.2385 0.9997 -2.0035  1.5264  False
0      5  -0.0162 1.0    -1.8636  1.8313  False
0      6   0.6254 0.9936 -2.0813  3.332   False
1      2   0.2666 0.8883 -0.3802  0.9135  False
1      3   0.2477 0.904  -0.3746  0.87    False
1      4   0.3696 0.6279 -0.2784  1.0176  False
1      5   0.592  0.3767 -0.2554  1.4395  False
1      6   1.2335 0.6225 -0.9184  3.3855  False
2      3  -0.0189 1.0    -0.3924  0.3545  False
2      4   0.103  0.9906 -0.3118  0.5179  False
2      5   0.3254 0.8025 -0.3605  1.0112  False
2      6   0.9669 0.822  -1.1267  3.0605  False
3      4   0.1219 0.9628 -0.2535  0.4974  False
3      5   0.3443 0.7254 -0.3185  1.0071  False
3      6   0.9858 0.8055 -1.1004  3.072   False
4      5   0.2224 0.9634 -0.4646  0.9093  False
4      6   0.8639 0.8879 -1.2301  2.9579  False
5      6   0.6415 0.9764 -1.5225  2.8056  False
-----

```

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



Resultados

No se encontraron diferencias significativas en el desempeño académico entre los grupos de diferentes niveles de actividad física. La actividad física, en este caso, no parece estar asociada con los puntajes en los exámenes según la prueba de Tukey.

1 Discapacidad aprendizaje

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

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

```
[30]: # 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'] == 'no']['puntaje_examen'],
    df[df['discapacidad_aprendizaje'] == 'si']['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.

Resultados

Los estudiantes con discapacidades de aprendizaje tienen un desempeño significativamente diferente (menor en promedio) en comparación con aquellos que no presentan discapacidades.

```

[31]: # 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")

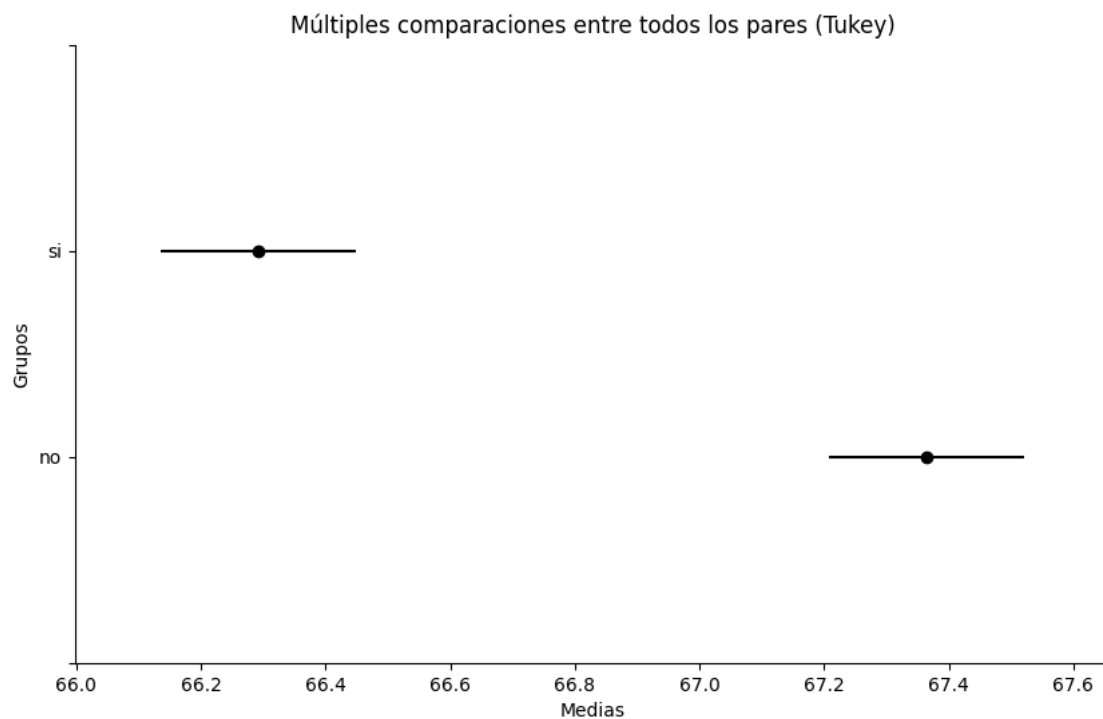
plt.gca().spines['right'].set_visible(False) # derecha
plt.gca().spines['top'].set_visible(False) # superior
plt.title("Múltiples comparaciones entre todos los pares (Tukey)")

```

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
-----
```

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



Resultados

Diferencia significativa entre estudiantes con (sí) y sin (no) discapacidades.

#Nivel estudio padres

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

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

```
[33]: # 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.

Resultados

El nivel educativo de los padres influye significativamente en el desempeño académico de los estudiantes.

```

[34]: # 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)

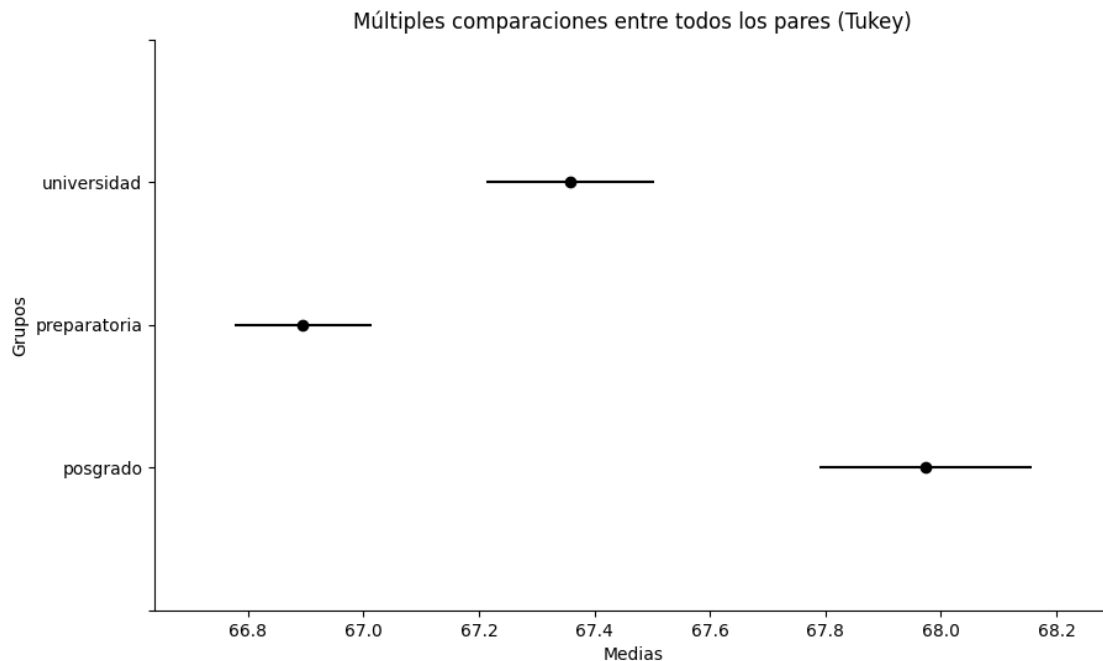
# Gráfico de las diferencias entre grupos
tukey.plot_simultaneous(ylabel="Grupos", xlabel="Medias")

```

```
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
=====
```

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



Resultados

Diferencias significativas entre los niveles preparatoria, universidad y posgrado. Los estudiantes con padres que tienen estudios de posgrado obtuvieron mejores puntajes.

#Distancia

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

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

```
[36]: # 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.

Resultados

La distancia a la escuela afecta los puntajes académicos. Estudiantes que viven más cerca de la escuela obtienen mejores resultados.

```
[37]: # 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)
```

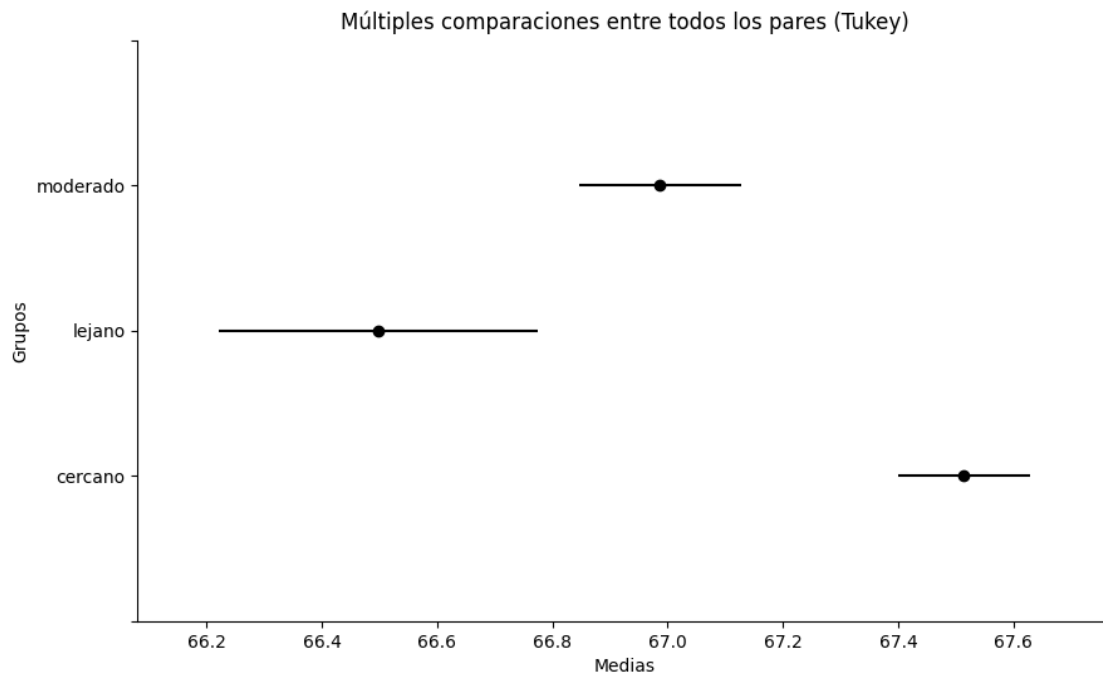
```
# Mostrar los resultados
print(tukey)

# 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
-----
```

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



Resultados

Diferencias significativas entre los niveles cercano, moderado y lejano. Estar más cerca de la escuela está relacionado con mejores puntajes.

2 Genero

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

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

```
[39]: # 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.

Resultados

No se encontraron diferencias significativas en los puntajes académicos según el género. Tanto hombres como mujeres obtuvieron resultados similares.

```
[40]: # 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['genero'],
    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)")

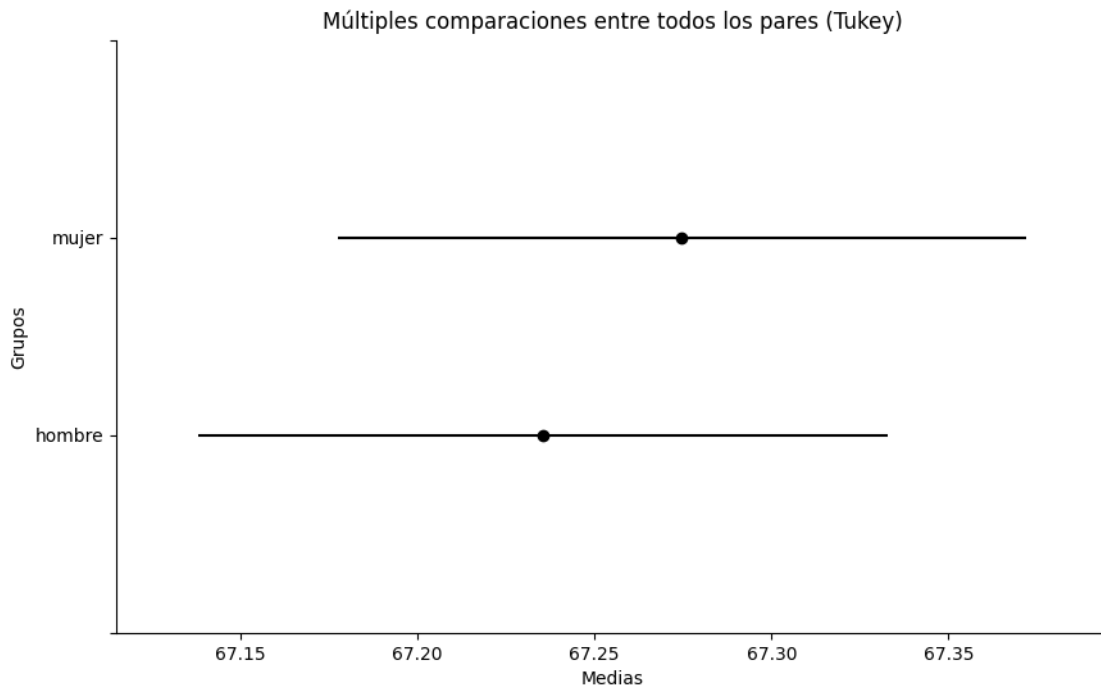
```

```

Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====
group1 group2 meandiff p-adj  lower  upper  reject
-----
hombre  mujer   0.0391 0.6937 -0.1555 0.2337  False
-----

```

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



Resultados

No hay diferencias significativas en el desempeño académico entre hombres y mujeres según la prueba de Tukey.

#Conclusión El análisis realizado reveló que varios factores tienen un impacto significativo en los puntajes académicos de los estudiantes:

Participación parental: Los estudiantes con padres más involucrados presentan puntajes considerablemente más altos. Acceso a recursos: Tener acceso a recursos materiales y tecnológicos se asocia con un mejor desempeño académico.

Calidad del profesorado: Los profesores con mayor calidad generan un entorno más propicio para el aprendizaje.

Motivación e influencia de compañeros: La motivación intrínseca y las influencias positivas contribuyen directamente al éxito académico. Por otro lado, variables como la actividad física y el género no mostraron un impacto significativo, lo que sugiere que estos factores podrían depender de contextos más específicos o de interacciones con otras variables no incluidas en este análisis.

En general, este estudio destaca la importancia de invertir en recursos educativos, fortalecer la calidad del profesorado y fomentar una participación activa de los padres en el proceso educativo. Estos hallazgos proporcionan información valiosa para el diseño de estrategias educativas y políticas públicas que promuevan un aprendizaje más equitativo y efectivo.