



EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO
NACIONAL DE MÉXICO

Instituto Tecnológico de Aguascalientes



INGENIERÍA
EN GESTIÓN EMPRESARIAL

UNIDAD IV

DISEÑO EXPERIMENTAL PARA UN FACTOR

ESTADISTICA INFERENCIAL II

JOSÉ CRUZ ESPARZA MUÑOZ

MARIANA LEDESMA CRUZ

Aguascalientes, Ags., 09 de diciembre del 2024



CamScanner

hp3xqee8a

December 9, 2024

0.1 INTRODUCCIÓN

0.2 En este trabajo se realizara un analisis de los factores que pueden influir es las calificaciones de los exámenes de los estudiantes, para realizarlo se consideraran distintas variables entre las cuales esta la participacion de los padres, el acceso a recursos, actividades extracurriculares, el nivel de motivación, acceso a internet, su tipo de escuela, los ingresos familiares, la influencia de los compañeros, la discapacidad de aprendizaje y el genero. Con el objetivo de identificar si todos estos factores tienen un impacto significativo en el rendimiento academico, utilizando métodos como ANOVA prueba de Tukey.

0.3 METODOLOGÍA

0.4 Datos: Se analizaron 6,607 registros de estudiantes que incluyen información sobre su entorno, hábitos y desempeño académico.

0.5 Herramientas: Se usaron Python y bibliotecas como scipy y statsmodels para realizar los análisis estadísticos.

0.6 Análisis:

0.7 -ANOVA: Para determinar si existen diferencias significativas entre los grupos definidos por cada variable.

0.8 -Pruebas de Tukey: Para identificar diferencias específicas entre pares de grupos.

0.9 Variables: Se analizaron factores como participación parental, acceso a recursos, actividades extracurriculares, y otras mencionadas previamente.

0.10 Pasos a seguir:

0.11 Paso 1: Obtencion de los datos para el analisis, donde se utiizara el siguiente codigo (import pandas as pd, import scipy.stats as stats)

0.12 Paso 2: Se utilizara ANOVA para efentiricar si existe dierencia entre los distintos grupos de las varbles:

0.13 Variables:

0.14

0.15 -Participacion parental:

##——-Grupos:

##Bajo

##Medio ##Elevado

##-Acceso a recursos: ##——-Grupos: ##Elevado ##Medio ##Bajo

##-Actividades Extacurriculares: ##——-Grupos: ##No ##Si

##-Nivel motivación: ##——-Grupos: ##Bajo ##Medio ##Elevado

##-Acceso a internet: ##——-Grupos: ##Si ##No

##-Ingreso Familiar: ##——-Grupos: ##Bajo ##Medio ##Elevado
 ##-Tipo escuela: ##——-Grupos: ##Privado ##Publico
 ##-Influencia compañeros: ##——-Grupos: ##Positivo ##Negativo ##Neutro
 ##-Discapacidad de aprendizaje: ##——-Grupos: ##No ##Si
 ##-Genero: ##——-Grupos: ##Hombre ##Mujer
 ##Paso 3: Aplicar la Pruebas de Tukey a cada variable.
 ##Paso 4: Obtener resultados.

0.16 MARCO TEÓRICO

0.17 La educación es un factor clave en el desarrollo de los individuos y de la sociedad. Diversos estudios han mostrado que variables externas como el apoyo parental, los recursos disponibles y la motivación personal pueden afectar significativamente el desempeño académico. Por otro lado, características como el género, el tipo de escuela o la presencia de discapacidades también han sido estudiadas para entender cómo influyen en los resultados escolares. Este proyecto se basa en técnicas estadísticas para determinar si estas diferencias son significativas y cómo pueden ser utilizadas para mejorar la calidad educativa.

##El análisis de varianza (ANOVA) de un factor es un método estadístico para examinar las diferencias en las medias de tres o mas grupos. ##Usualmente, el ANOVA de un factor se emplea cuando tenemos una unica variable o factor independiente y el objetivo es investigar si las variebles o diferentes niveles de este factor tienen efecto medible sobre una variable dependiente. ##Este método estadístico se utiliza para probar la hipótesis nula (H_0) de que tes o mas medias poblacionales son iguales frente a la hipóteisis alternativa (H_1) que al menos una de las medias es diferente.

##La prueba Tukey es una prueba estadística que se utiliza para comparar las medias de los tratamientos en un ANOVA. La prueba Tukey se usa en experimentos que implican un número elevado de coparaciones. ##Es de fácil cálculo puesto que se define un solo comparador, resultante del producto del error estándar de la media por el valor tabular en la tabla de Tukey usando como nemrador el número de tratamientos y como dominador los grados de ibertab de error.

```
[ ]: import pandas as pd
import scipy.stats as stats
```

```
[ ]: df = pd.read_csv('https://raw.githubusercontent.com/m4r14n4l3/Datos-pro.4/refs/heads/main/dataset%20(1).csv')
```

```
[ ]: df
```

```
[ ]:      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]

1 Participación parental

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

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

```
[ ]: # 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:}\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: 5.875479153325444e-37

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```

[ ]: from statsmodels.stats.multicomp import pairwise_tukeyhsd
import matplotlib.pyplot as plt

nivel_de_significancia = 0.05

# Prueba de Tukey
# Changed 'licenciatura' to 'nivel_educativo_de_los_padres' in the groups_
↪argument
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

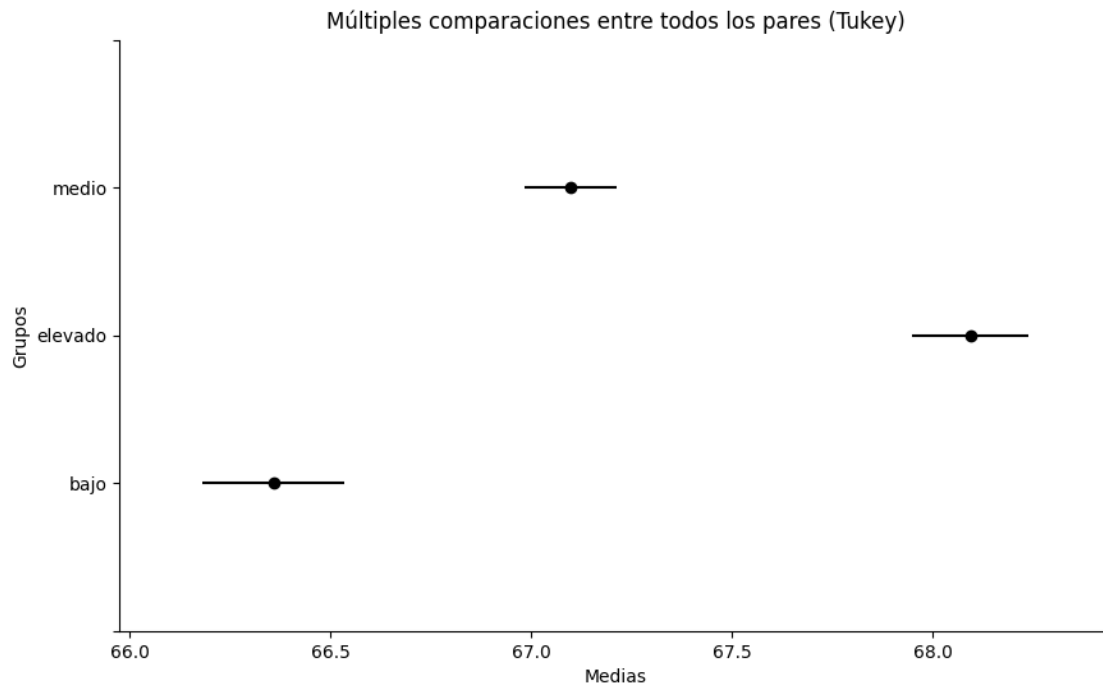
```

=====
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
```

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



1.1 RESULTADOS:

1.2 La participacion parental tiene influencia entre el nivele (bajo, medio, elevado) de los estudiantes se puede observar que los estudiantes con mayor participaci3n parental son los que obtienen mejores puntajes, por otro lado los que no tienen tanta participacion parental tienen un puntaje en las calificaciones mas bajo.

2 Acceso a recursos

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

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

```
[ ]: # Si todos los grupos tienen una distribuci3n 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:.2e}\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: 1.5332234901623123e-36

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```

[ ]: from statsmodels.stats.multicomp import pairwise_tukeyhsd
import matplotlib.pyplot as plt

nivel_de_significancia = 0.05

# Prueba de Tukey
# Changed 'licenciatura' to 'nivel_educativo_de_los_padres' in the groups_
↪argument
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)")

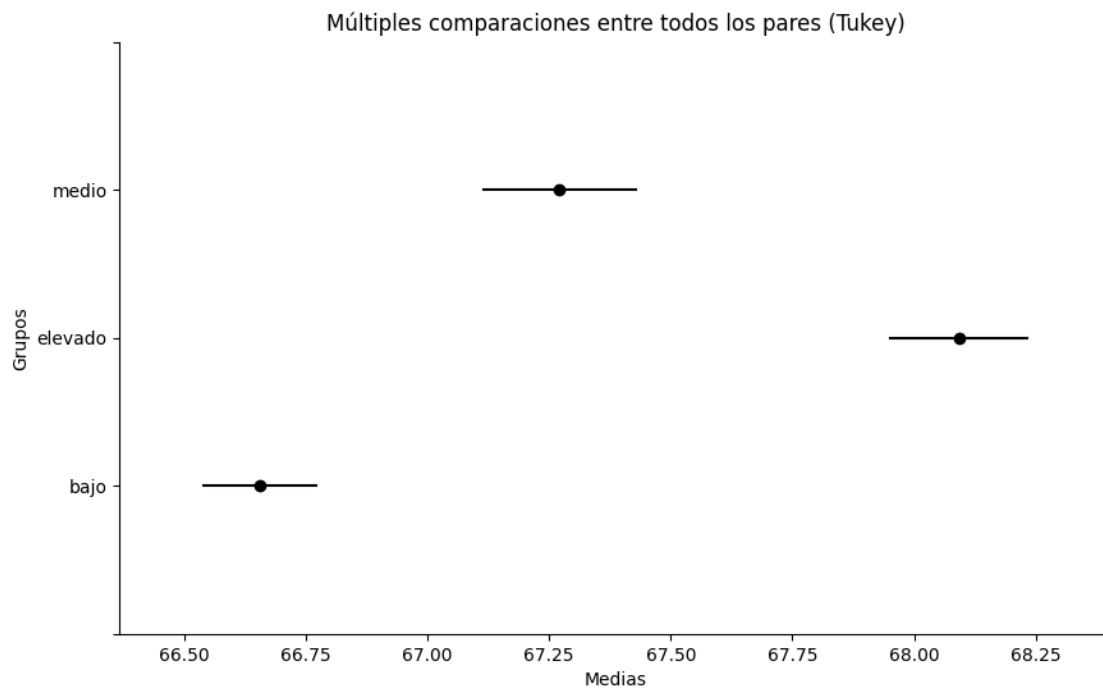
```

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

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



2.1 ACCESO A RECURSOS

2.2 En el caso del factor acceso a recursos se observa que existe diferencia entre los niveles (elevado, medio, bajo) en los que los estudiantes tienen el acceso a recursos, esto nos demuestra que aquellos estudiantes que tienen un nivel elevado del uso del internet tienen puntajes más altos por ende los que tienen un nivel bajo su puntaje es más bajo en las calificaciones.

3 Actividades extracurriculares

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

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

```
[ ]: # 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:.2f}\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: 1.626677707731656e-07

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```

[ ]: from statsmodels.stats.multicomp import pairwise_tukeyhsd
import matplotlib.pyplot as plt

nivel_de_significancia = 0.05

# Prueba de Tukey
# Changed 'licenciatura' to 'nivel_educativo_de_los_padres' in the groups
↪argument
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")

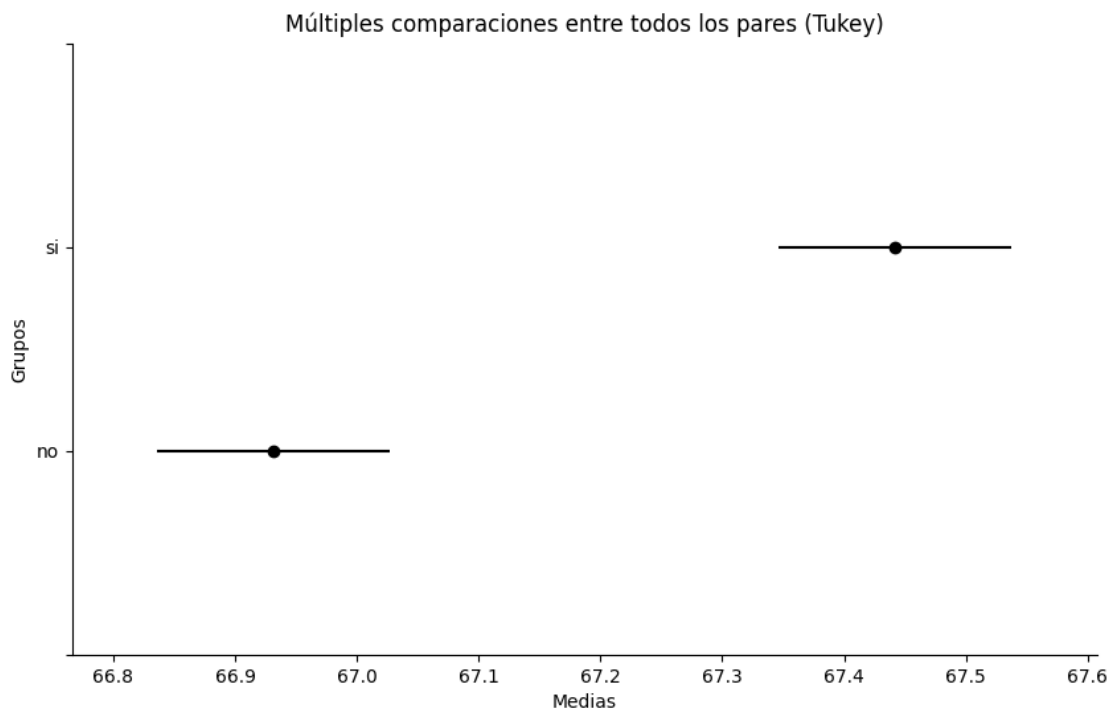
plt.gca().spines['right'].set_visible(False) # derecha
plt.gca().spines['top'].set_visible(False)   # superior
plt.title("Múltiples comparaciones entre todos los pares (Tukey)")

```

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
-----
```

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



3.1 ACTIVIDADES EXTRACURRICILARES

3.2 La participación de los estudiantes en actividades extracurriculares y el no participar en ellas tienen una diferencia significativa, pero el análisis de este factor nos demuestra que el hecho de que los estudiantes participen en estas actividades mejora su calificación en el examen en cambio los que no participan en las actividades su calificación es baja.

4 Nivel motivación

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

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

```
[ ]: # 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:.2e}")

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: 7.492319380241663e-12

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```
[ ]: from statsmodels.stats.multicomp import pairwise_tukeyhsd
import matplotlib.pyplot as plt

nivel_de_significancia = 0.05

# Prueba de Tukey
# Changed 'licenciatura' to 'nivel_educativo_de_los_padres' in the groups
↪argument
tukey = pairwise_tukeyhsd(endog=df['puntaje_examen'],
↪groups=df['nivel_motivacion'], alpha=nivel_de_significancia)

# Mostrar los resultados
print(tukey)

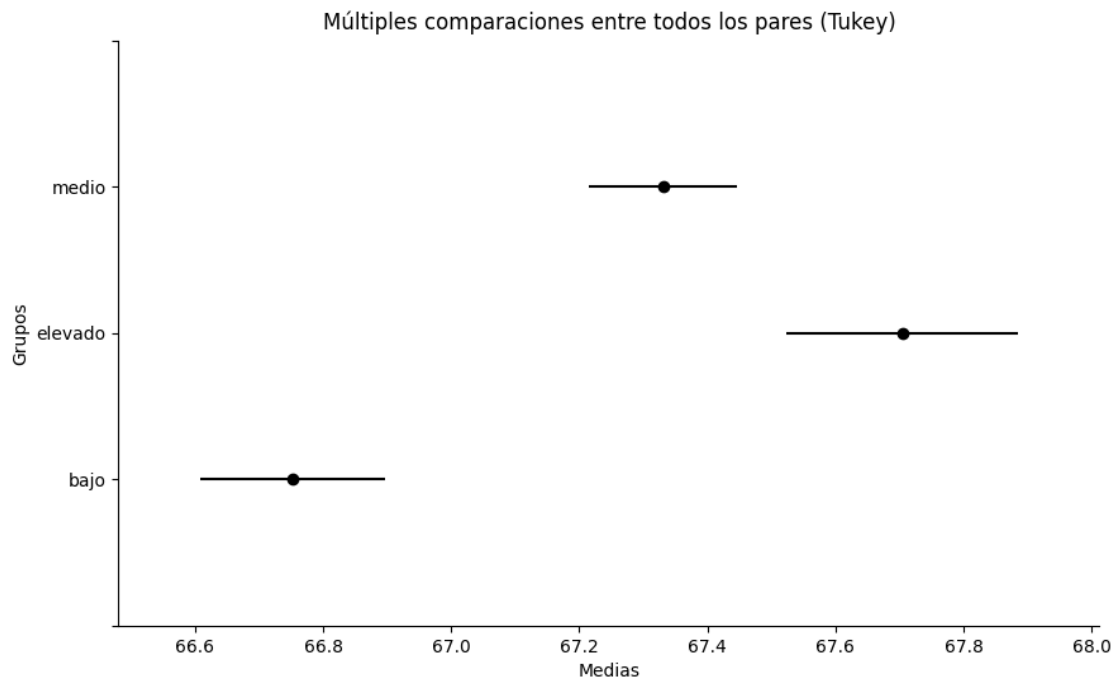
# Gráfico de las diferencias entre grupos
tukey.plot_simultaneous(ylabel="Grupos", xlabel="Medias")
```

```
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

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



4.1 NIVEL DE MOTIVACIÓN

4.2 De acuerdo a tres niveles (medio, elevado, bajo) los cuales son los niveles de motivación que tiene los estudiantes se encuentra que existe una diferencia significativa entre ellos, lo que nos indica que si los estudiantes tienen un nivel elevado de motivación se verá reflejado en sus calificaciones y estas serán mejores, pero si no existe motivación su calificación será mala.

5 Acceso a internet

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

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

```
[ ]: # 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:.2f}\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: 2.8385046310284837e-05

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```
[ ]: from statsmodels.stats.multicomp import pairwise_tukeyhsd
import matplotlib.pyplot as plt

nivel_de_significancia = 0.05
```

```

# Prueba de Tukey
# Changed 'licenciatura' to 'nivel_educativo_de_los_padres' in the groups
↳ argument
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)")

```

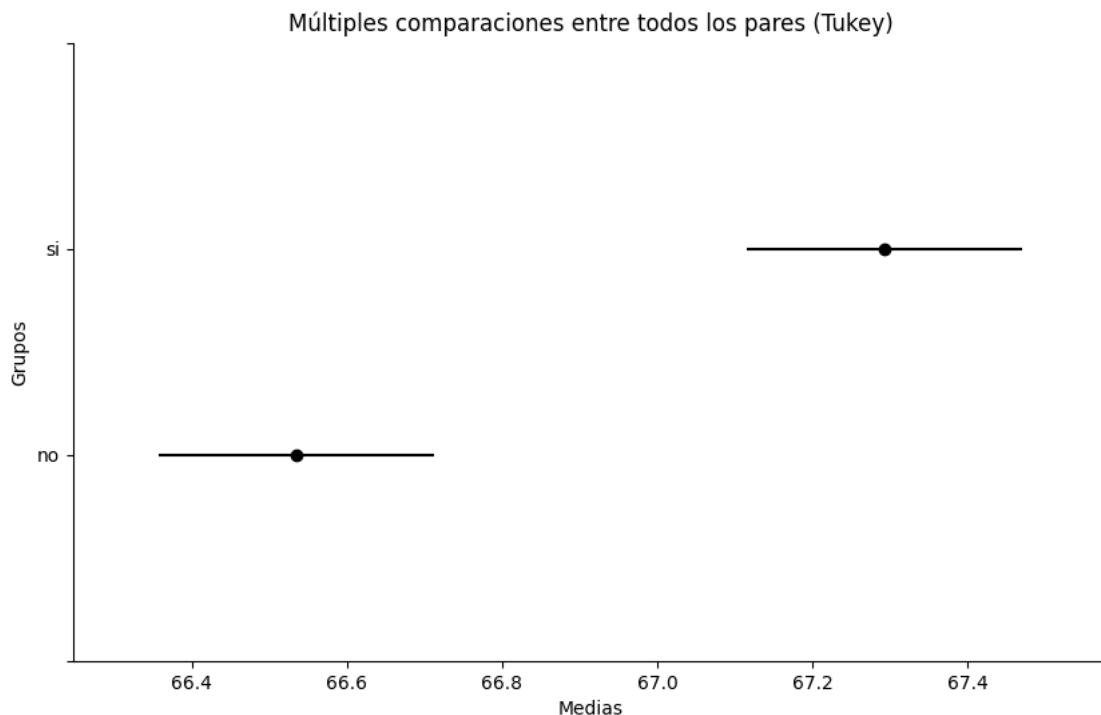
Multiple Comparison of Means - Tukey HSD, FWER=0.05

```

=====
group1 group2 meandiff p-adj lower upper reject
-----
no      si      0.7578    0.0 0.4032 1.1125   True
-----

```

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



5.1 ACCESO A INTERNET

5.2 Entre los factores de que los estudiantes tengan acceso a internet y el no tener acceso a internet existe una diferencia significativa, los resultados nos arrojan que todo aquel estudiante que tiene acceso a internet obtendra un porcentaje de calificación más alto por el contrario aquel que no tenga acceso a internet obtendra una calificación mas baja.

6 Ingreso familiar

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

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

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

nivel_de_significancia = 0.05

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

# Interpretar los resultados
print("Resultados del ANOVA:")
print(f"F-Estadístico: {anova_result.statistic:.2f}")
print(f"Valor p: {anova_result.pvalue:.2f}\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: 1.3143686049770217e-13

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```
[ ]: from statsmodels.stats.multicomp import pairwise_tukeyhsd
import matplotlib.pyplot as plt

nivel_de_significancia = 0.05
```

```

# Prueba de Tukey
# Changed 'licenciatura' to 'nivel_educativo_de_los_padres' in the groups
↳ argument
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)")

```

```

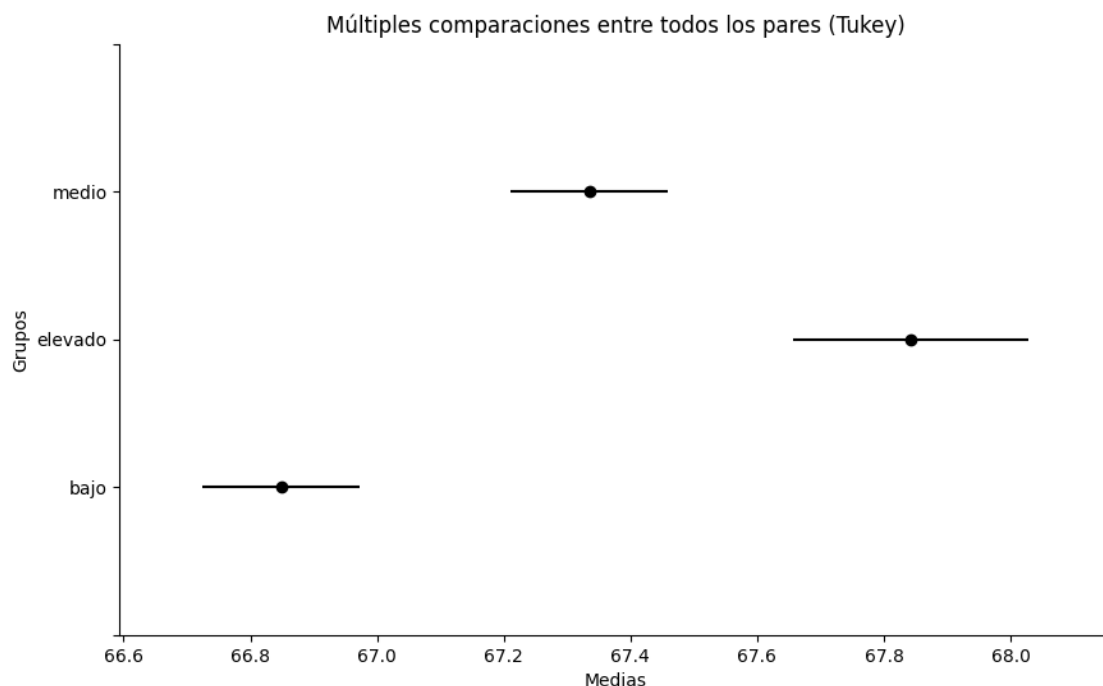
Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====
group1 group2 meandiff p-adj lower upper reject
-----
bajo elevado 0.994 0.0 0.6844 1.3035 True
bajo medio 0.4865 0.0 0.238 0.7351 True
elevado medio -0.5074 0.0004 -0.8171 -0.1977 True
=====

```

```

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

```



6.1 INGRESO FAMILIAR

6.2 Los ingresos de las familias estan clasificados en: nivel medio, nivel elevado y nivel bajo entre los cales existe una diferencia significativa, pero el hecho de que una familia tenga ingresos elevados tiene efectos posotivos en los estudiantes obteniendo mejores resultados, pero si los ingresos de las familias estan en un nivel bajo el estudiante tendra calificaciones bajas.

7 Tipo escuela

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

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

```
[ ]: # 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'] == 'privado']['puntaje_examen'],
    df[df['tipo_escuela'] == 'publico']['puntaje_examen']
)

# Interpretar los resultados
print("Resultados del ANOVA:")
print(f"F-Estadístico: {anova_result.statistic:.2f}")
print(f"Valor p: {anova_result.pvalue:.2f}\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: 0.52

Valor p: 0.47231811262159584

Conclusión: No hay diferencias significativas entre los tratamientos.

```
[ ]: from statsmodels.stats.multicomp import pairwise_tukeyhsd
import matplotlib.pyplot as plt

nivel_de_significancia = 0.05

# Prueba de Tukey
# Changed 'licenciatura' to 'nivel_educativo_de_los_padres' in the groups
↳ argument
tukey = pairwise_tukeyhsd(endog=df['puntaje_examen'],
↳ groups=df['tipo_escuela'], alpha=nivel_de_significancia)

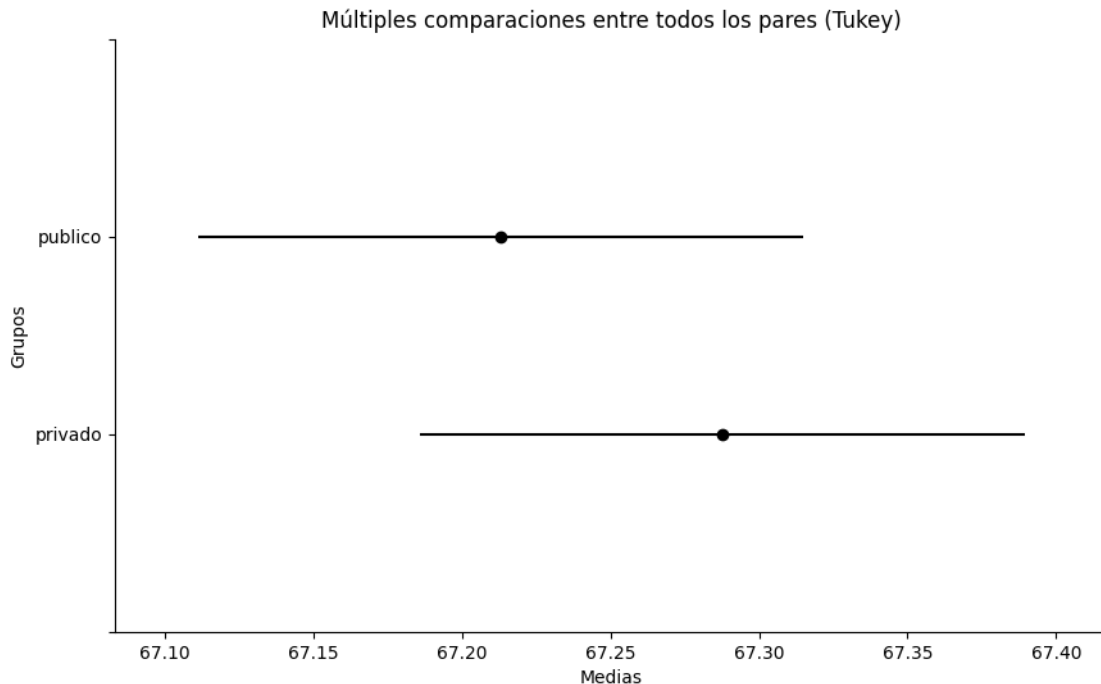
# Mostrar los resultados
print(tukey)

# 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 -0.0748 0.4723 -0.2788 0.1292 False
-----
```

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



7.1 TIPO DE ESCUELA

7.2 Al analizar el tipo de escuela de los estudiantes los resultados fueron que este factor no afecta en que el estudiante se encuentre en una escuela publica o una escuela privada los resultados en las calificaciones no tendra tanta variación.

8 Influencia compañeros

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

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

```
[ ]: # 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:}\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: 34.05

Valor p: 1.9368909270985136e-15

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```

[ ]: from statsmodels.stats.multicomp import pairwise_tukeyhsd
import matplotlib.pyplot as plt

nivel_de_significancia = 0.05

# Prueba de Tukey
# Changed 'licenciatura' to 'nivel_educativo_de_los_padres' in the groups
↪argument
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)")

```

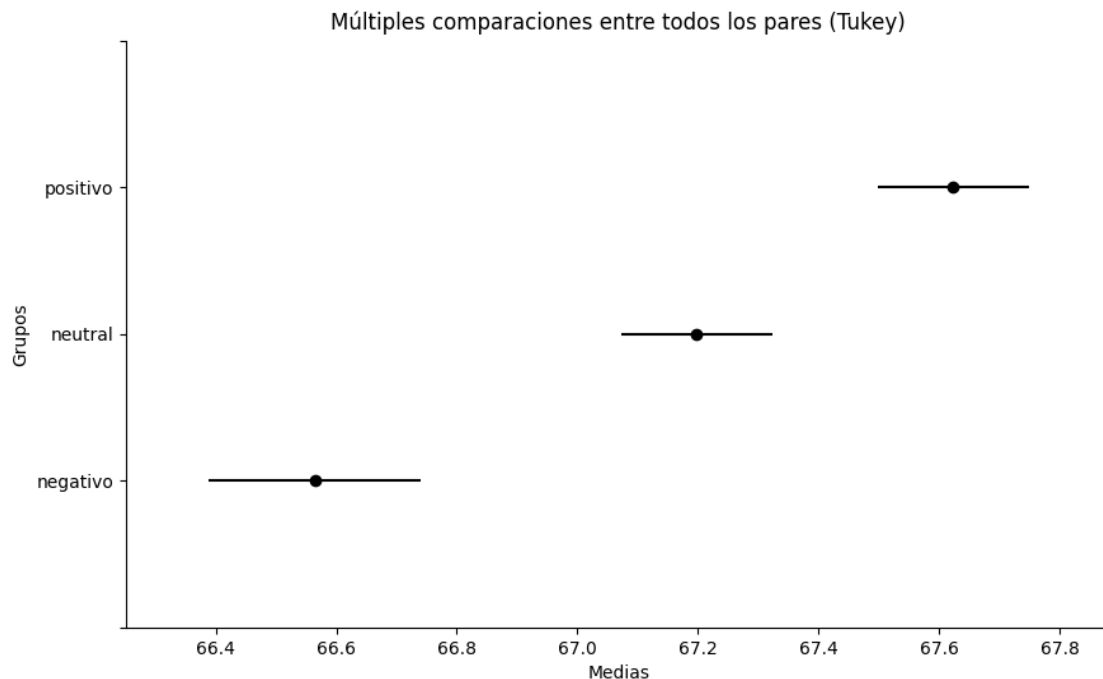
Multiple Comparison of Means - Tukey HSD, FWER=0.05

```

=====
group1  group2  meandiff  p-adj  lower  upper  reject
-----
negativo  neutral    0.6336    0.0   0.331  0.9363    True
negativo positivo    1.0589    0.0  0.7572  1.3606    True
neutral positivo    0.4253  0.0002  0.1743  0.6763    True
-----

```

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



8.1 INFLUENCIA DE COMPAÑEROS

Anova no muestra que existe diferencia significativa entre los grupos de niveles de influencia de motivación positiva, negativa y neutral, por lo que en tukey podemos ver que una influencia positiva mejora los puntajes de los estudiantes mientras que una influencia negativa los disminuye.

9 Discapacidad aprendizaje

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

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

```
[ ]: # 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:.2e}\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: 48.14

Valor p: 4.338623930738945e-12

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```

[ ]: from statsmodels.stats.multicomp import pairwise_tukeyhsd
import matplotlib.pyplot as plt

nivel_de_significancia = 0.05

# Prueba de Tukey
# Changed 'licenciatura' to 'nivel_educativo_de_los_padres' in the groups_
↪argument
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

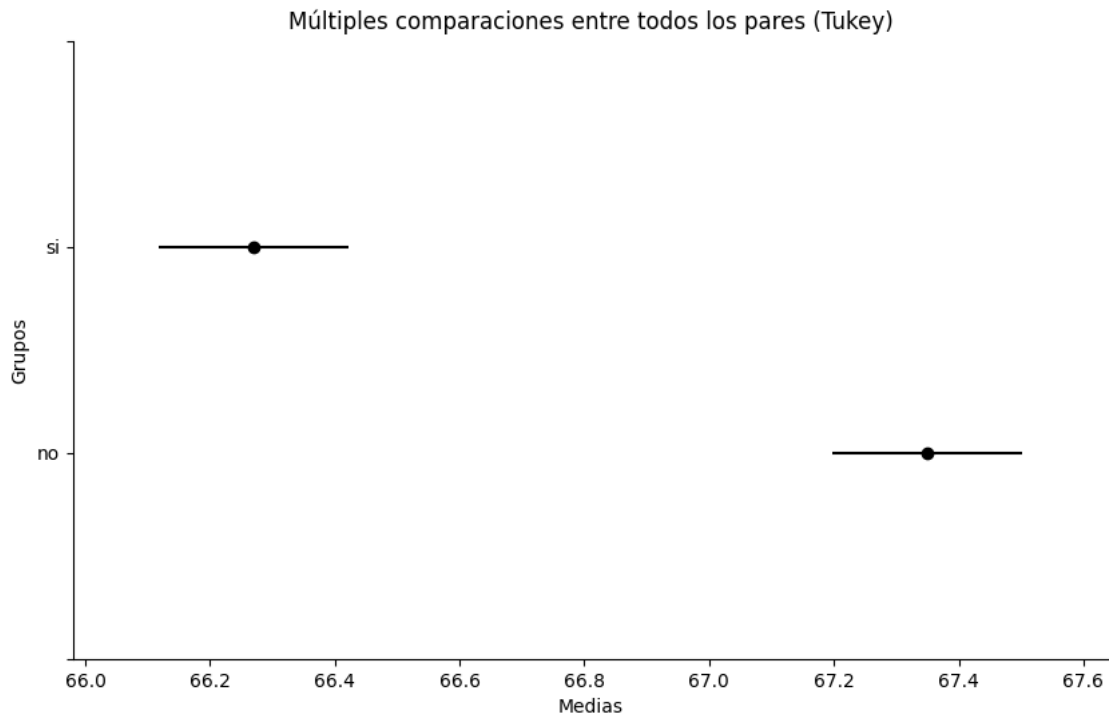
```

=====
group1 group2 meandiff p-adj lower upper reject
-----
no      si    -1.0786   0.0  -1.3834 -0.7739   True
-----

```



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



##DISCAPACIDAD DE APRENDIZAJE

##Todos aquellos estudiantes que no tienen ninguna discapacidad obtienen mejores resultados que los que tienen discapacidad de aprendizaje, nos muestra que existe una diferencia significativa entre ambos grupos.

10 Genero

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

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

```
[ ]: # 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:}\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: 0.03

Valor p: 0.8688153297341272

Conclusión: No hay diferencias significativas entre los tratamientos.

```

[ ]: from statsmodels.stats.multicomp import pairwise_tukeyhsd
import matplotlib.pyplot as plt

nivel_de_significancia = 0.05

# Prueba de Tukey
# Changed 'licenciatura' to 'nivel_educativo_de_los_padres' in the groups_
↪argument
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)")

```

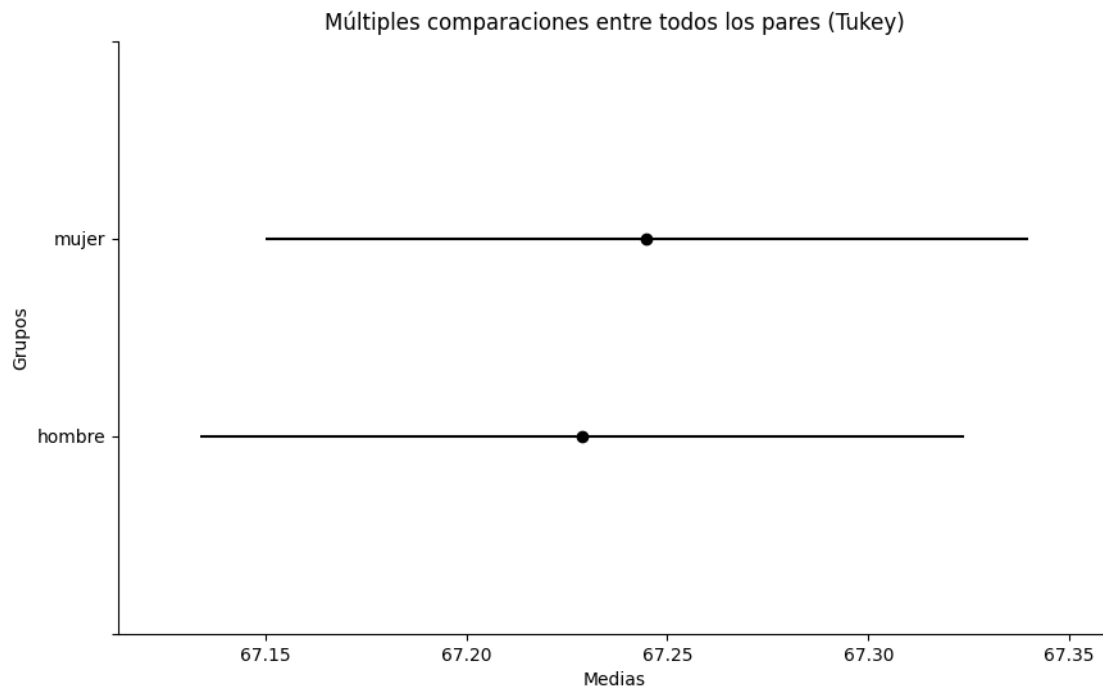
Multiple Comparison of Means - Tukey HSD, FWER=0.05

```

=====
group1 group2 meandiff p-adj  lower  upper reject
-----
hombre  mujer    0.016 0.8688 -0.1739 0.206  False
-----

```

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



##GENERO

##Anova no muestra que no existe diferencia significativa entre los grupos de hombres y mujeres por lo que esto no tiene influencia en la calificación de los estudiantes.

10.1 CONCLUSIÓN:

##Con la realización de este trabajo se puede observar que la mayoría de las pruebas Anovas realizadas de en la mayoría de ellas la diferencia es significativa por lo que las pruebas Tukey muestran algunos de los grupos con mayor significancia.

##Este análisis confirmó que factores como la participación parental, el acceso a recursos y el nivel de motivación tienen un impacto significativo en el rendimiento académico. También destaca la importancia de las actividades extracurriculares, el acceso a internet y la influencia positiva de los compañeros. Por otro lado, no se encontraron diferencias en el desempeño según el tipo de escuela o el género. Estos resultados subrayan la necesidad de intervenciones que fortalezcan los factores positivos y reduzcan las desigualdades en el entorno educativo.