

ei-u4-01-22150541

November 28, 2024

```
[ ]: # Pues este pequeño código suprime los warnings:
```

```
import warnings
# Suprimir todos los warnings
warnings.filterwarnings('ignore')
```

```
[ ]: import pandas as pd
df = pd.read_csv('https://raw.githubusercontent.com/diana462/Unidad-4/refs/heads/main/Student_Performance_new.csv')
df.drop(columns='Unnamed: 0', inplace=True)
df.columns
```

```
[ ]: Index(['etnia', 'nivel_educativo_de_los_padres', 'tipo_de_comida',
        'curso_de_preparación_para_el_examen', 'porcentaje_de_matematicas',
        'porcentaje_de_lectura', 'porcentaje_de_escritura', 'sexo'],
        dtype='object')
```

```
[ ]: df["promedio"] = (df['porcentaje_de_matematicas'] +
        df['porcentaje_de_escritura'] + df['porcentaje_de_lectura']) / 3
df
```

```
[ ]:      etnia nivel_educativo_de_los_padres tipo_de_comida \
0    group_B          licenciatura      estandar
1    group_C      algo_de_universidad      estandar
2    group_B          maestria      estandar
3    group_A      tecnico_asociado  subsidiado
4    group_C      algo_de_universidad      estandar
..     ...
995  group_E          maestria      estandar
996  group_C      preparatoria  subsidiado
997  group_C      preparatoria  subsidiado
998  group_D      algo_de_universidad      estandar
999  group_D      algo_de_universidad  subsidiado

      curso_de_preparación_para_el_examen  porcentaje_de_matematicas \
0                      nada                                0.72
1                  completado                                0.69
```

2	nada	0.90
3	nada	0.47
4	nada	0.76
..
995	completado	0.88
996	nada	0.62
997	completado	0.59
998	completado	0.68
999	nada	0.77

	porcentaje_de_lectura	porcentaje_de_escritura	sexo	promedio
0	0.72	0.74	mujer	0.726667
1	0.90	0.88	mujer	0.823333
2	0.95	0.93	mujer	0.926667
3	0.57	0.44	hombre	0.493333
4	0.78	0.75	hombre	0.763333
..
995	0.99	0.95	mujer	0.940000
996	0.55	0.55	hombre	0.573333
997	0.71	0.65	mujer	0.650000
998	0.78	0.77	mujer	0.743333
999	0.86	0.86	mujer	0.830000

[1000 rows x 9 columns]

Nivel educativo de los padres

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

```
[ ]: array(['licenciatura', 'algo_de_universidad', 'maestria',
          'tecnico_asociado', 'preparatoria', 'algo_de_preparatoria'],
          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_educativo_de_los_padres'] == 'licenciatura']['promedio'],
    df[df['nivel_educativo_de_los_padres'] == '
    ↪'algo_de_universidad']['promedio'],
    df[df['nivel_educativo_de_los_padres'] == 'maestria']['promedio'],
    df[df['nivel_educativo_de_los_padres'] == 'tecnico_asociado']['promedio'],
    df[df['nivel_educativo_de_los_padres'] == 'preparatoria']['promedio'],
```

```

    df[df['nivel_educativo_de_los_padres'] == '
    ↪ algo_de_preparatoria']['promedio'],
)

# 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: 10.75

Valor p: 0.00000

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
tukey = pairwise_tukeyhsd(endog=df['promedio'],
    ↪ groups=df['nivel_educativo_de_los_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)")

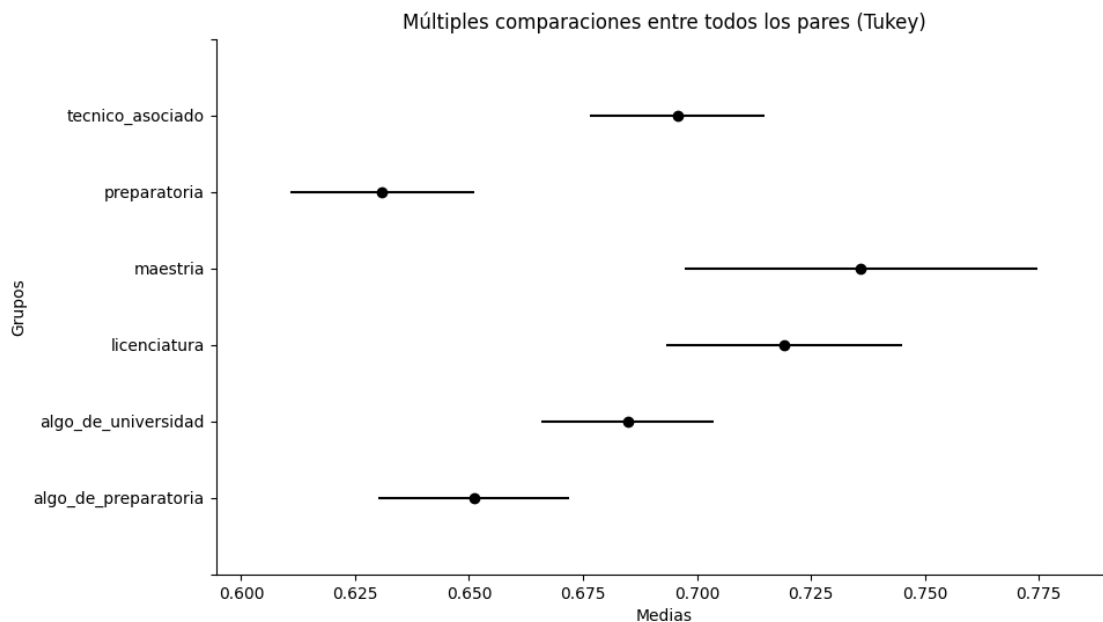
```

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
algo_de_preparatoria	algo_de_universidad	0.0337	0.1509	-0.0061	0.0735	False
algo_de_preparatoria	licenciatura	0.0682	0.0006	0.021	0.1153	True

algo_de_preparatoria	maestria	0.0849	0.0007	0.0252	0.1446	True
algo_de_preparatoria	preparatoria	-0.0201	0.7288	-0.0612	0.021	False
algo_de_preparatoria	tecnico_asociado	0.0446	0.0183	0.0047	0.0845	True
algo_de_universidad	licenciatura	0.0345	0.2479	-0.0107	0.0796	False
algo_de_universidad	maestria	0.0512	0.1201	-0.0069	0.1093	False
algo_de_universidad	preparatoria	-0.0538	0.0011	-0.0926	-0.015	True
algo_de_universidad	tecnico_asociado	0.0109	0.9618	-0.0266	0.0485	False
licenciatura	maestria	0.0168	0.9748	-0.0466	0.0801	False
licenciatura	preparatoria	-0.0883	0.0	-0.1346	-0.042	True
licenciatura	tecnico_asociado	-0.0235	0.6743	-0.0688	0.0217	False
maestria	preparatoria	-0.105	0.0	-0.164	-0.046	True
maestria	tecnico_asociado	-0.0403	0.3567	-0.0985	0.0179	False
preparatoria	tecnico_asociado	0.0647	0.0	0.0258	0.1037	True

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



Interpretación: En base a la grafica podemos observar, que muchas de las líneas no se superponen, lo que sugiere que hay diferencias significativas entre los promedios de varios grupos. Sin embargo, los resultados sugieren que el nivel educativo está relacionado con el promedio de la variable nivel educativo de los padres.

Etnia

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

```
[ ]: array(['group_B', 'group_C', 'group_A', 'group_D', 'group_E'],
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['etnia'] == 'group_B']['promedio'],
    df[df['etnia'] == 'group_C']['promedio'],
    df[df['etnia'] == 'group_A']['promedio'],
    df[df['etnia'] == 'group_D']['promedio'],
    df[df['etnia'] == 'group_E']['promedio'],
)

# 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: 9.10

Valor p: 0.00000

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
tukey = pairwise_tukeyhsd(endog=df['promedio'], groups=df['etnia'],
    ↪alpha=nivel_de_significancia)

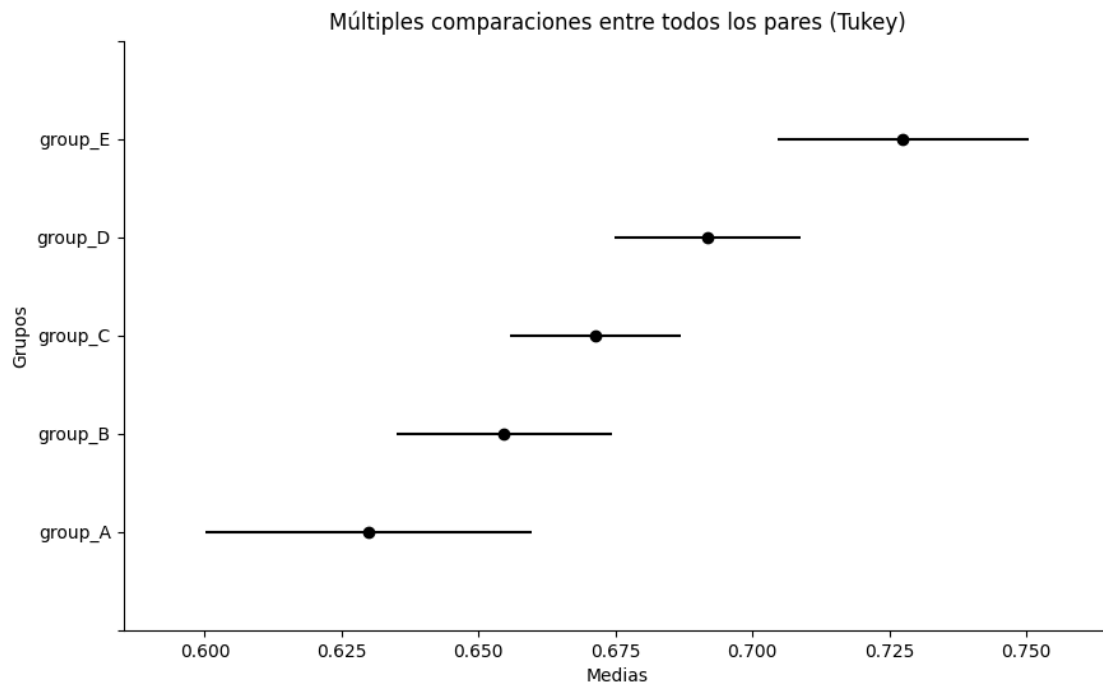
# Mostrar los resultados
print(tukey)

# 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
-----
group_A group_B 0.0248 0.6447 -0.0245 0.074 False
group_A group_C 0.0414 0.1006 -0.0046 0.0874 False
group_A group_D 0.0619 0.0031 0.0148 0.1089 True
group_A group_E 0.0976 0.0 0.0456 0.1496 True
group_B group_C 0.0166 0.6953 -0.0185 0.0518 False
group_B group_D 0.0371 0.0445 0.0006 0.0736 True
group_B group_E 0.0728 0.0 0.0301 0.1155 True
group_C group_D 0.0205 0.4036 -0.0115 0.0524 False
group_C group_E 0.0562 0.0008 0.0173 0.0951 True
group_D group_E 0.0357 0.1076 -0.0044 0.0759 False
-----
```

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



Interpretación: Con base en la grafica podemos observar, que las líneas no se superponen, lo que sugiere que hay diferencias significativas en los promedios de la variable entre varias etnias.

Curso de preparación para el examen

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

```
[ ]: array(['nada', 'completado'], 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['curso_de_preparación_para_el_examen'] == 'nada']['promedio'],
    df[df['curso_de_preparación_para_el_examen'] == 'completado']['promedio'],
)

# 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: 70.41

Valor p: 0.00000

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
tukey = pairwise_tukeyhsd(endog=df['promedio'],
    ↪groups=df['curso_de_preparación_para_el_examen'],
    ↪alpha=nivel_de_significancia)

# Mostrar los resultados
```

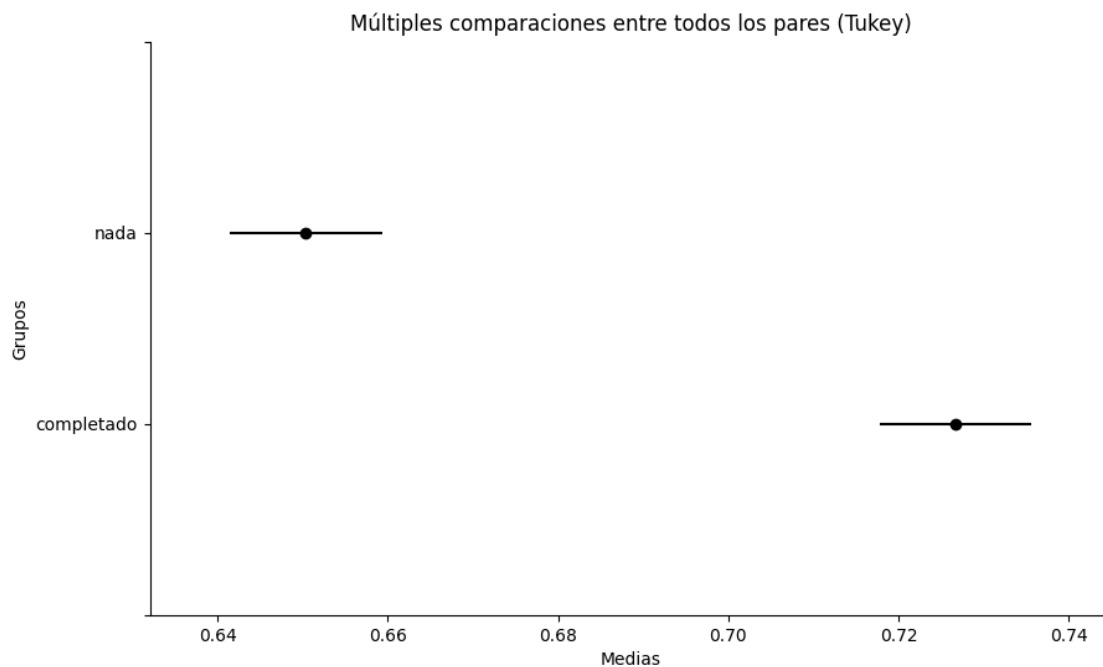
```
print(tukey)

# 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
-----
completado  nada -0.0763  0.0 -0.0942 -0.0585  True
-----
```

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



Interpretación: Existe una diferencia significativa en las medias entre los grupos, en base a los resultados podemos inferir que el curso preparatorio tiene un efecto significativo en el rendimiento de los estudiantes en el examen. Los estudiantes que completaron el curso obtuvieron, en promedio, mejores resultados que aquellos que no lo hicieron, es decir el curso preparatorio parece ser una herramienta efectiva para mejorar el desempeño en el examen.

Sexo


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

```
[ ]: array(['mujer', 'hombre'], 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['sexo'] == 'mujer']['promedio'],
    df[df['sexo'] == 'hombre']['promedio'],
)

# 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: 17.39

Valor p: 0.00003

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
tukey = pairwise_tukeyhsd(endog=df['promedio'], groups=df['sexo'],_
    ↪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

```

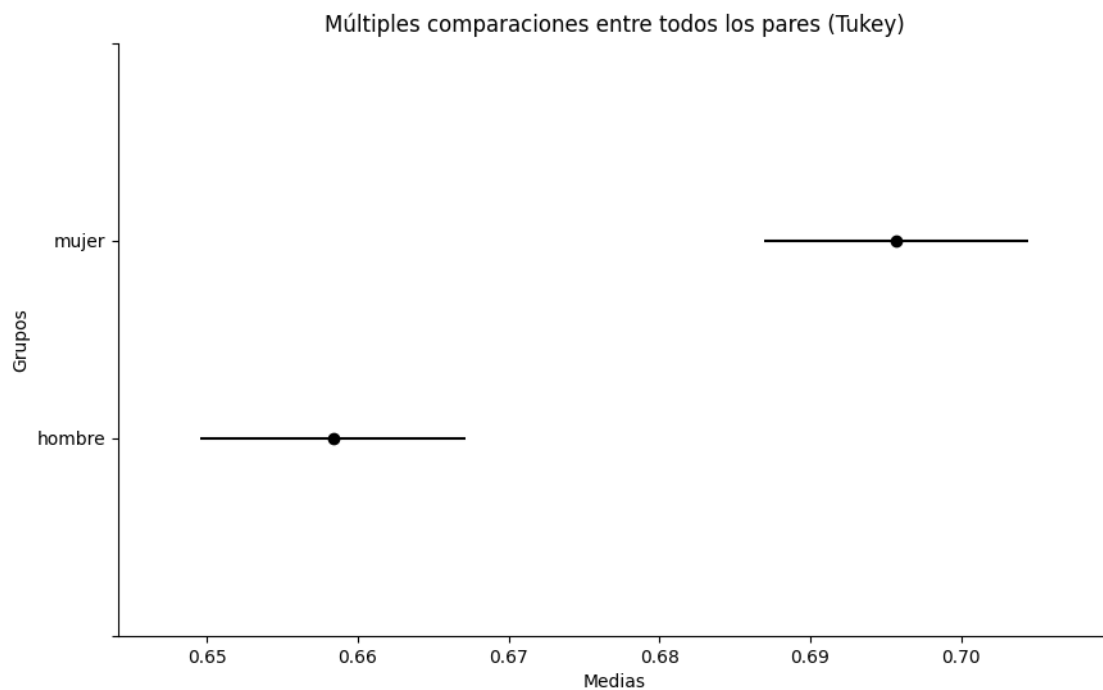
=====
group1 group2 meandiff p-adj lower  upper  reject
-----
hombre  mujer    0.0373   0.0 0.0198 0.0549    True
-----

```

```

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

```



Interpretación: Existe una diferencia significativa en las medias entre hombres y mujeres, es decir que en promedio, esto significa que, en promedio, uno de los géneros obtuvo un valor más alto en esta variable.

Tipo de comida

```

[ ]: df['tipo_de_comida'].unique()

```

```

[ ]: array(['estandar', 'subsidiado'], 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_de_comida'] == 'estandar']['promedio'],
    df[df['tipo_de_comida'] == 'subsidiado']['promedio'],
)

# 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: 91.68

Valor p: 0.00000

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
tukey = pairwise_tukeyhsd(endog=df['promedio'], groups=df['tipo_de_comida'],
    ↪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
```

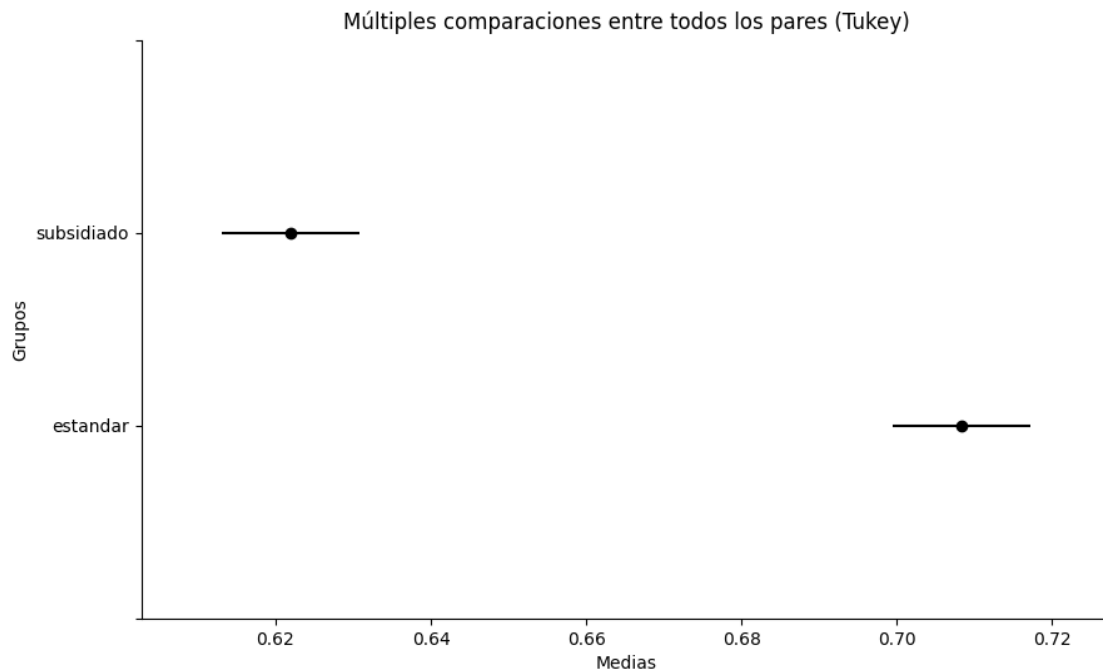
```
plt.title("Múltiples comparaciones entre todos los pares (Tukey)")
```

```

Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====
group1    group2    meandiff p-adj    lower    upper    reject
-----
estandar  subsidiado -0.0864    0.0 -0.1041 -0.0687    True
=====

```

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



Interpretacion: Podemos inferir que existe una diferencia significativa en las medias entre la comida subsidiada y la estándar. Es decir que en promedio, uno de los tipos de comida obtuvo un valor más alto en esta variable.

Conclusión:

En base a los resultados obtenidos nos podemos percatar que en todos los hallazgos hay diferencia significativa, dado que permite comparar las medias de varios para determinar si existen diferencias significativas entre ellos.