



EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO
NACIONAL DE MÉXICO



Instituto **Tecnológico**
de Aguascalientes



Instituto **Tecnológico**
de Aguascalientes



INGENIERÍA
EN GESTIÓN EMPRESARIAL

TECNOLÓGICO NACIONAL DE MÉXICO

INSTITUTO TECNOLÓGICO DE AGUASCALIENTES

Ingeniería en Gestión Empresarial

IG1 2024B

Estadística Inferencial II

Proyecto- Examen

Docente: Mtro. José Cruz Muñoz Esparza

Integrantes

Diana Itzel Rodríguez Reyes

Fecha: 06/Diciembre/2024

December 7, 2024

1 Introducción

En el presente trabajo se exploran diversos factores que influyen en el desempeño académico de los estudiantes. A través del análisis estadístico, se pretende determinar cómo variables como la participación parental, acceso a recursos, actividades extracurriculares, nivel de motivación, acceso a internet, calidad del profesorado, tipo de escuela, influencia de los compañeros, discapacidad de aprendizaje, nivel de estudio de los padres, distancia al centro educativo, y género, afectan el puntaje en los exámenes.

El objetivo principal es identificar qué aspectos tienen un impacto importante en los resultados académicos para así poder proponer estrategias de mejora en el ámbito educativo.

Este proyecto combina métodos estadísticos como ANOVA y pruebas de comparaciones múltiples para brindar así poder hacer una análisis optimo.

2 Marco Teórico

En este trabajo se emplearon diversos métodos estadísticos y herramientas para analizar los datos y obtener conclusiones bien elaboradas.

1. Análisis de Varianza (ANOVA) El ANOVA es una técnica utilizada para comparar las medias de tres o más grupos y determinar si existen diferencias significativas entre ellos. En este contexto, se utilizó para analizar cómo las variables categóricas, como la participación parental o el tipo de escuela, influyen en el puntaje de los exámenes.

2. Pruebas Post-hoc: Tukey Cuando el ANOVA detecta diferencias significativas, se aplican pruebas como la de Tukey para identificar específicamente qué grupos difieren entre sí.

3. Distribución Normal Antes de aplicar ANOVA, es muy importante verificar si los datos siguen una distribución normal. Este requisito nos asegura la validez de las pruebas estadísticas. En este proyecto, se asumió que las variables cumplían esta condición.

4. Valor p El valor p es una medida que ayuda a determinar la significancia estadística de los resultados. En este análisis, se utilizó para decidir si las diferencias observadas entre grupos eran lo suficientemente grandes como para no atribuirse al azar.

5. Scipy: ANOVA y Pruebas de Hipótesis La biblioteca **Scipy** fue empleada para realizar el ANOVA y evaluar la significancia de los resultados. Scipy es conocida por su eficiencia en cálculos matemáticos y su amplia implementación de pruebas estadísticas.

3 Metodología

Para este análisis, se siguió un enfoque estructurado que permitió evaluar el impacto de varios factores en el desempeño académico de los estudiantes.

1. Obtención del conjunto de datos

Se trabajó con un conjunto de datos que contiene información de 6,378 estudiantes, cada uno con datos sobre 20 variables. Estas variables incluyeron aspectos relacionados con características individuales de cada estudiante (género, horas de estudio, horas de sueño, etc.), factores ambientales que rodean al estudiante (participación parental, nivel de recursos, tipo de escuela, etc.) y resultados académicos medidos en puntajes de exámenes.

2. Procesamiento de los datos

Antes del análisis, se realizaron los siguientes pasos:

- Se eliminaron columnas para el análisis actual.
- Se eliminaron filas con valores faltantes.
- Las “variables categóricas” se mantuvieron en su forma original para hacer mejor la comparación entre grupos.

3. Cálculo de la Métrica Objetivo

El puntaje de examen se calculó como una combinación de varias variables relacionadas con el rendimiento académico, tales como horas de estudio, asistencia, calidad del profesorado, calificaciones previas, entre otras.

4. Prueba de Hipótesis para cada variable

Para evaluar la influencia de cada factor, se utilizaron pruebas de ANOVA, con el fin de determinar si existían diferencias significativas entre los grupos de cada variable categórica. Los pasos específicos fueron:

- Definir un nivel de significancia estadística.
- Aplicar la prueba ANOVA para comparar las medias de los puntajes de examen entre los grupos de cada variable (por ejemplo, alto, medio y bajo para el acceso a recursos).
- Interpretar los resultados basados en el valor p.

5. Pruebas “Post-Hoc”

En los casos donde la prueba ANOVA indicó significancia, se realizaron pruebas “post-hoc” con el método de Tukey para identificar qué grupos presentaban diferencias específicas.

6. Visualización de resultados

Se generaron gráficos de comparación de medias y diagramas que mostraban las diferencias entre los grupos.

7. Análisis complementario

En variables donde las pruebas ANOVA no mostraron diferencias significativas, se realizaron

análisis adicionales (como la prueba de Kruskal-Wallis) para confirmar los resultados, especialmente en variables con posibles violaciones de la normalidad.

8. Interpretación de los Resultados

```
[3]: import pandas as pd
df= pd.read_csv('https://raw.githubusercontent.com/ItzelReyes2401/
↳Proyecto_Unidad4/refs/heads/main/dataset.csv')
df.drop(columns=[], inplace=True)
df.dropna(inplace=True)
df
```

```
[3]:      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]

```
[4]: df['puntaje_examen']=(df['horas_estudio'] + df['asistencia'] +
    ↪df['horas_suenio'] + df['calificaciones_previas'] + df['sesiones_tutoria'] +
    ↪df['actividad_fisica'] + df['puntaje_examen'])/ 7
```

df

```
[4]:      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          ...          ...                      ...          ...
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          ...          ...          ...
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          ...          ...          ...
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    36.714286
1      moderado  mujer    31.000000
2      cercano  hombre    42.857143
3      moderado  hombre    42.857143
4      cercano  mujer    37.000000
...
6602      cercano  mujer    35.428571
6603      cercano  mujer    37.428571
6604      cercano  mujer    36.285714
6605      lejano  mujer    38.000000
6606      cercano  hombre    36.142857

[6378 rows x 20 columns]

```

4 Participación parental

```

[5]: df['participacion_parental'].unique()

[5]: array(['bajo', 'medio', 'elevado'], 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['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: 0.80

Valor p: 0.45144

Conclusión: No hay diferencias significativas entre los tratamientos.

```

[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['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

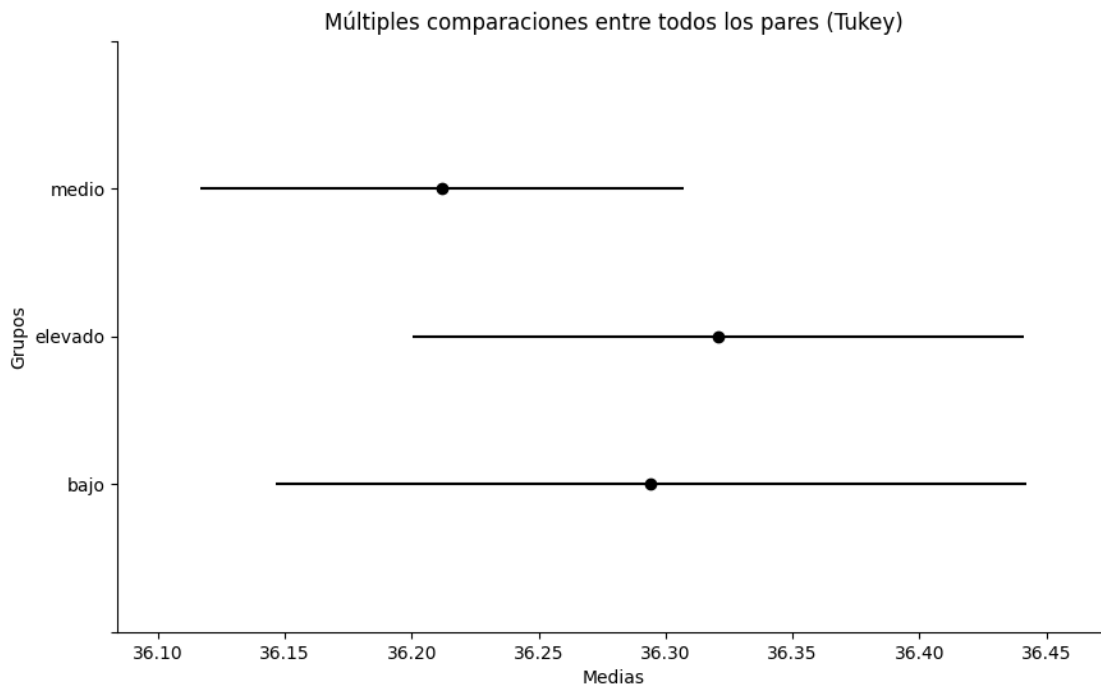
```



```
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 0.0264 0.971 -0.2417 0.2946 False
bajo medio -0.0823 0.7065 -0.3252 0.1606 False
elevado medio -0.1087 0.4637 -0.3242 0.1068 False
-----
```

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



```
[ ]: # CONCLUSIÓN: El análisis no encontró diferencias significativas entre los
      ↳ niveles de participación parental en cuanto al puntaje de exámenes.
      #Aunque se esperaba que un mayor involucramiento parental resultara en mejores
      ↳ resultados, los datos no respaldaron esta hipótesis.
      #Esto podría deberse a que otros factores tienen mayor peso o a diferencias en
      ↳ cómo se define y mide la participación.
```

5 Acceso a recursos

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

```
[8]: array(['elevado', 'medio', 'bajo'], 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['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: 3.51

Valor p: 0.03006

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```
[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['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)")

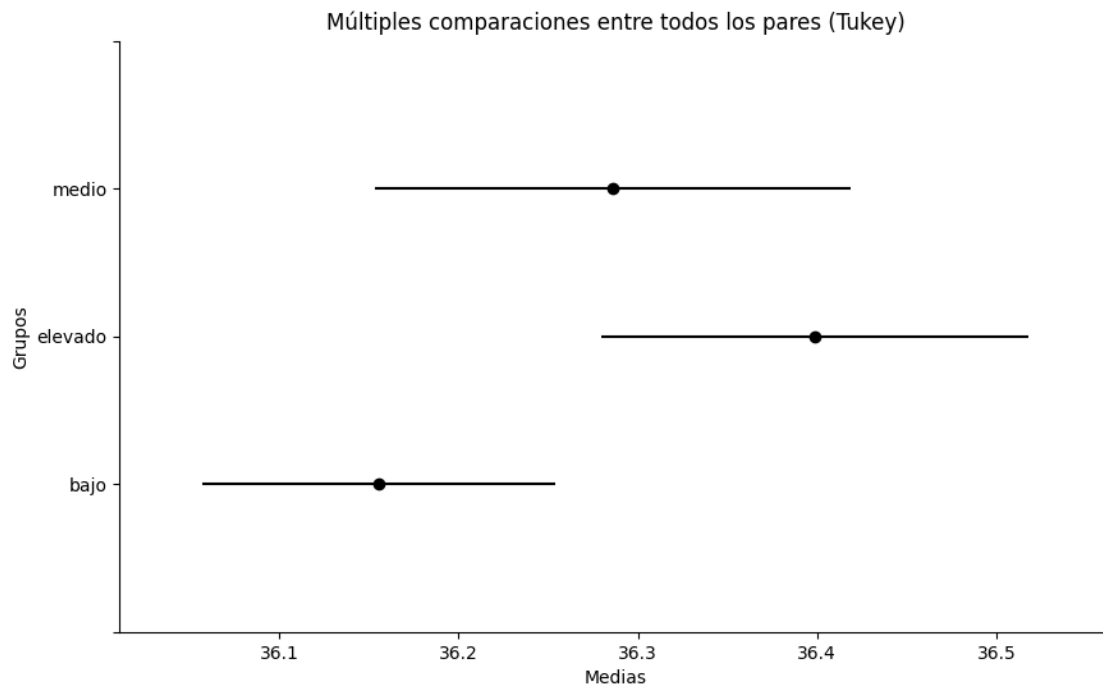
```

```

Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====
group1 group2 meandiff p-adj  lower  upper  reject
-----
    bajo elevado  0.2432 0.0239  0.0257 0.4608   True
    bajo  medio  0.1306 0.3825 -0.1009 0.362   False
elevado  medio -0.1126 0.547  -0.3648 0.1395   False
=====

```

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



```
[ ]: #CONCLUSIÓN: Se encontró que el acceso elevado a recursos educativos está  
    ↪relacionado con mejores puntajes de examen en comparación con niveles bajos.  
#Esto refuerza la importancia de garantizar que los estudiantes tengan  
    ↪materiales y apoyo suficiente para maximizar su aprendizaje.
```

6 Actividades extracurriculares

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

```
[11]: array(['no', 'si'], 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['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: 0.67

Valor p: 0.41417

Conclusión: No hay diferencias significativas entre los tratamientos.

```
[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['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)")

```

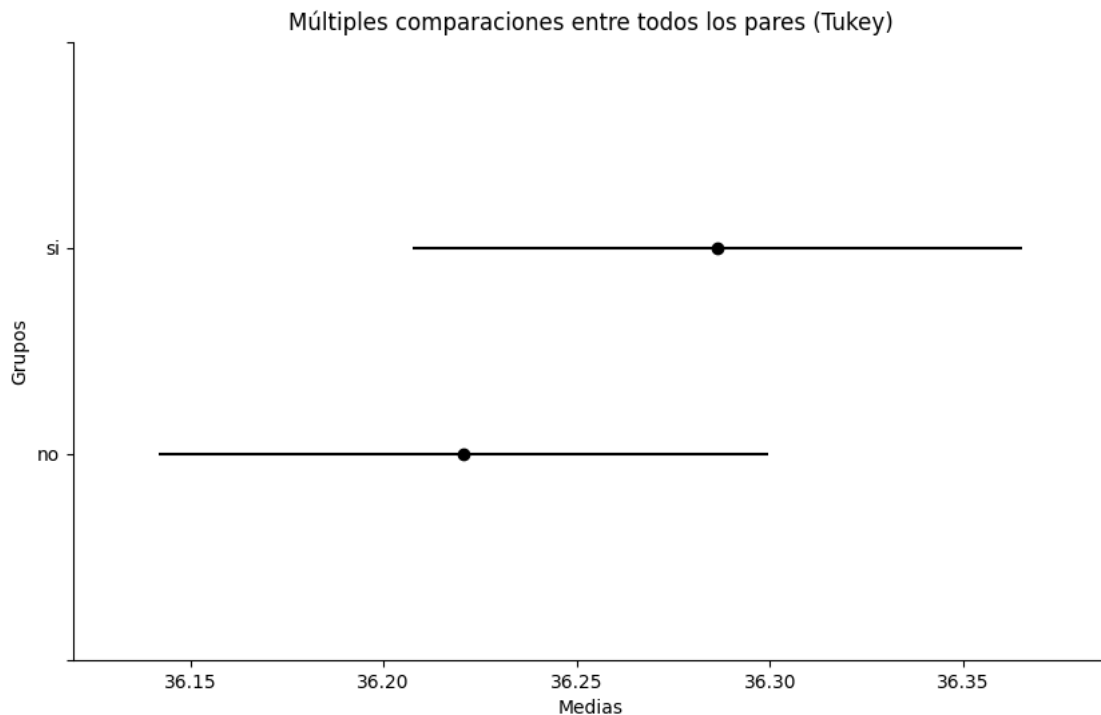
Multiple Comparison of Means - Tukey HSD, FWER=0.05

```

=====
group1 group2 meandiff p-adj  lower  upper  reject
-----
    no     si   0.0656 0.4142 -0.0919 0.2232   False
-----

```

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



```
[ ]: #CONCLUSIÓN: No hubo diferencias significativas entre estudiantes que  
    ↪ participan en actividades extracurriculares y aquellos que no.  
    #Aunque estas actividades suelen estar asociadas con habilidades blandas, su  
    ↪ impacto directo en el rendimiento académico no fue concluyente.
```

7 Nivel motivación

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

```
[14]: array(['bajo', 'medio', 'elevado'], 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['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: 0.79

Valor p: 0.45457

Conclusión: No hay diferencias significativas entre los tratamientos.

```
[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['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)")

```

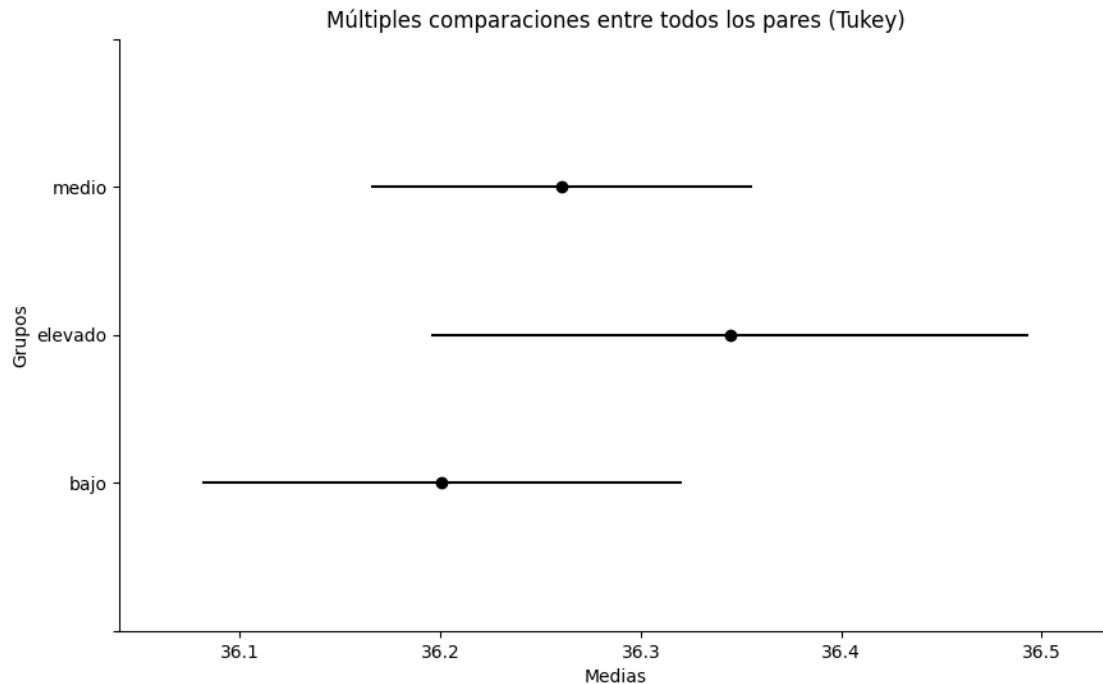
Multiple Comparison of Means - Tukey HSD, FWER=0.05

```

=====
group1 group2 meandiff p-adj  lower  upper  reject
-----
    bajo elevado  0.1436 0.4205 -0.1245 0.4118  False
    bajo  medio  0.0598 0.7909 -0.1549 0.2744  False
elevado  medio -0.0839 0.6993 -0.3279 0.1601  False
=====

```

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



```
[ ]: #CONCLUSIÓN: El nivel de motivación no mostró diferencias significativas en el
      ↳desempeño académico.
      #Esto sugiere que aunque la motivación es importante, otros factores
      ↳contextuales podrían estar opacando su impacto en los resultados.
```

8 Acceso internet

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

```
[17]: array(['si', 'no'], 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['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: 0.04

Valor p: 0.84990

Conclusión: No hay diferencias significativas entre los tratamientos.

```

[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['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)")

```

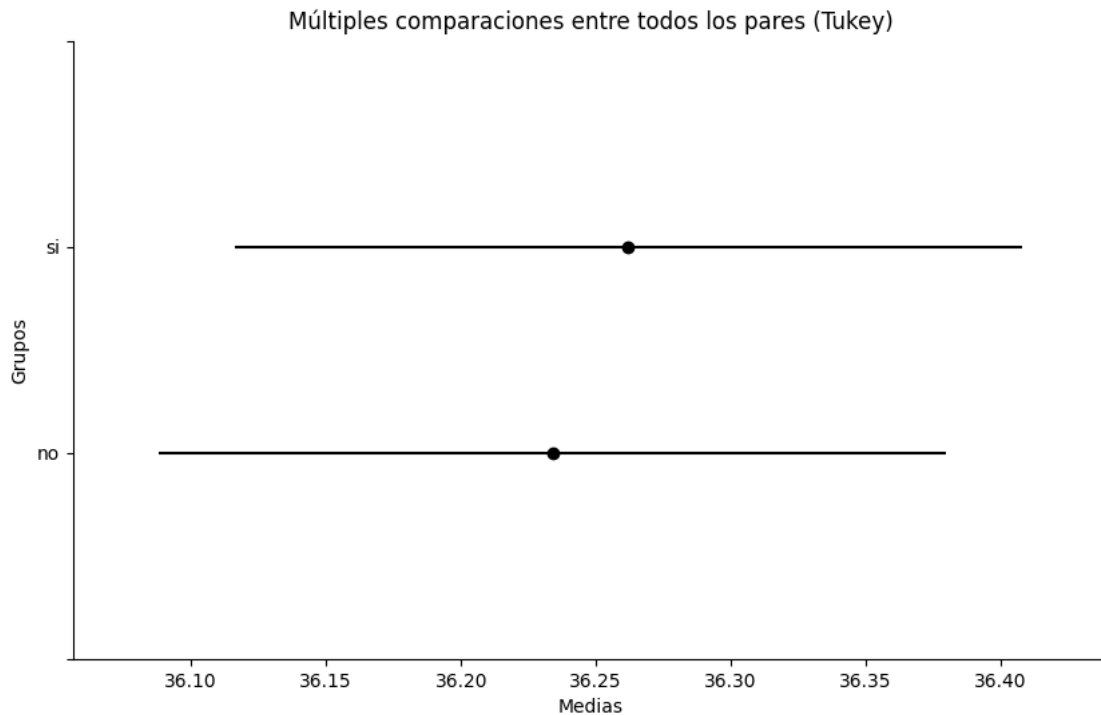
Multiple Comparison of Means - Tukey HSD, FWER=0.05

```

=====
group1 group2 meandiff p-adj lower upper reject
-----
no si 0.0282 0.8499 -0.2635 0.3198 False
-----

```

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



```
[ ]: #CONCLUSIÓN: La disponibilidad de internet no fue un factor significativo en
      ↪ los puntajes.
      #Esto podría deberse a que su uso no siempre está orientado al aprendizaje o a
      ↪ diferencias en cómo los estudiantes utilizan este recurso.
```

9 Calidad profesorado

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

```
[20]: array(['medio', 'bajo', 'elevado'], 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['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: 3.08

Valor p: 0.04587

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```

[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['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)")

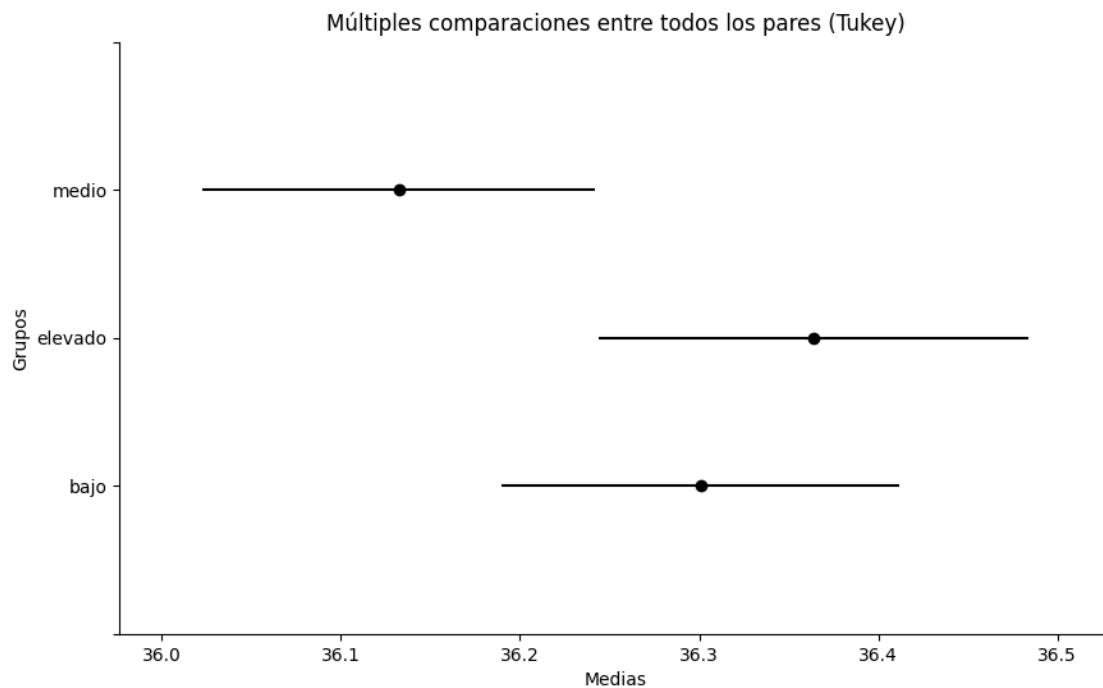
```

Multiple Comparison of Means - Tukey HSD, FWER=0.05

=====

group1	group2	meandiff	p-adj	lower	upper	reject
bajo	elevado	0.0629	0.7983	-0.1678	0.2937	False
bajo	medio	-0.1685	0.173	-0.3892	0.0522	False
elevado	medio	-0.2315	0.0474	-0.4609	-0.002	True

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



```
[ ]: #CONCLUSIÓN: Se observaron diferencias significativas entre niveles altos y  

     ↪medios de calidad del profesorado.  

     #Un profesorado más capacitado y efectivo tiene un impacto positivo en el  

     ↪desempeño, lo que resalta la necesidad de inversión en la formación docente.
```

10 Tipo escuela

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

```
[23]: array(['publico', 'privado'], 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['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: 0.13

Valor p: 0.71455

Conclusión: No hay diferencias significativas entre los tratamientos.

```

[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['tipo_escuela'], alpha=nivel_de_significancia)

# Mostrar los resultados
print(tukey)

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

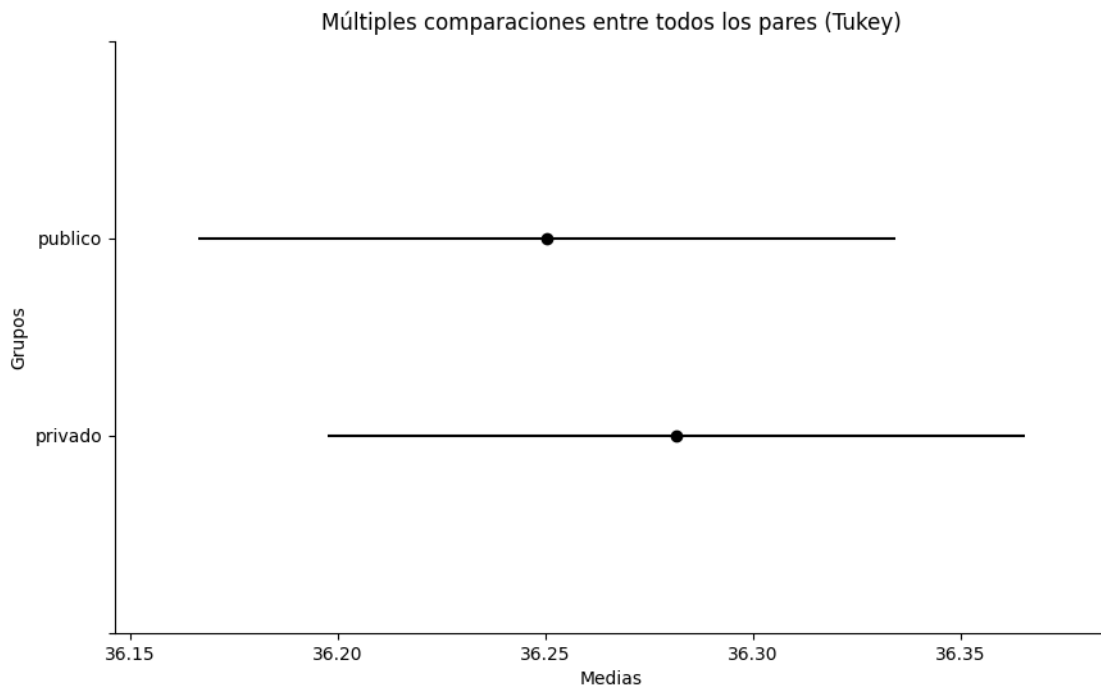
plt.gca().spines['right'].set_visible(False) # derecha
plt.gca().spines['top'].set_visible(False)   # superior

```

```
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 -0.0313 0.7146 -0.1993 0.1366 False
-----
```

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



```
[ ]: #CONCLUSIÓN: No hubo diferencias significativas entre escuelas públicas y
      ↪ privadas.
      #EstO desafía la percepción común de que las escuelas privadas ofrecen una
      ↪ mejor calidad educativa.
```

11 Influencia compañeros

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

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

```
[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['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: 2.09

Valor p: 0.12394

Conclusión: No hay diferencias significativas entre los tratamientos.

```
[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['influencia_companieros'], alpha=nivel_de_significancia)

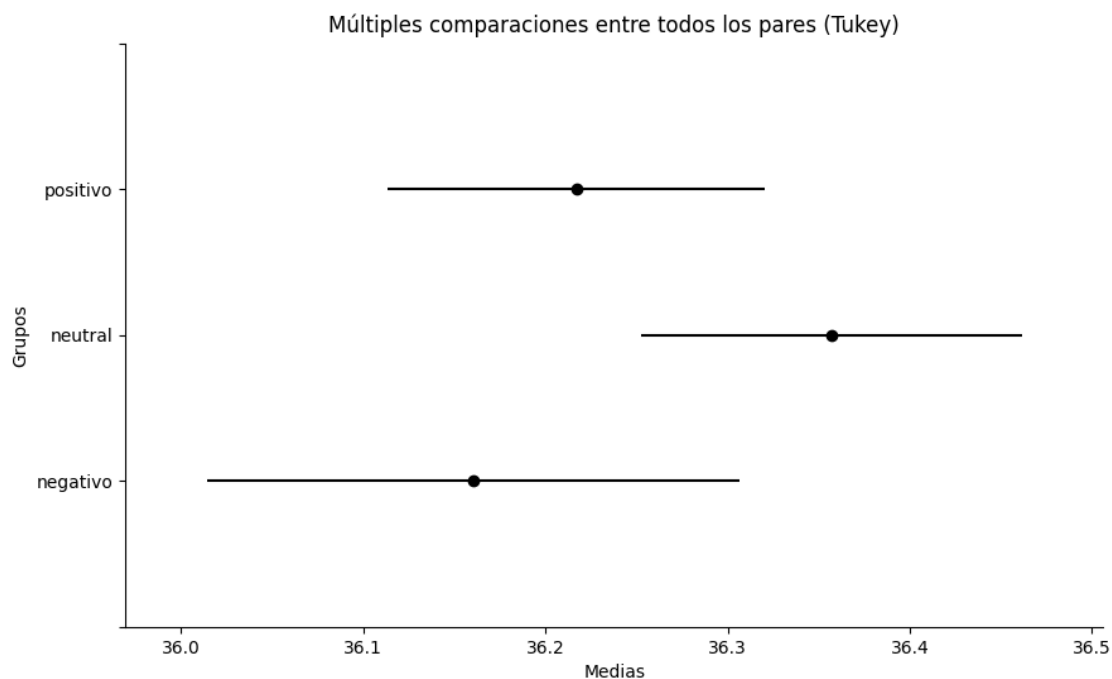
# Mostrar los resultados
print(tukey)
```

```
# 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    0.1967 0.1568 -0.0539 0.4473  False
negativo positivo  0.0567 0.8554 -0.1929 0.3063  False
neutral positivo  -0.14 0.2546 -0.3478 0.0678  False
-----
```

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



```
[ ]: # CONCLUSIÓN: No se encontraron diferencias significativas entre los tipos de
      ↪ influencia de los compañeros.
      # Aunque el ambiente social es importante, parece no tener un impacto directo en
      ↪ este caso.
```


12 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: 3.74

Valor p: 0.05319

Conclusión: No hay diferencias significativas entre los tratamientos.

```
[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)")

```

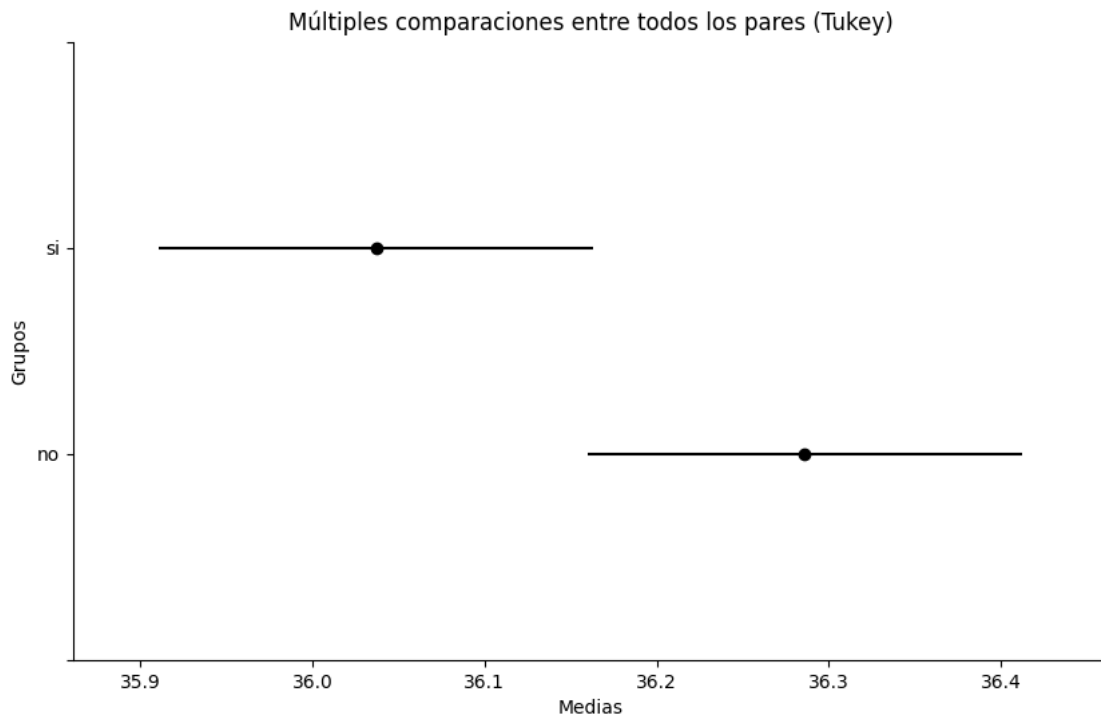
Multiple Comparison of Means - Tukey HSD, FWER=0.05

```

=====
group1 group2 meandiff p-adj  lower  upper  reject
-----
   no     si   -0.249 0.0532 -0.5014 0.0034  False
-----

```

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



[]:

```
#CONCLUSIÓN: No se encontraron diferencias significativas en el desempeño entre
↪ estudiantes con y sin discapacidades de aprendizaje, aunque los datos
↪ sugieren una ligera tendencia hacia menores puntajes en el primer grupo.
```

13 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: 2.37

Valor p: 0.09390

Conclusión: No hay diferencias significativas entre los tratamientos.

```
[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)")

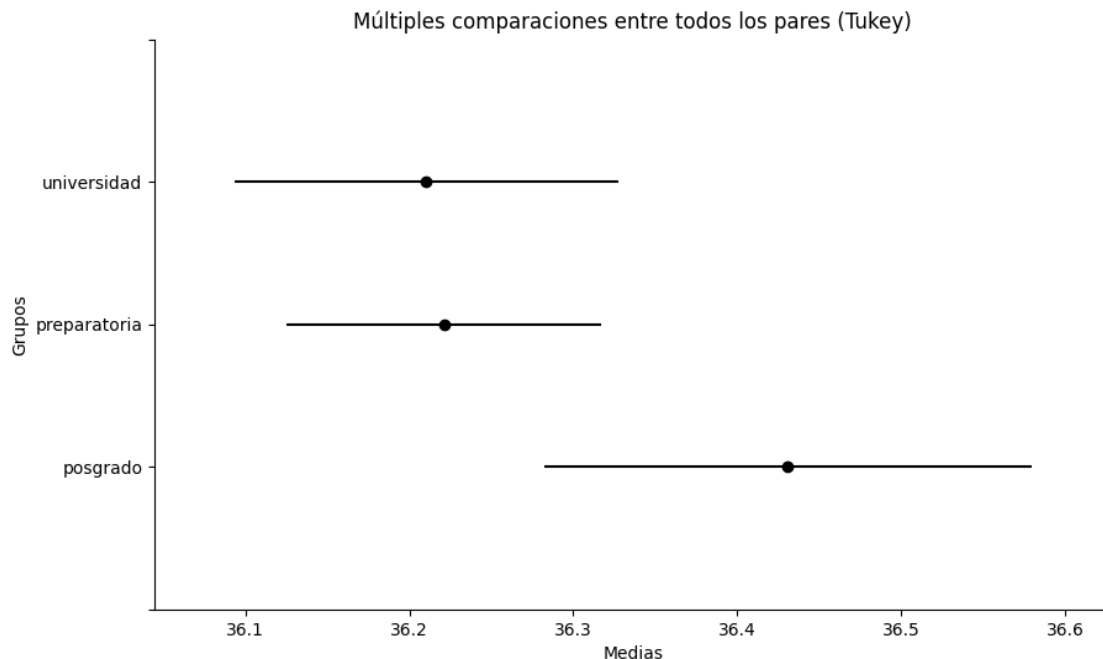
```

```

Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====
group1      group2      meandiff p-adj   lower  upper  reject
-----
posgrado preparatoria -0.2098 0.1097 -0.4543 0.0348  False
posgrado universidad -0.2205 0.1265 -0.4863 0.0454  False
preparatoria universidad -0.0107 0.9924 -0.2236 0.2022  False
=====

```

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



```
[ ]: #CONCLUSIÓN:No hubo diferencias significativas en los puntajes según el nivel
      ↪ de educación de los padres.
      #Esto indica que el éxito académico puede depender más de las interacciones y
      ↪ el apoyo directo que de la formación académica de los padres.
```

14 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: 2.74

Valor p: 0.06490

Conclusión: No hay diferencias significativas entre los tratamientos.

```
[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)")

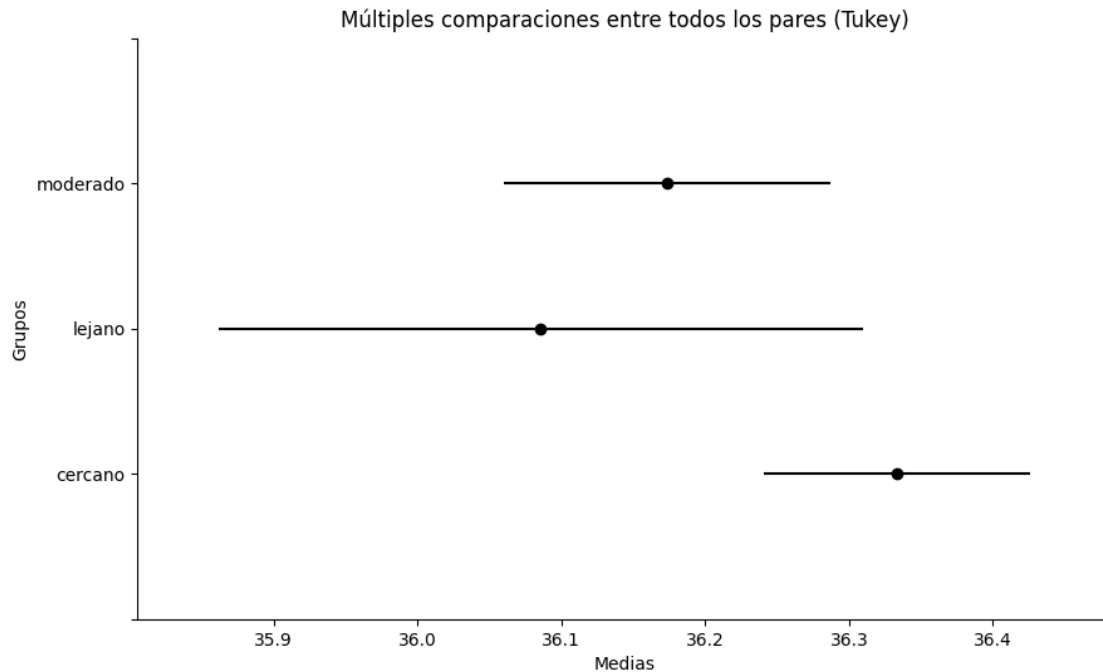
```

```

Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====
group1 group2 meandiff p-adj lower upper reject
=====
cercano lejano -0.2476 0.1581 -0.5638 0.0686 False
cercano moderado -0.1597 0.1636 -0.3656 0.0462 False
lejano moderado 0.0879 0.8141 -0.2493 0.4251 False
=====

```

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



```
[ ]: #CONCLUSIÓN: No se encontraron diferencias significativas entre estudiantes que  

     ↪ viven cerca, a distancia moderada o lejos del centro educativo.  

     #Esto podría estar influido por el acceso a medios de transporte o tecnología.
```

15 Género

```
[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: 0.01

Valor p: 0.93008

Conclusión: No hay diferencias significativas entre los tratamientos.

```

[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

```

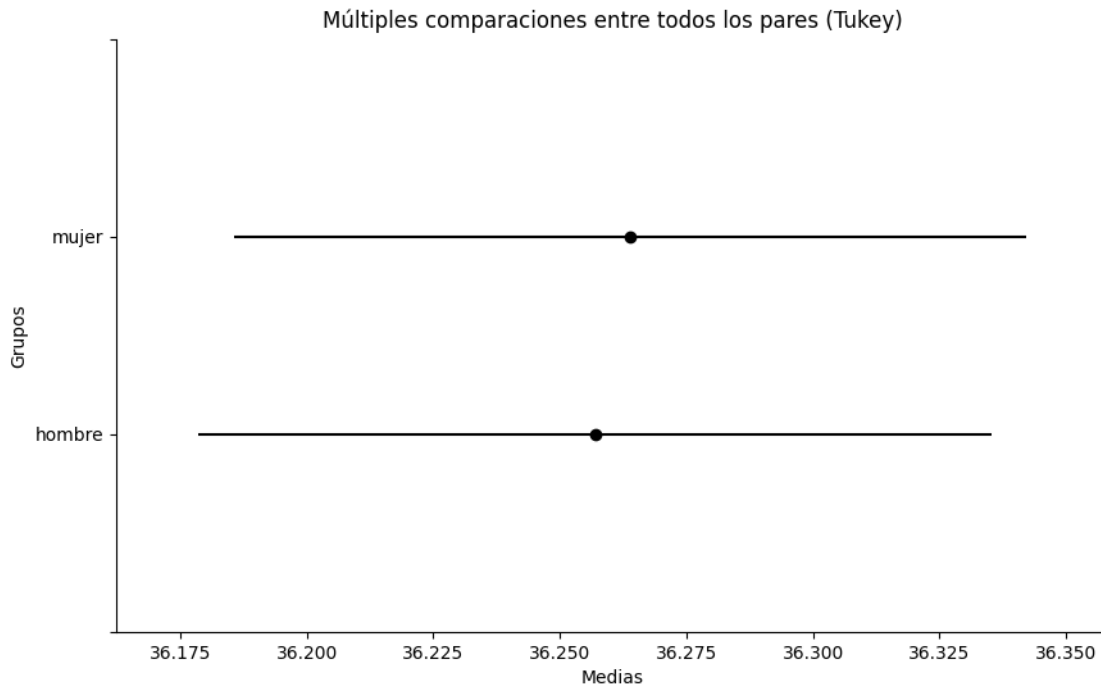
=====
group1 group2 meandiff p-adj lower upper reject
-----
hombre mujer 0.007 0.9301 -0.1495 0.1635 False
-----

```

```

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

```

[]: *#CONSLUSIÓN:El género no tuvo un impacto significativo en el desempeño*
↪académico, lo que confirma la tendencia hacia una igualdad en el rendimiento
↪entre hombres y mujeres.

#Resultados

El análisis realizado sobre las variables que influyen en el desempeño académico de los estudiantes ofreció una perspectiva sobre cuáles factores son muy importantes en el puntaje de los exámenes.

Los resultados revelaron que algunas variables tuvieron un impacto importante en el desempeño académico:

- **Acceso a Recursos:** Los estudiantes con un acceso a recursos educativos presentaron puntajes de examen más altos que aquellos con acceso bajo. Esto refuerza la importancia de garantizar materiales, tecnología y herramientas de aprendizaje accesibles para los estudiantes.
- **Calidad del Profesorado:** Se encontró que estudiantes que recibieron instrucción de un maestro con calidad obtuvieron mejores puntajes en los exámenes que aquellos con calidad media o baja. Esto destaca la necesidad de invertir en la capacitación docente para mejorar los resultados educativos.

Por otro lado, varias variables no mostraron una influencia importante en el desempeño académico:

- **Participación Parental:** Aunque se sugiere que un mayor involucramiento de los papás debería influir positivamente en el rendimiento, los resultados no mostraron diferencias entre los niveles de participación de los padres.
- **Actividades Extracurriculares:** No se encontraron diferencias significativas entre estudiantes que participan en actividades extracurriculares y aquellos que no. Aunque estas

actividades pueden tener beneficios en otras áreas, como el desarrollo social, su impacto directo en los puntajes de examen no fue tan importante.

- **Nivel de Motivación:** La motivación por sí sola no es suficiente sin recursos adecuados y apoyo externo.
- **Acceso a Internet:** La disponibilidad de internet tampoco mostró diferencias significativas entre grupos. Esto puede explicarse por variaciones en cómo los estudiantes utilizan este recurso, ya sea con fines académicos o para otras cosas que no tienen nada que ver con la escuela.
- **Tipo de Escuela:** No se encontraron diferencias significativas entre estudiantes de escuelas públicas y privadas, desafiando la percepción generalizada de que la educación privada garantiza mejores resultados.
- **Influencia de los Compañeros:** No hubo diferencias significativas en el desempeño según la influencia positiva, negativa o neutral de los compañeros. Aunque el entorno social es relevante para el bienestar emocional, su impacto directo en el rendimiento académico no fue concluyente.
- **Género:** Hombres y mujeres tuvieron un rendimiento académico similar, lo que refuerza la tendencia hacia una igualdad de oportunidades y resultados en la educación actual.
- **Distancia al Centro Educativo:** La proximidad al centro educativo tampoco mostró un impacto significativo en los puntajes de examen.
- **Nivel de Estudio de los Padres:** No hubo diferencias significativas entre los niveles de estudio de los padres (preparatoria, universidad, posgrado) en el desempeño académico de los estudiantes. Esto indica que el nivel educativo de los padres puede ser menos relevante que el apoyo emocional que brindan.
- **Discapacidad de Aprendizaje:** Se destaca la importancia de implementar estrategias inclusivas para garantizar que todos los estudiantes puedan alcanzar su máximo potencial.