

December 9, 2024

## 1 Introduccion

En este trabajo se analizara la variable de puntaje de examen utilizando ciertas pruebas como la prueba ANOVA y la prueba Tukey para de esta manera poder analizar si hay algun factor que pueda afectar de manera negativa a los alumnos y que debido a esto se reduzca el puntaje del examen de los mismos.

## 2 Marco Teórico

Prueba ANOVA:

La prueba ANOVA (Análisis de Varianza) es una técnica estadística que se utiliza para determinar si existen diferencias significativas entre las medias de tres o más grupos. Es especialmente útil en situaciones donde se evalúan múltiples factores o niveles de una variable independiente. Su resultado principal es el estadístico F y el valor p, que indican si las diferencias observadas son estadísticamente significativas. En este trabajo, la ANOVA se aplica para analizar la influencia de factores como el nivel de participación parental, acceso a recursos, actividades extracurriculares, entre otros, sobre el puntaje de los exámenes.

Prueba de Tukey:

Cuando la prueba ANOVA identifica diferencias significativas, se realiza un análisis post hoc, como la prueba de Tukey. Este método permite comparar las medias de todos los pares de grupos, indicando cuáles difieren significativamente. Es particularmente valioso en estudios con múltiples niveles de comparación, como los diferentes niveles de calidad del profesorado o motivación de los estudiantes.

Factores evaluados:

- Participación parental:  
Influencia de la implicación de los padres en el rendimiento académico de los estudiantes.
- Acceso a recursos:  
Impacto de contar con herramientas y materiales adecuados.
- Actividades extracurriculares:  
Relación entre actividades complementarias y el desempeño académico.

- Calidad del profesorado y tipo de escuela:  
Variables relacionadas con el entorno educativo.
- Factores personales y del hogar:  
Nivel de motivación, ingreso familiar, acceso a internet, entre otros.

### 3 Metodología

En este trabajo se realizarán diversas pruebas para poder analizar las variables que se van a emplear y de esta manera poder comparar estas variables con el puntaje de examen que en este caso es la variable en la que hay interés.

En este caso se realizará la prueba de ANOVA para poder verificar si hay alguna diferencia significativa entre las variables comparadas, también se empleará la prueba de Tukey para poder comparar las variables de estudio, además de poder ver de una manera más gráfica los resultados del problema.

Lo primero que se hará es una preparación de los datos, para esto se realizará una limpieza y preparación de los datos, que incluye:

- Eliminación de valores faltantes (NaN) en variables clave.
- Codificación de variables categóricas para su análisis.
- Verificación de la normalidad de las distribuciones y la homogeneidad de varianzas como requisitos para la ANOVA.

Lo siguiente que se realizará es un análisis estadístico, para esto se empleará lo siguiente.

- ANOVA: Se aplica para cada variable independiente categórica contra la variable dependiente (puntaje\_examen).
- Prueba de Tukey: Se realiza únicamente cuando ANOVA detecta diferencias significativas (valor  $p < 0.05$ ).
- Visualización: Los resultados se representan mediante gráficos que ilustran las diferencias entre grupos.

Finalmente se realizará una interpretación de resultados para cada una de las variables que se va a comparar, en este caso cada análisis incluye:

- Estadístico F y valor p para la ANOVA.
- Comparaciones específicas entre pares de grupos mediante la prueba de Tukey.
- Identificación de los factores que influyen significativamente en el puntaje de los exámenes y aquellos que no.

### 4 Resultados

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

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

df = pd.read_csv('https://raw.githubusercontent.com/Joseve-Sebastian/Unidad-4/
↪refs/heads/main/dataset.csv') # en github se sube el .csv y luego le pegas
↪y donde dice "Raw" le das click y copias el link de ese

df.columns
```

```
[82]: Index(['horas_estudio', 'asistencia', 'participacion_parental',
'acceso_a_recursos', 'actividades_extracurriculares', 'horas_suenio',
'calificaciones_previas', 'nivel_motivacion', 'acceso_internet',
'sesiones_tutoria', 'ingreso_familiar', 'calidad_profesorado',
'tipo_escuela', 'influencia_companieros', 'actividad_fisica',
'discapacidad_aprendizaje', 'nivel_estudio_padres', 'distancia',
'genero', 'puntaje_examen'],
dtype='object')
```

```
[83]: df
```

```
[83]:
```

	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

[6607 rows x 20 columns]

#### 4.1 participacion\_parental

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

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

```
[85]: # 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: 84.49

Valor p: 0.00000

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```
[86]: 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)

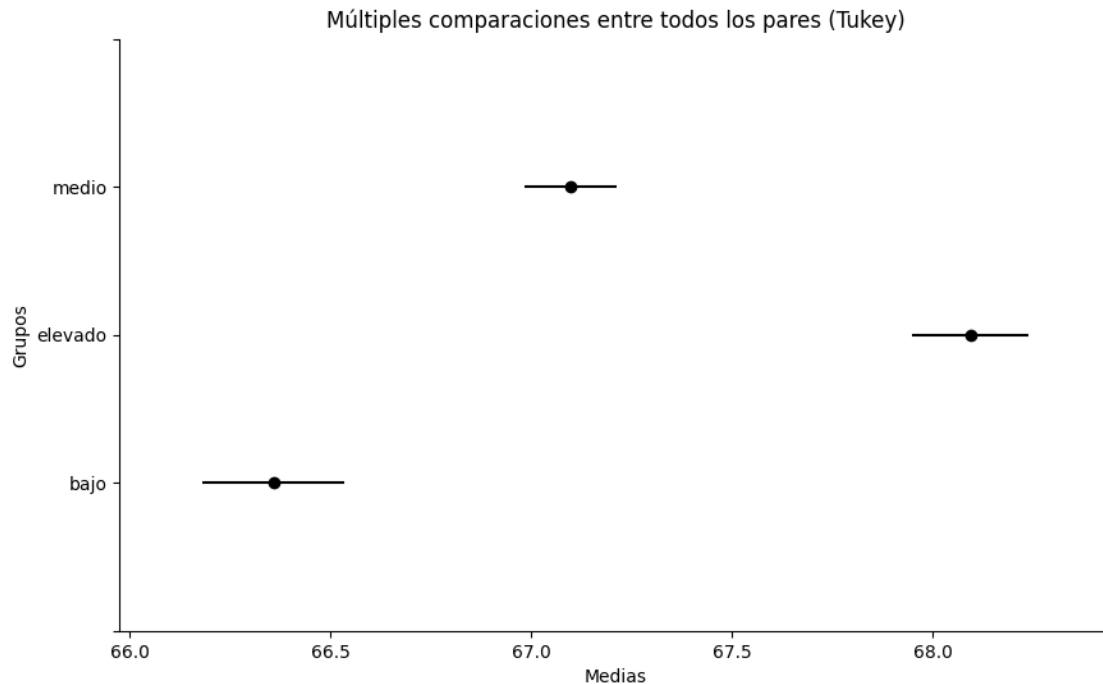
# Mostrar los resultados
print(tukey)

# 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 1.7345 0.0 1.4133 2.0557 True
bajo medio 0.7399 0.0 0.4487 1.0311 True
elevado medio -0.9946 0.0 -1.2528 -0.7364 True
-----
```

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



## 4.2 acceso\_a\_recursos

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

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

```
[88]: # 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: 83.50

Valor p: 0.00000

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```

[89]: 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    1.437    0.0 1.1762 1.6978   True
    bajo  medio    0.616    0.0 0.3384 0.8936   True
elevado  medio   -0.821    0.0 -1.123 -0.519   True
-----

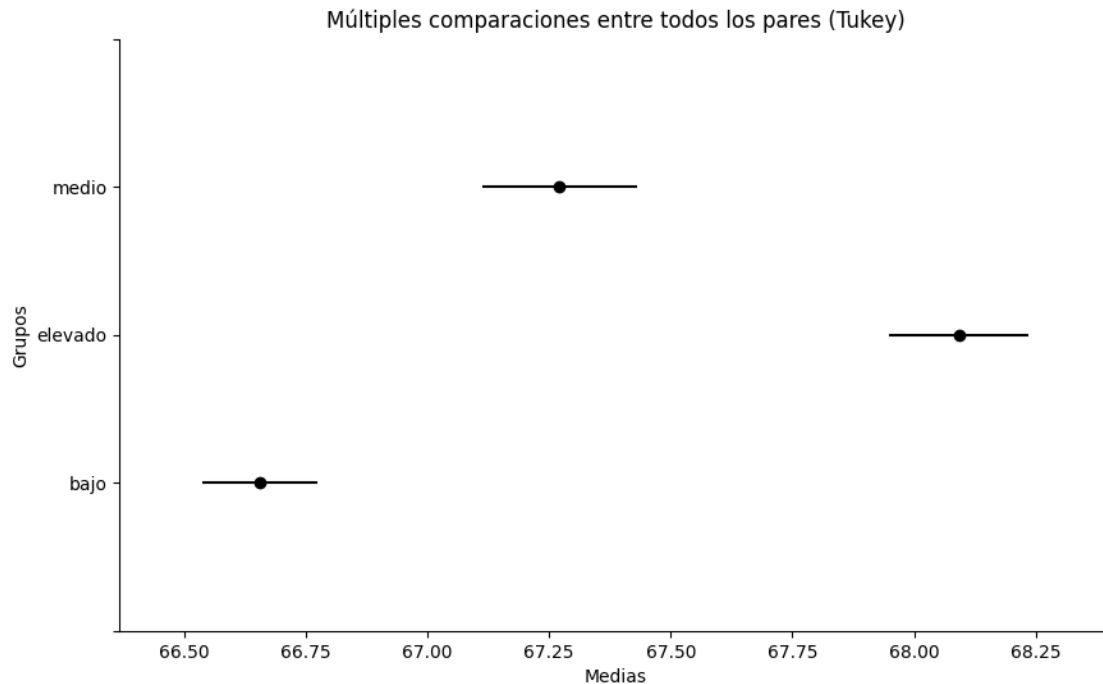
```

```

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

```





### 4.3 actividades\_extracurriculares

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

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

```
[91]: 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: 27.49

Valor p: 0.00000

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```
[92]: 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)

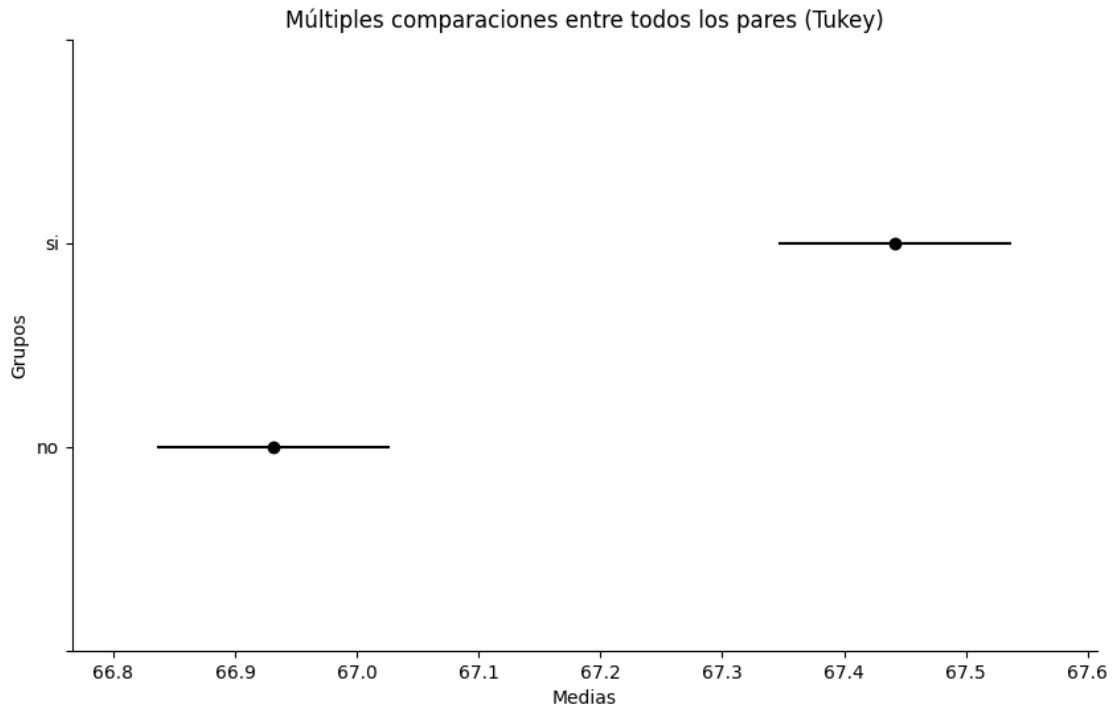
# Gráfico de las diferencias entre grupos
tukey.plot_simultaneous(ylabel="Grupos", xlabel="Medias")

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      0.5104   0.0 0.3196 0.7012   True
-----
```

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



#### 4.4 nivel\_motivacion

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

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

```
[94]: 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: 25.72

Valor p: 0.00000

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```
[95]: nivel_de_significancia = 0.05

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

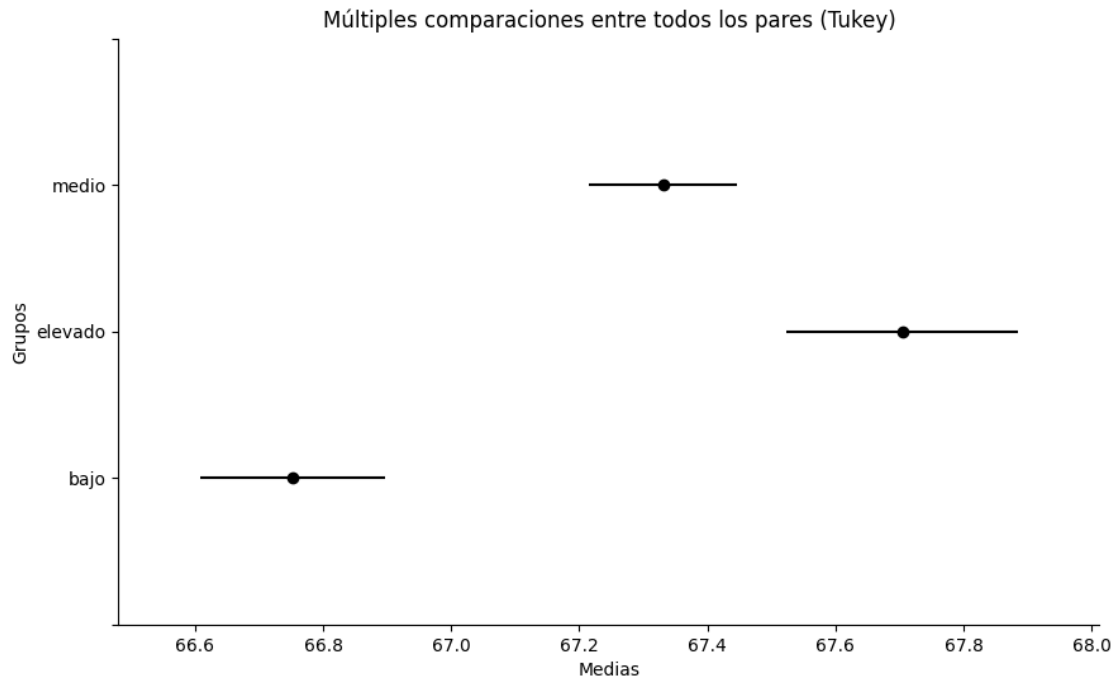
# Mostrar los resultados
print(tukey)

# 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  0.9521    0.0 0.6278  1.2765   True
    bajo  medio  0.5785    0.0 0.3191  0.8378   True
elevado  medio -0.3737  0.0085 -0.669 -0.0783   True
-----
```

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



#### 4.5 acceso\_internet

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

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

```
[97]: 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: 17.55

Valor p: 0.00003

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```
[98]: 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)

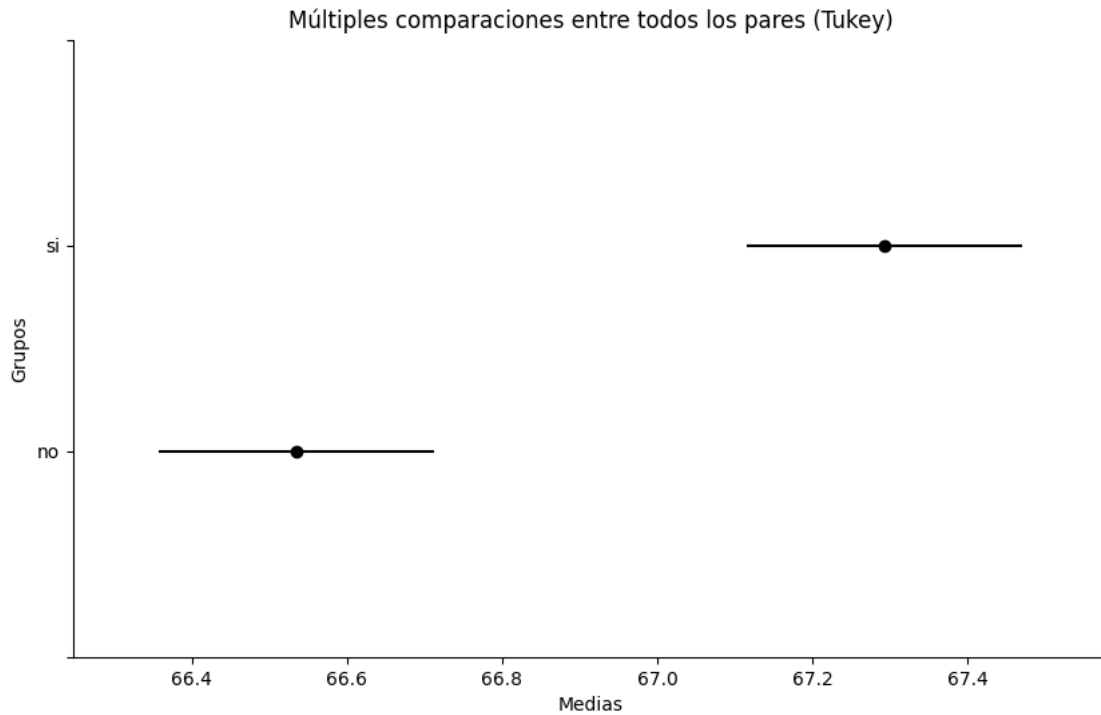
# Gráfico de las diferencias entre grupos
tukey.plot_simultaneous(ylabel="Grupos", xlabel="Medias")

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      0.7578   0.0 0.4032 1.1125   True
-----
```

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



#### 4.6 ingreso\_familiar

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

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

```
[100]: 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: 29.79

Valor p: 0.00000

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```

[101]: 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   0.994    0.0  0.6844  1.3035   True
    bajo  medio   0.4865    0.0  0.238  0.7351   True
elevado  medio  -0.5074  0.0004 -0.8171 -0.1977   True
=====

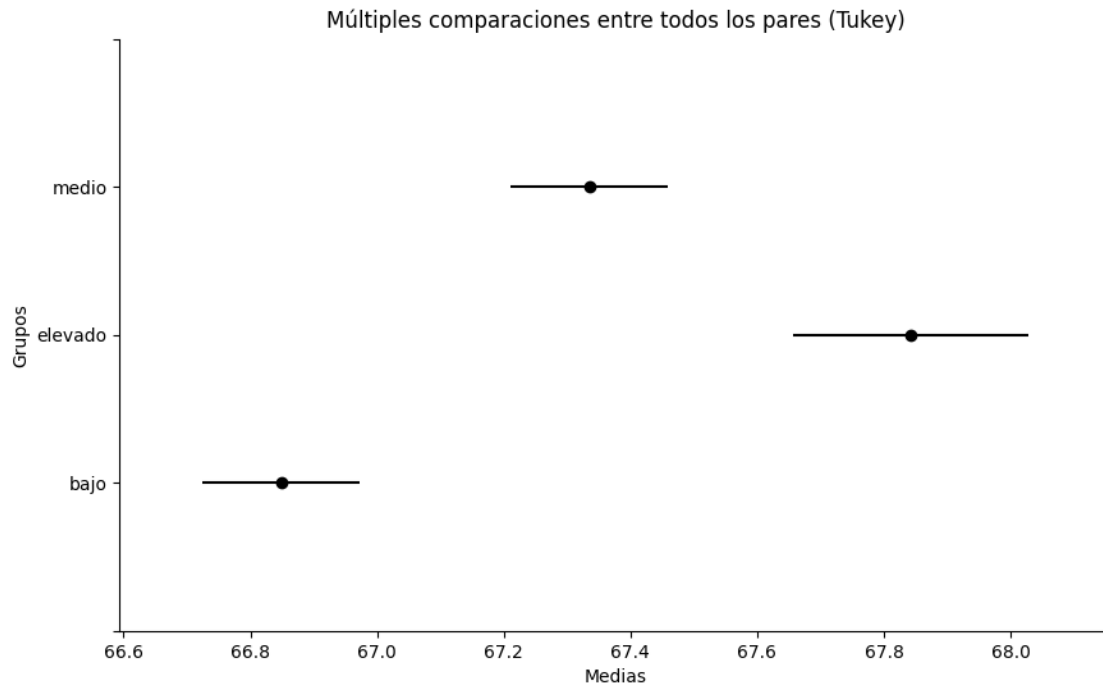
```

```

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

```





## 4.7 calidad\_profesorado

```
[102]: #para ver si hay valores nan
print(df.isnull().sum())
```

horas_estudio	0
asistencia	0
participacion_parental	0
acceso_a_recursos	0
actividades_extracurriculares	0
horas_suenio	0
calificaciones_previas	0
nivel_motivacion	0
acceso_internet	0
sesiones_tutoria	0
ingreso_familiar	0
calidad_profesorado	78
tipo_escuela	0
influencia_companieros	0
actividad_fisica	0
discapacidad_aprendizaje	0
nivel_estudio_padres	90
distancia	67
genero	0
puntaje_examen	0

dtype: int64

```
[103]: #eliminar los valores nan
df = df.dropna(subset=['calidad_profesorado', 'puntaje_examen'])

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: 17.88

Valor p: 0.00000

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```
[64]: nivel_de_significancia = 0.05

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

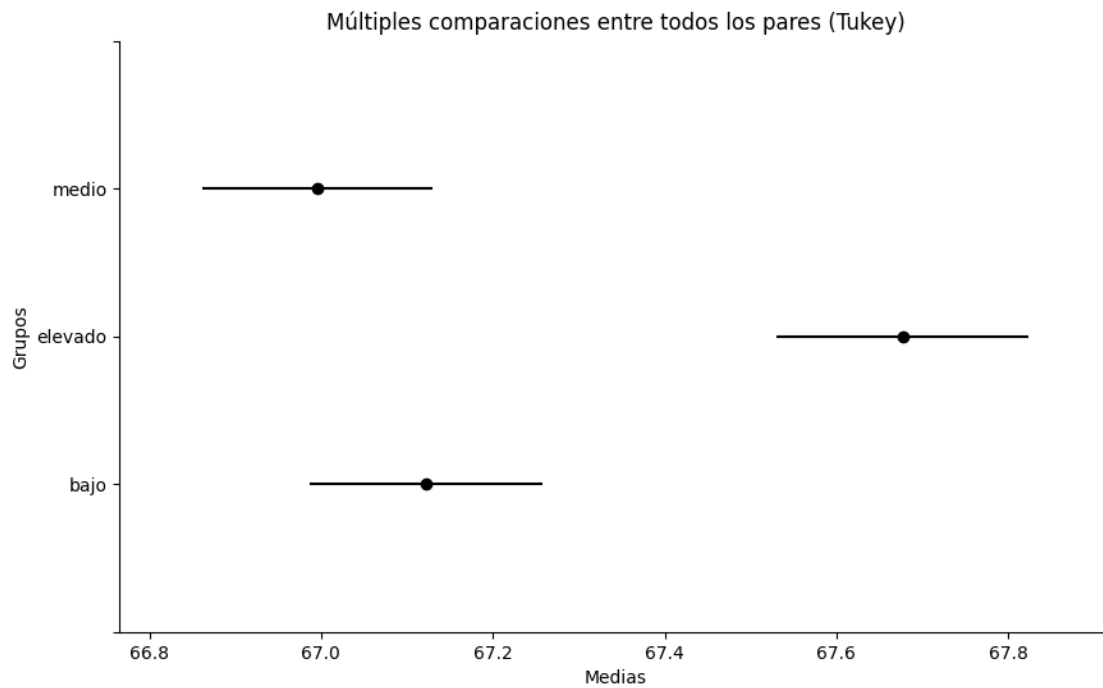
# Mostrar los resultados
print(tukey)

# 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  0.5545    0.0  0.2726  0.8365   True
bajo medio   -0.1267  0.5127 -0.3962  0.1428  False
elevado medio -0.6813    0.0 -0.9616 -0.4009   True
=====
```

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



#### 4.8 tipo\_escuela

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

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

```
[66]: 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: 0.58

Valor p: 0.44455

Conclusión: No hay diferencias significativas entre los tratamientos.

## 4.9 influencia\_companieros

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

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

```
[68]: 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: 33.41

Valor p: 0.00000

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```

[69]: 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)")

```

```

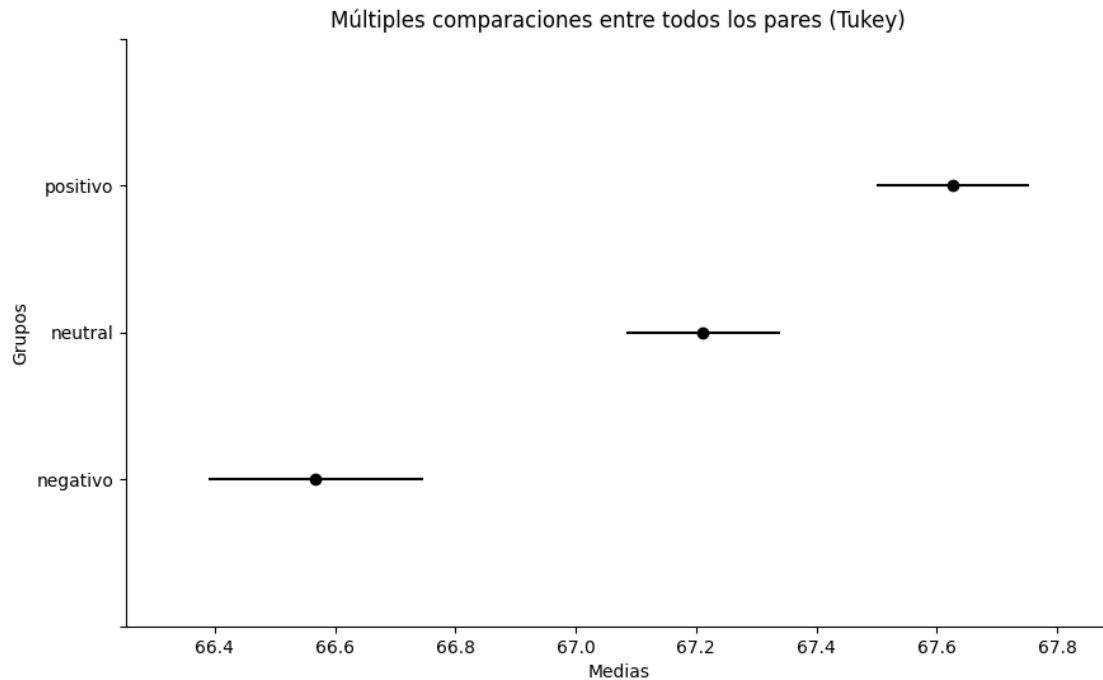
Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====
group1  group2  meandiff  p-adj  lower  upper  reject
-----
negativo  neutral    0.6441    0.0  0.3389  0.9492    True
negativo positivo    1.0587    0.0  0.7545  1.3629    True
neutral positivo    0.4146  0.0004  0.1614  0.6678    True
-----

```

```

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

```



#### 4.10 discapacidad\_aprendizaje

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

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

```
[71]: 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: 47.10

Valor p: 0.00000

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```

[72]: 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

```

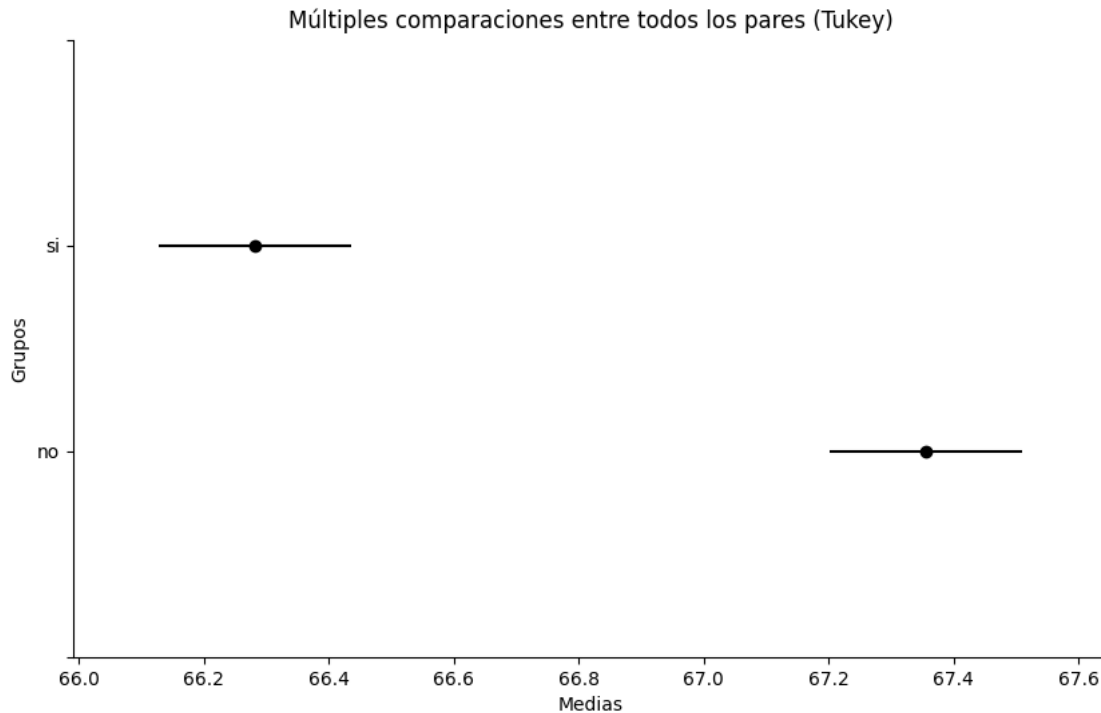
=====
group1 group2 meandiff p-adj lower upper reject
-----
    no     si  -1.0746   0.0 -1.3815 -0.7677   True
-----

```

```

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

```



#### 4.11 nivel\_estudio\_padres

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

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

```
[74]: df = df.dropna(subset=['nivel_estudio_padres', 'puntaje_examen'])
```

```
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: 36.13  
Valor p: 0.00000

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```

[75]: 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)")

```

```

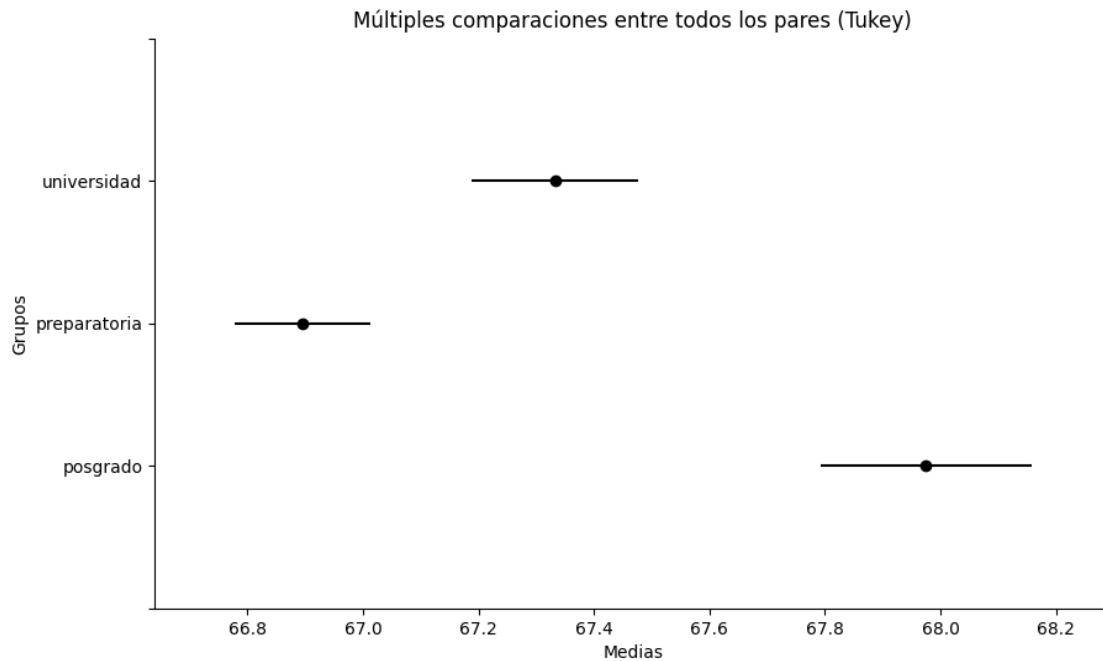
Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====
  group1      group2  meandiff p-adj  lower  upper  reject
-----
  posgrado preparatoria -1.0792   0.0 -1.3798 -0.7786   True
  posgrado  universidad  -0.642   0.0 -0.9684 -0.3155   True
preparatoria  universidad  0.4373 0.0003  0.1759  0.6986   True
-----

```

```

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

```



## 4.12 distancia

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

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

```
[77]: df = df.dropna(subset=['distancia', 'puntaje_examen'])
```

```
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.

```

[78]: 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)")

```

```

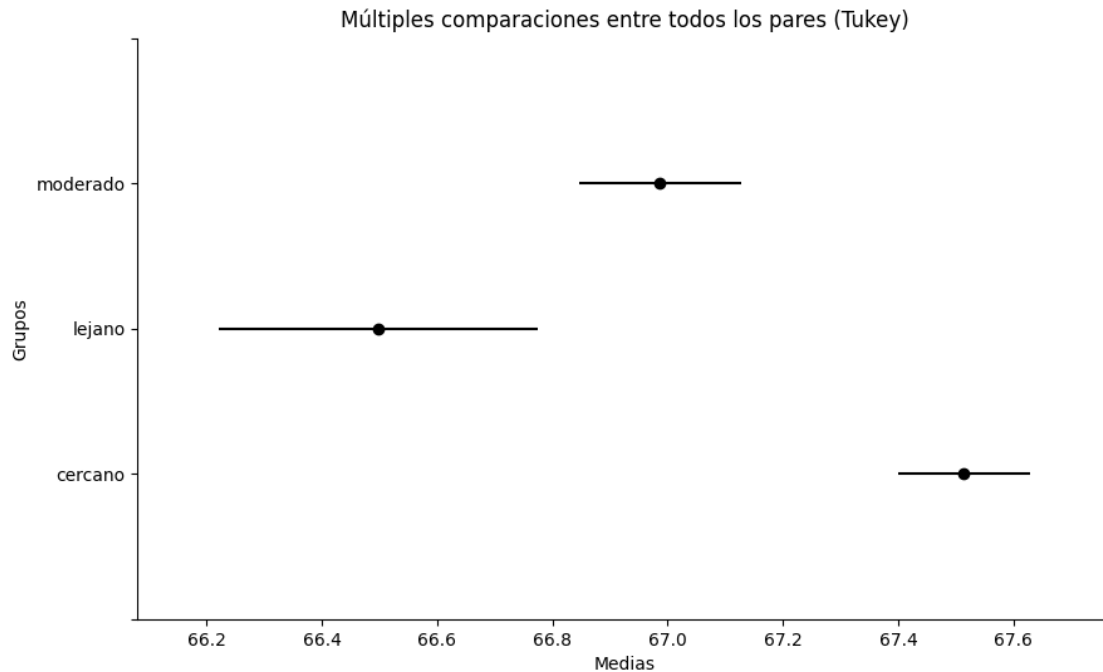
Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====
group1 group2 meandiff p-adj lower upper reject
-----
cercano lejano -1.0154 0.0 -1.407 -0.6237 True
cercano moderado -0.5272 0.0 -0.7822 -0.2722 True
lejano moderado 0.4882 0.017 0.0705 0.9059 True
-----

```

```

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

```



### 4.13 genero

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

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

```
[80]: 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.

## 5 Conclusion

Tomando en cuenta todas las variables analizadas se puede concluir que la mayoría de los factores analizados si influyen en la calificación del examen, pero hay ciertas variables como el genero y el tipo de escuela que no influyen en la variable que se esta analizando que es la calificación del examen, tomando en cuenta esto seria recomendable revisar los factores que si influyen en la calificación del examen y tratar de hacer algo para poder solucionar estas problemáticas