

December 10, 2024

## 1 INTRODUCCIÓN

En el presente trabajo estaremos realizando distintos ejercicios con la prueba ANOVA y Tukey siendo utilizada si dentro de la base de datos existen diferencias significativas entre las variables de ellos. El trabajo consiste en una serie de datos que se busca realizar el estudio según la prueba de examen basado en datos cualitativos de cierta población. Se estará realizando la herramienta de Python debido que visto anteriormente con trabajos pasados sabemos que esta es de gran ayuda para realizar los trabajos de manera más clara y sencilla ya que los beneficios que ofrece al momento de ejecutar los códigos con la información necesaria nos otorga los datos y resultados más concisos, en este proyecto se llevarán a cabo la práctica de las pruebas ya mencionadas anteriormente, ya que se sabe que en conjunto las pruebas ANOVA y Tukey son herramientas para explorar las relaciones y diferencias entre varios grupos.

## 2 MARCO TEÓRICO

### *Prueba ANOVA*

La prueba ANOVA (Análisis de la Varianza) es una técnica estadística utilizada para comparar las medias de tres o más grupos independientes y determinar si al menos uno de esos grupos es significativamente diferente de los demás. A través de este análisis, se evalúa si las diferencias observadas en las medias de los grupos son mayores de lo que se esperaría por casualidad. ANOVA genera un valor F (estadístico F), que es la razón entre la variabilidad entre los grupos y la variabilidad dentro de los grupos. Si el valor de F es mayor que un valor crítico determinado por una tabla de distribución F, se rechaza la hipótesis nula, lo que sugiere que existen diferencias significativas entre las medias de los grupos.

### *Prueba Tukey*

La prueba de Tukey, también conocida como prueba de Tukey HSD (Honest Significant Difference), es una técnica estadística utilizada para realizar comparaciones por pares entre las medias de varios grupos después de haber obtenido un resultado significativo en una prueba ANOVA. Su principal objetivo es identificar cuáles grupos son significativamente diferentes entre sí, controlando el error tipo I (la probabilidad de rechazar incorrectamente una hipótesis nula verdadera) cuando se realizan múltiples comparaciones. Cuando ANOVA muestra que hay diferencias significativas entre los grupos, la prueba de *Tukey* se aplica para examinar qué grupos específicamente difieren.

Valor  $p$

El valor p es una herramienta fundamental en la estadística inferencial que ayuda a determinar si los resultados de un estudio son lo suficientemente fuertes como para rechazar la hipótesis nula. Su interpretación depende de la comparación con el umbral de significancia, y su valor refleja la probabilidad de obtener los resultados observados bajo la suposición de que la hipótesis nula es cierta.

### 3 METODOLOGÍA

En el presente trabajo se estará realizando el estudio de analizar los factores que afectan al rendimiento académico de distintos estudiantes tomando en cuenta como única variable el puntaje del examen que se haya obtenido para cada uno, el estudio se llevará a cabo con ayuda de las dos pruebas ya mencionadas, ANOVA y Tukey con la base de datos cualitativos ya proporcionada.

Valores cualitativos para el estudio: - Participación parental: Nivel de participación de los padres en la educación del estudiante (Baja, Media, Alta). - Acceso a recursos: Disponibilidad de recursos educativos (Bajo, Medio, Alto). - Actividades extracurriculares: Participación en actividades extracurriculares (Sí, No). - Nivel motivación: Nivel de motivación del estudiante (Bajo, Medio, Alto). - Acceso internet: Disponibilidad de acceso a internet (Sí, No). - Ingreso familiar: Nivel de ingresos familiares (Bajo, Medio, Alto). - Calidad profesorado: Calidad de los profesores (Baja, Media, Alta). - Tipo escuela: Tipo de escuela asistida (Pública, Privada). - Influencia compañeros: Influencia de los compañeros en el rendimiento académico (Positiva, Neutral, Negativa). - Discapacidad aprendizaje: Presencia de discapacidades de aprendizaje (Sí, No). - Nivel estudios padres: Nivel educativo más alto de los padres (Preparatoria, Universidad, Posgrado). - Distancia: Distancia desde el hogar a la escuela (Cerca, Moderada, Lejos). género - Género: Género del estudiante (Masculino, Femenino).

### RESULTADOS

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

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

```
[1]:
```

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

6606	15	67	medio	bajo
------	----	----	-------	------

	actividades_extracurriculares	horas_suenio	calificaciones_previas	\
0	no	7	73	
1	no	8	59	
2	si	7	91	
3	si	8	98	
4	si	6	65	
...	...	...	...	
6602	no	7	76	
6603	no	8	81	
6604	si	6	65	
6605	si	6	91	
6606	si	9	94	

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

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

	actividad_fisica	discapacidad_aprendizaje	nivel_estudio_padres	\
0	3	no	preparatoria	
1	4	no	universidad	
2	4	no	posgrado	
3	4	no	preparatoria	
4	4	no	universidad	

...	...	...	...
6602	2	no	preparatoria
6603	2	no	preparatoria
6604	2	no	posgrado
6605	3	no	preparatoria
6606	4	no	posgrado

	distancia	genero	puntaje_examen
0	cercano	hombre	67
1	moderado	mujer	61
2	cercano	hombre	74
3	moderado	hombre	71
4	cercano	mujer	70

...	...	...	...
6602	cercano	mujer	68
6603	cercano	mujer	69
6604	cercano	mujer	68
6605	lejano	mujer	68
6606	cercano	hombre	64

[6378 rows x 20 columns]

```
[2]: df.columns
```

```
[2]: 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')
```

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

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

```
[4]: # 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: 80.46

Valor p: 3.086484723306131e-35

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

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

nivel_de_significancia = 0.05

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

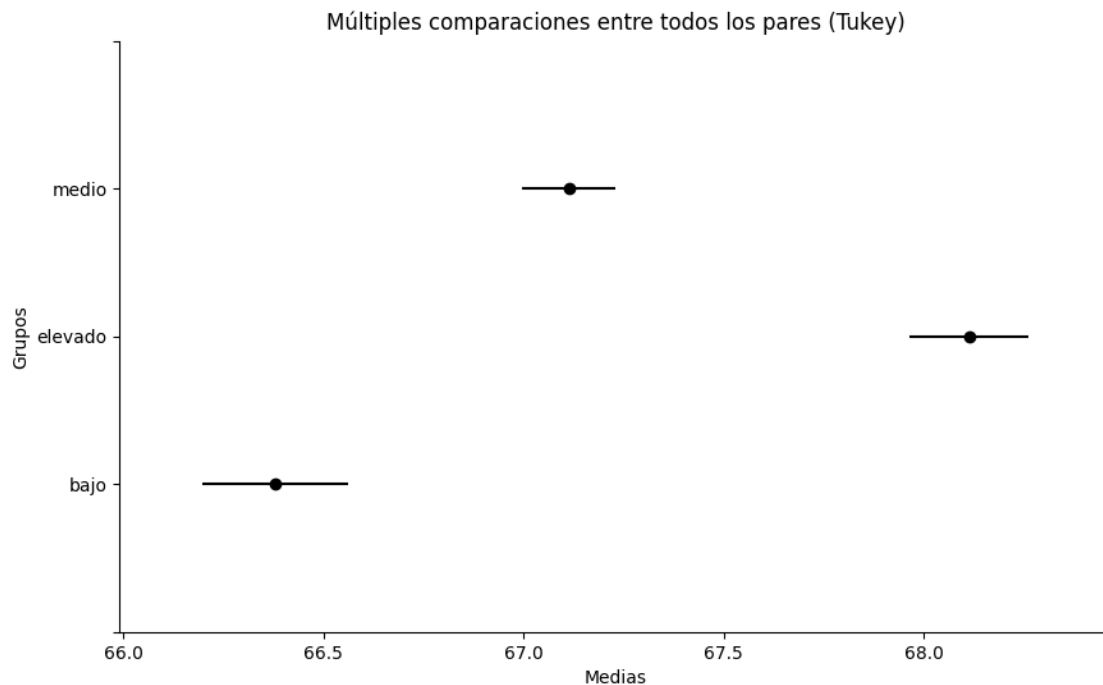
# Mostrar los resultados
print(tukey)

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

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

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

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



## CONCLUSIÓN

Con base con los resultados de la prueba ANOVA se puede observar que existen diferencias significativas entre los grupos de los datos debido a que el valor p de la prueba sobre pasa el asignado de 0.05 dandonos un resultado de 3.08 lo que nos lleva a la prueba Tukey la cual nos da como resultado que los tres pares de esta parte si cumplen con diferencias significativas.

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

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

```
[7]: # 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:.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: 77.74

Valor p: 4.409839909861347e-34

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```

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

nivel_de_significancia = 0.05

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

# Mostrar los resultados
print(tukey)

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

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

```

```

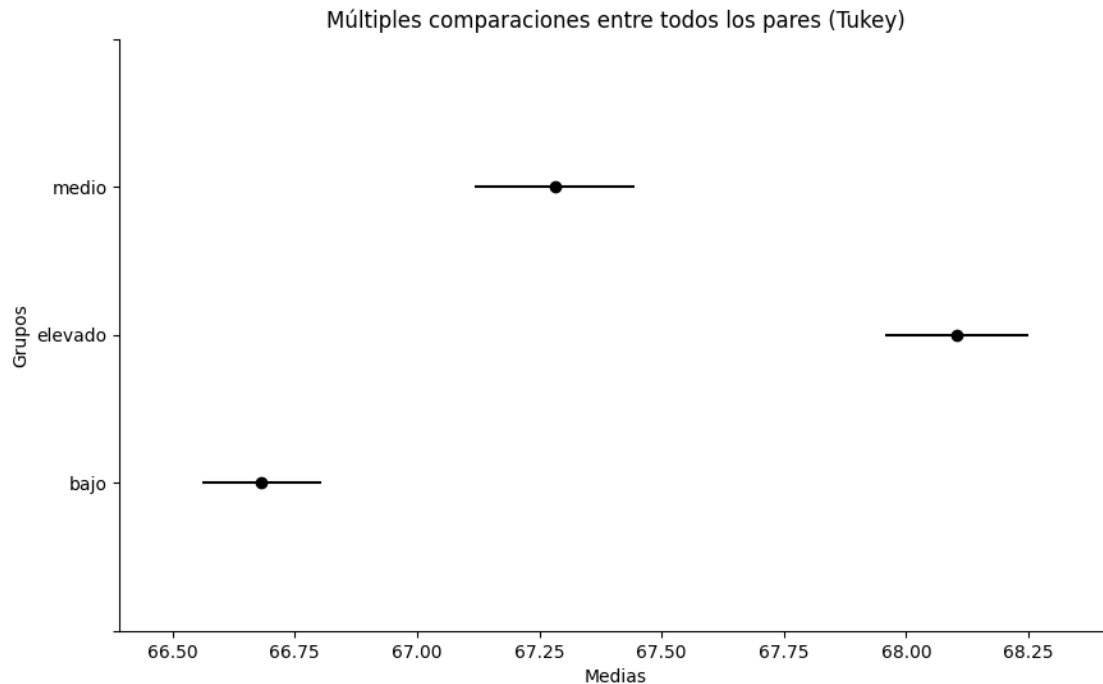
Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====
group1 group2 meandiff p-adj lower upper reject
-----
    bajo elevado   1.4212   0.0   1.1539   1.6885   True
    bajo  medio   0.6001   0.0   0.3157   0.8845   True
elevado  medio  -0.8211   0.0  -1.1309  -0.5113   True
=====

```

```

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

```



## CONCLUSIÓN

Para este estudio podemos ver en los resultados de la prueba ANOVA que el valor p nos arroja un resultado de 4.40 lo cual nos indica que existen diferencias significativas entre los datos, para la prueba tukey entre el grupo 1 y 2 de las comparaciones podemos ver que todas cumplen con diferencias significativas según los resultados de ANOVA

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

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

```
[10]: # 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:}\n")

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

```

Resultados del ANOVA:

F-Estadístico: 25.46

Valor p: 4.6465544032283915e-07

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```

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

nivel_de_significancia = 0.05

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

# Mostrar los resultados
print(tukey)

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

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

```

Multiple Comparison of Means - Tukey HSD, FWER=0.05

```

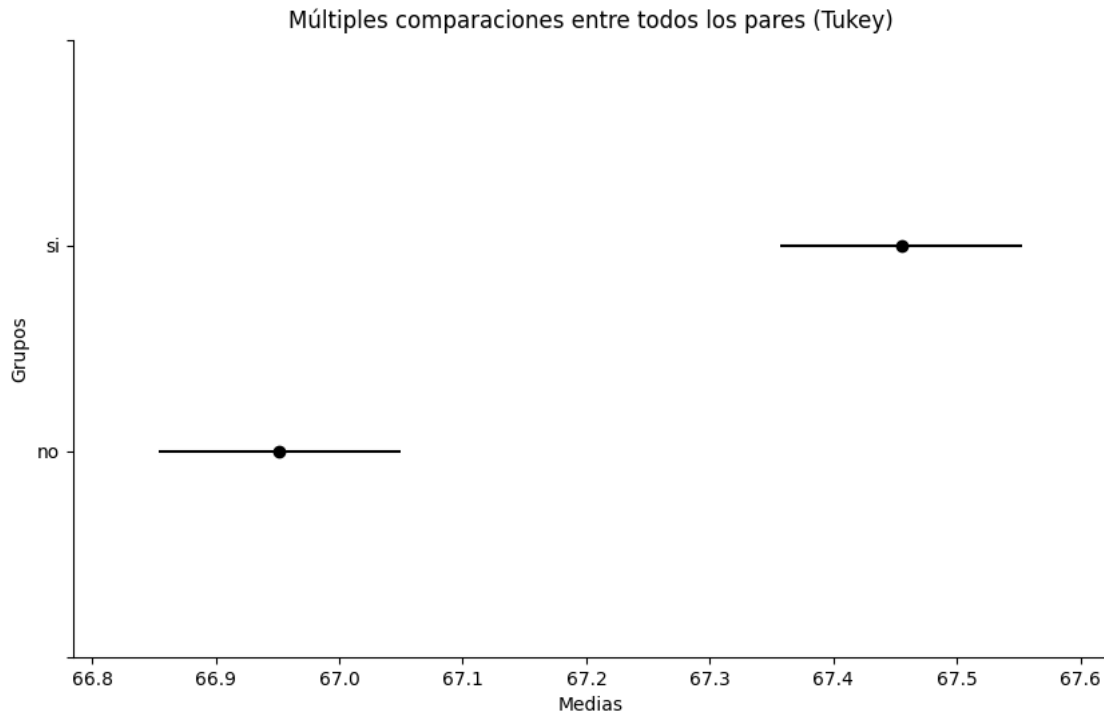
=====
group1 group2 meandiff p-adj lower upper reject
-----
    no     si   0.5032   0.0 0.3077 0.6987   True
-----

```

```

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

```



## CONCLUSIÓN

En este estudio de actividades extracurriculares tenemos como resultado en ANOVA 4.64 en su valor p debido a esto realizamos la prueba Tukey ya que la anterior prueba nos indicó que existen diferencias significativas en los grupos de datos lo que se confirma en los resultados de Tukey.

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

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

```
[13]: # 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:}\n")

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

```

Resultados del ANOVA:

F-Estadístico: 25.54

Valor p: 9.000204034394693e-12

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```

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

nivel_de_significancia = 0.05

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

# Mostrar los resultados
print(tukey)

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

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

```

```

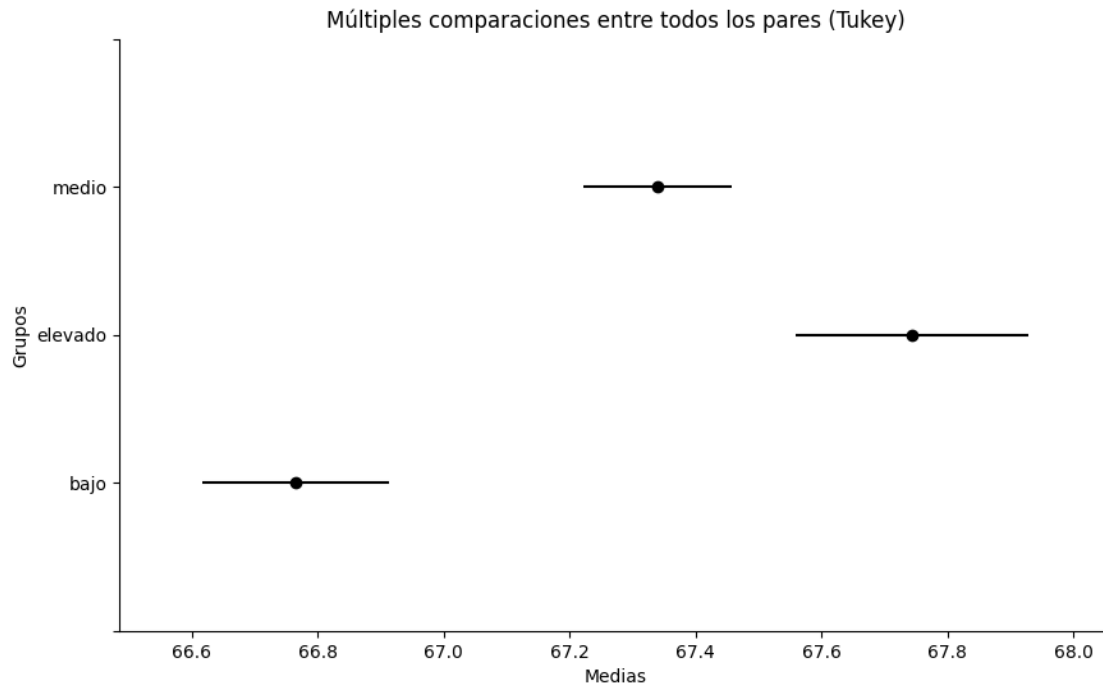
Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====
group1 group2 meandiff p-adj  lower  upper  reject
-----
    bajo elevado  0.9794    0.0  0.6474  1.3115    True
    bajo  medio  0.5744    0.0  0.3086  0.8402    True
elevado  medio -0.405  0.0048 -0.7071 -0.103    True
-----

```

```

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

```



## CONCLUSIÓN

Para el estudio de nivel motivación al realizar la prueba ANOVA nos indica que existen diferencias significativas ya que en valor p nos dio un valor de 9.00 lo cual sobre pasa el valor asignado para así mismo realizar la prueba Tukey la cual nos confirma de las diferencias entre las comparaciones de los datos

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

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

```
[16]: # 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: 16.71

Valor p: 4.4128283860948813e-05

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```

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

nivel_de_significancia = 0.05

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

# Mostrar los resultados
print(tukey)

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

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

```

Multiple Comparison of Means - Tukey HSD, FWER=0.05

```

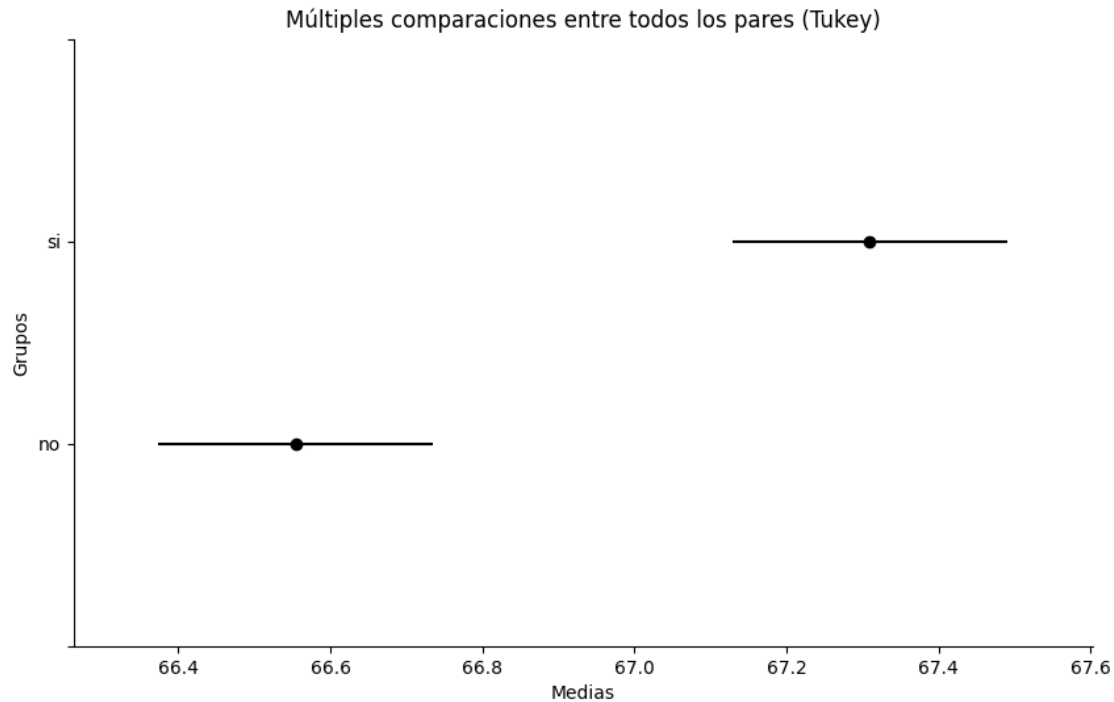
=====
group1 group2 meandiff p-adj lower upper reject
-----
    no     si   0.7549   0.0 0.3929 1.1169   True
-----

```

```

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

```



## CONCLUSIÓN

Para el estudio de acceso a internet realizamos ANOVA y como se puede observar el valor p nos da de un valor de 4.41 lo que nos indica que hay diferencias significativas y al realizar Tukey nos confirma los resultados obtenidos por lo que se recomienda a los estudiantes una mejora en ese aspecto.

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

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

```
[19]: # 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:}\n")

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

```

Resultados del ANOVA:

F-Estadístico: 28.78

Valor p: 3.59767227814745e-13

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```

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

nivel_de_significancia = 0.05

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

# Mostrar los resultados
print(tukey)

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

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

```

```

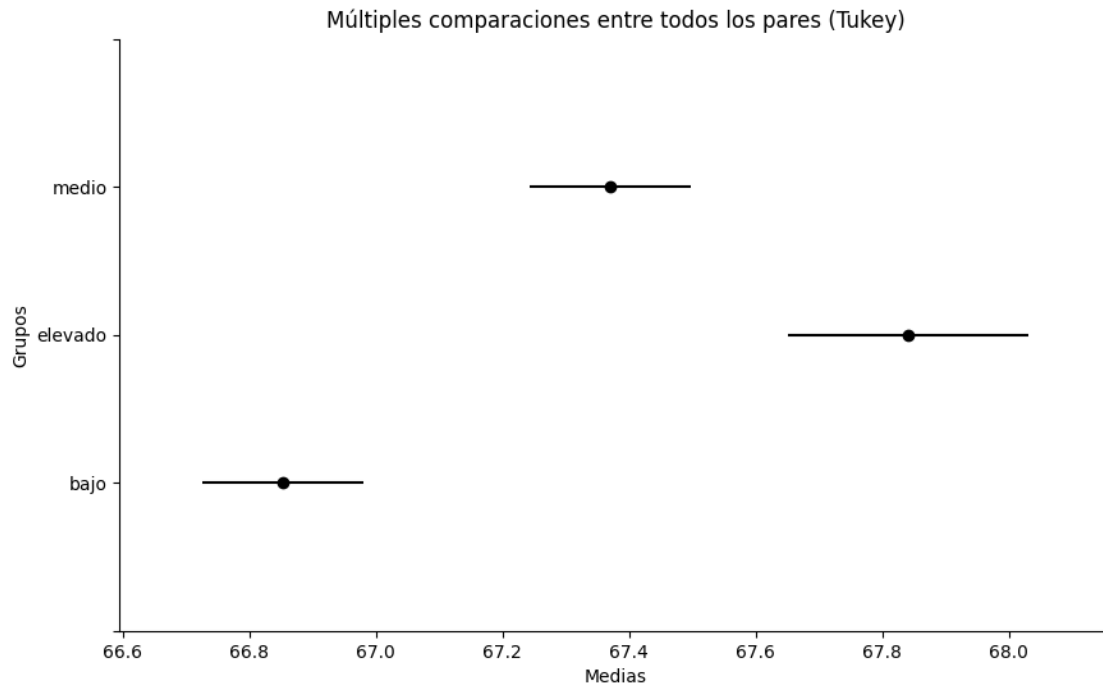
Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====
group1 group2 meandiff p-adj  lower  upper  reject
-----
    bajo elevado  0.9882    0.0  0.6717  1.3048   True
    bajo  medio  0.5178    0.0  0.2631  0.7725   True
elevado  medio -0.4705 0.0015 -0.7873 -0.1536   True
=====

```

```

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

```



## CONCLUSIÓN

Para este caso podemos ver los resultados obtenidos ya que en ANOVA nos indica que si hay diferencias significativas debido a su valor p de 3.59 lo que se confirma con la prueba Tukey debido a esto se buscaría una mejora para las familias.

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

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

```
[22]: # 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:}\n")

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

```

Resultados del ANOVA:

F-Estadístico: 17.12

Valor p: 3.844782377251745e-08

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```

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

nivel_de_significancia = 0.05

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

# Mostrar los resultados
print(tukey)

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

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

```

```

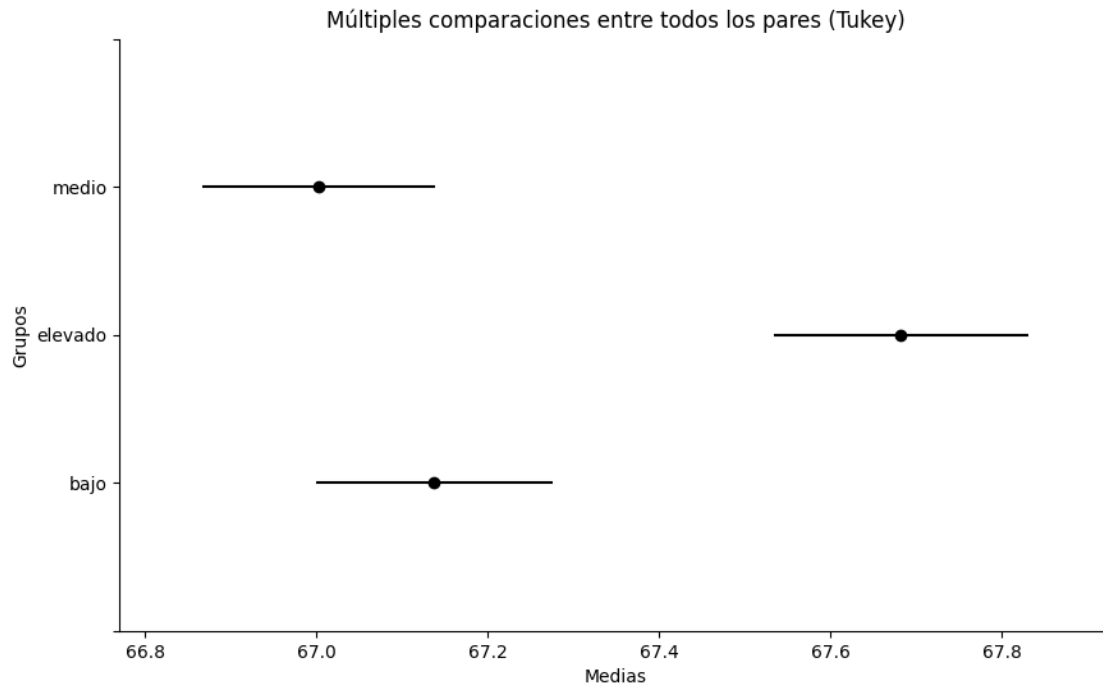
Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====
group1 group2 meandiff p-adj  lower  upper  reject
-----
    bajo elevado  0.5447    0.0  0.2585  0.8309   True
    bajo  medio -0.1345  0.4821 -0.4082  0.1392  False
elevado  medio -0.6792    0.0 -0.9638 -0.3947   True
-----

```

```

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

```



## CONCLUSIÓN

En esta parte podemos ver que el valor p de ANOVA sobre pasa el valor asignado antes de realizar la prueba por lo que si existen pruebas significativas entre los datos y con ayuda de la prueba Tukey podemos ver que es verdadero para las comparaciones de datos.

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

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

```
[25]: # 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:.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: 0.75

Valor p: 0.3854987810260675

Conclusión: No hay diferencias significativas entre los tratamientos.

```

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

nivel_de_significancia = 0.05

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

# Mostrar los resultados
print(tukey)

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

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

```

```

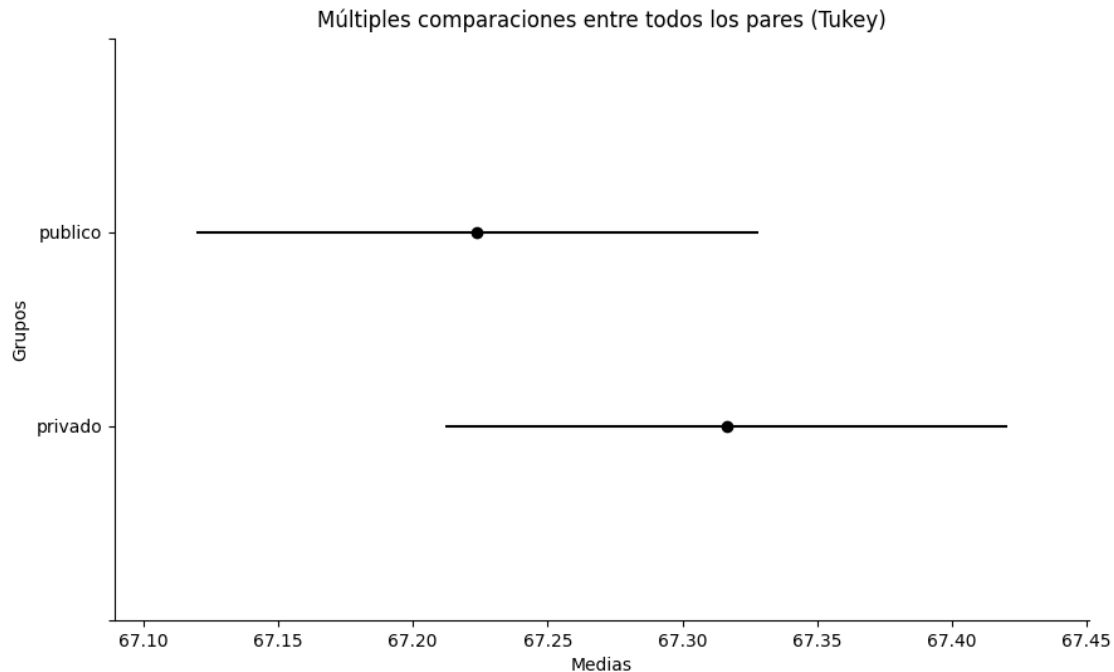
Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====
group1 group2 meandiff p-adj  lower  upper  reject
-----
privado publico -0.0924 0.3855 -0.3011 0.1163  False
-----

```

```

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

```



## CONCLUSIÓN

El tipo de escuela también es un factor como podemos ver ya que dependiendo de que tipo de escuela se encuentre la población se estudia si afecta o no en los resultados como se puede observar en ANOVA y Tukey las cuales nos indican que no existen diferencias significativas entre los datos.

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

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

```
[28]: # 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:.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: 32.18

Valor p: 1.2426130744876118e-14

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```

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

nivel_de_significancia = 0.05

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

# Mostrar los resultados
print(tukey)

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

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

```

```

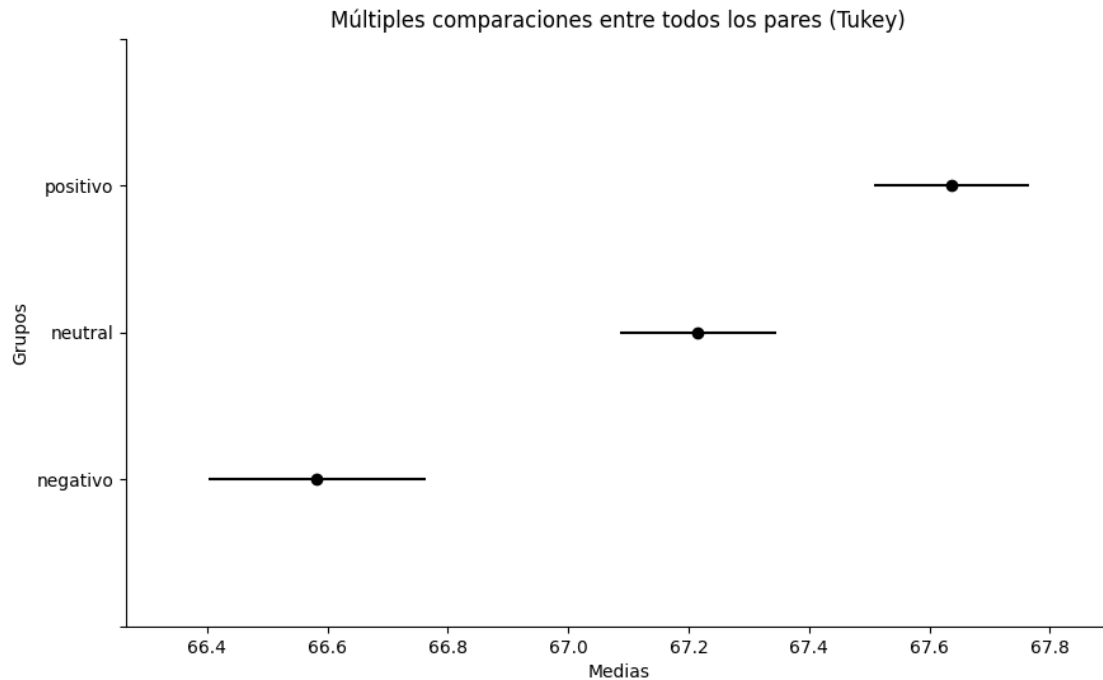
Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====
group1  group2  meandiff  p-adj  lower  upper  reject
-----
negativo  neutral    0.6329    0.0  0.3229  0.9429    True
negativo positivo  1.0538    0.0  0.745  1.3626    True
neutral positivo   0.4209  0.0004  0.1638  0.6779    True
-----

```

```

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

```



## CONCLUSIÓN

La influencia de los compañeros de igual manera se observa que es un factor que puede o no afectar en los resultados esperados y con ayuda de ANOVA podemos ver que si hay diferencias significativas en los resultados y con ayuda de Tukey confirmamos los resultados de ANOVA entre las comparaciones de datos.

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

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

```
[44]: # 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}\n")

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

```

Resultados del ANOVA:

F-Estadístico: 35.95

Valor p: 2.9904980898726883e-16

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```

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

nivel_de_significancia = 0.05

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

# Mostrar los resultados
print(tukey)

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

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

```

```

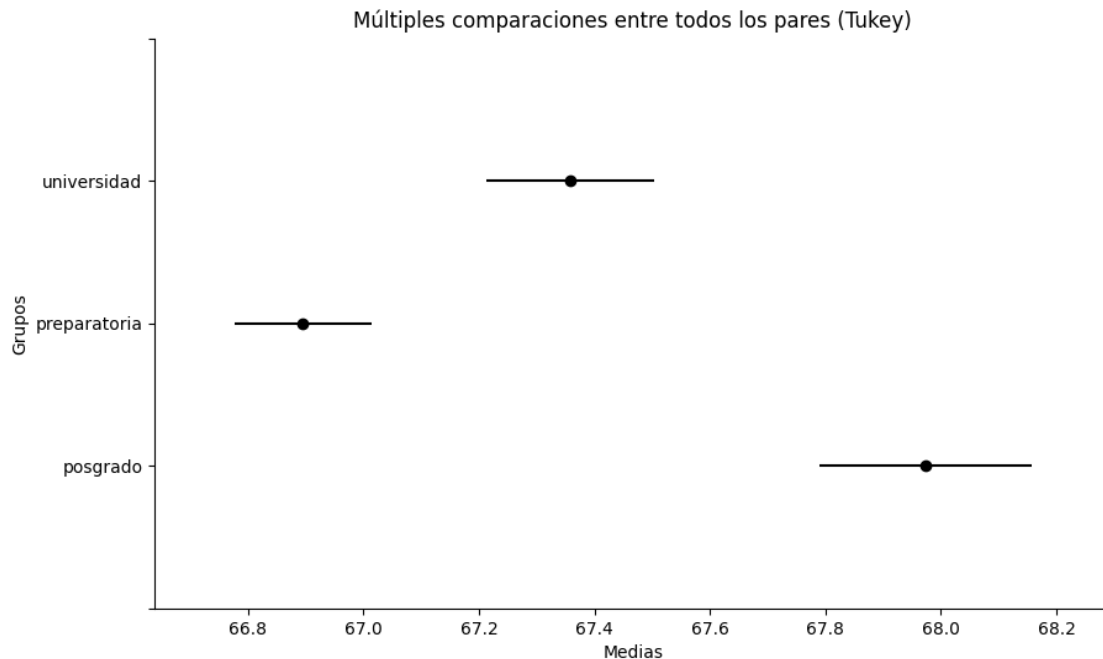
Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====
group1      group2      meandiff p-adj   lower  upper  reject
-----
posgrado preparatoria -1.0778    0.0 -1.3801 -0.7754   True
posgrado universidad  -0.6142    0.0 -0.9429 -0.2856   True
preparatoria universidad  0.4635 0.0001  0.2002  0.7268   True
=====

```

```

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

```



## CONCLUSIÓN

Nivel de estudio de los padres puede ser factor para que los resultados varíen, debido a ANOVA y Tukey podemos observar que si existen diferencias significativas en los datos ya que para el valor p nos da un valor de 2.99 que es un valor alto.

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

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

```
[50]: # 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}\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: 1.6303083634793744e-11

Conclusión: Hay diferencias significativas entre al menos dos tratamientos.

```

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

nivel_de_significancia = 0.05

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

# Mostrar los resultados
print(tukey)

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

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

```

```

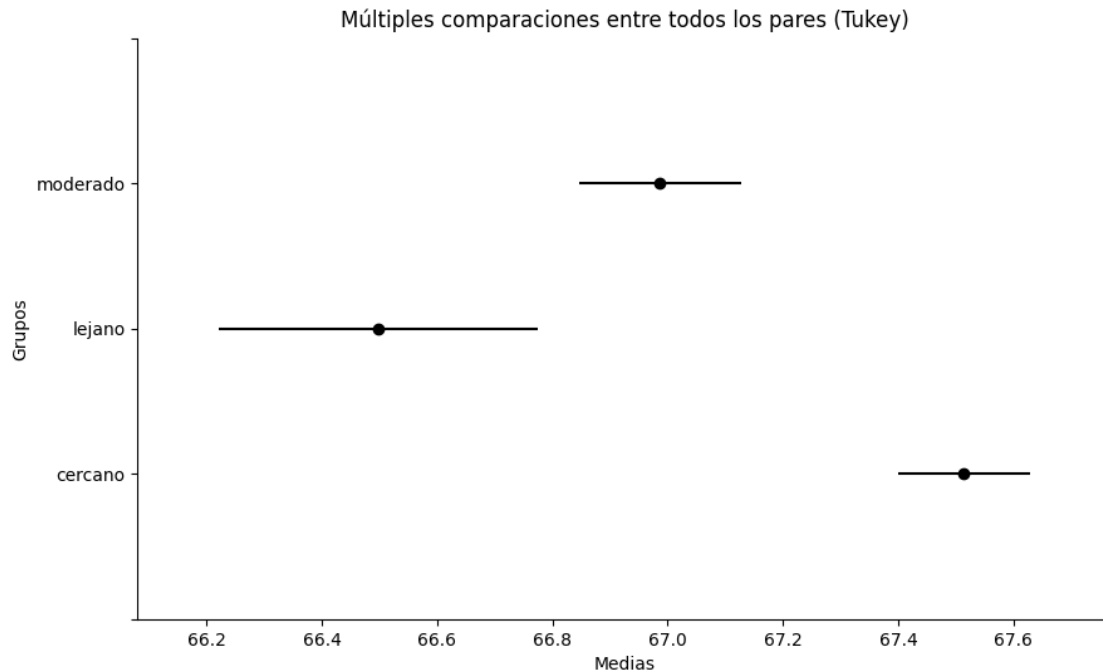
Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====
group1 group2 meandiff p-adj lower upper reject
-----
cercano lejano -1.0154 0.0 -1.407 -0.6237 True
cercano moderado -0.5272 0.0 -0.7822 -0.2722 True
lejano moderado 0.4882 0.017 0.0705 0.9059 True
-----

```

```

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

```



## CONCLUSIÓN

Para este estudio podemos observar que la distancia si es un factor para la diferencia significativa de los datos pues ANOVA nos da un valor p de 1.63 lo que podemos observar más claro en la prueba Tukey.

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

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

```
[54]: # 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: 0.16

Valor p: 0.693695501200843

Conclusión: No hay diferencias significativas entre los tratamientos.

```

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

nivel_de_significancia = 0.05

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

# Mostrar los resultados
print(tukey)

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

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

```

Multiple Comparison of Means - Tukey HSD, FWER=0.05

```

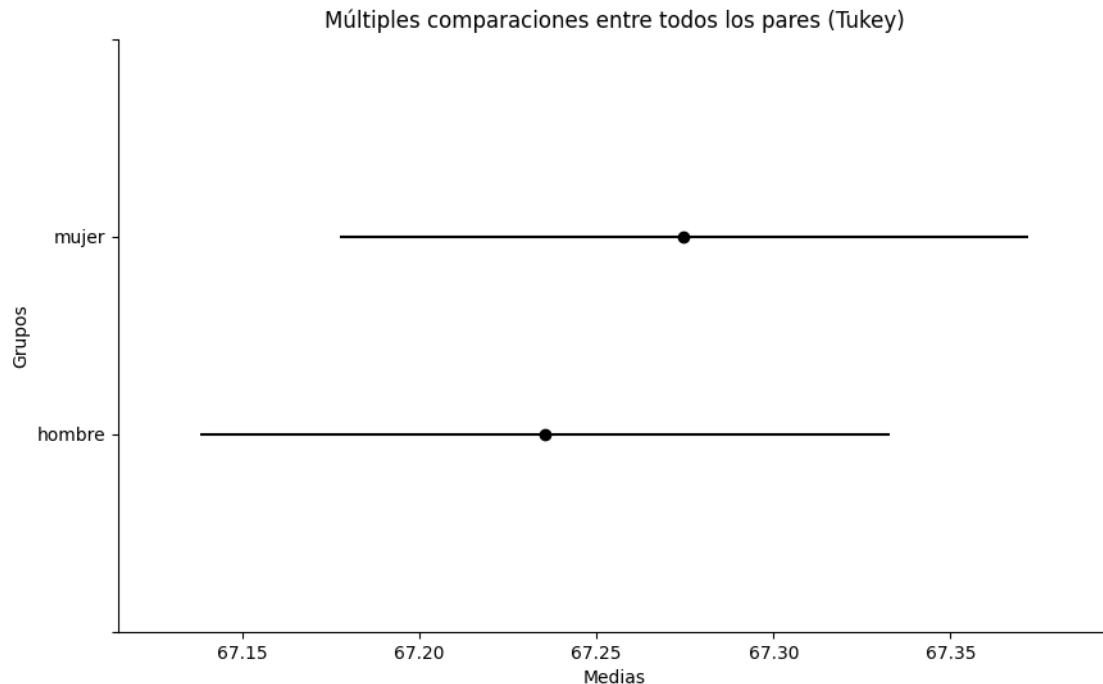
=====
group1 group2 meandiff p-adj  lower  upper  reject
-----
hombre  mujer    0.0391 0.6937 -0.1555 0.2337  False
-----

```

```

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

```



## CONCLUSIÓN

En esta parte vemos qué la prueba ANOVA nos indica que no hay una diferencia significativa entre los datos ya que su valor  $p$  es de 0.69 lo cual confirmamos con la prueba Tukey, lo que nos indica de igual manera que el género no es un factor de diferencias entre los resultados del nivel académico.

## 4 CONCLUSIÓN GENERAL

Finalmente podemos ver que la mayor parte de estos datos cualitativos nos dan diferencias significativas entre las comparaciones de sus grupos esto para que estudio de dichos datos y saber con exactitud los factores que afectan para el rendimiento académico de los estudiantes, con ayuda de las pruebas ANOVA y Tukey nos quedaron más claros la relación de los datos y cómo estas pruebas estudian cada grupo de datos para llegar a los resultados deseados que se necesitan para dicho estudio, para concluir podemos decir que ambas pruebas si son de gran utilidad para poder hacer diferenciaciones para cualquier base de datos que se desee estudiar pues ambas son herramientas que se complementan y visto así podemos concluir que la mayoría de datos cumplen con diferencias significativas las cuales todos sus grupos dentro de los datos son responsables de dichas diferencias lo que nos permite llegar a conclusiones específicas de las relaciones entre los grupos.