In [117…
```
import pandas as pd
from scipy.stats import pearsonr, spearmanr, ttest_ind, f_oneway
```

In [118…
```
df = pd.read_csv('./data/Student_performance_data.csv')
df
```

Out[118…

	StudentID	Age	Gender	Ethnicity	ParentalEducation	StudyTimeWeekly	Absences	Tutoring	ParentalSupport	Extracurricula
0	1001	17	1	0	2	19.833723	7	1	2	
1	1002	18	0	0	1	15.408756	0	0	1	
2	1003	15	0	2	3	4.210570	26	0	2	
3	1004	17	1	0	3	10.028829	14	0	3	
4	1005	17	1	0	2	4.672495	17	1	3	
...	...	...	...	...	...	...	...	...	...	
2387	3388	18	1	0	3	10.680555	2	0	4	
2388	3389	17	0	0	1	7.583217	4	1	4	
2389	3390	16	1	0	2	6.805500	20	0	2	
2390	3391	16	1	1	0	12.416653	17	0	2	
2391	3392	16	1	0	2	17.819907	13	0	2	

2392 rows × 15 columns

Uso de 'GPA' como Variable Objetivo

Razón para elegir 'GPA' sobre 'GradeClass':

- Al usar el GPA, obtenemos una medida continua que permite una mayor precisión en los análisis estadísticos. Las diferencias en el rendimiento académico se pueden capturar con mayor exactitud que con una variable categórica.
- Muchas de las pruebas estadísticas empleadas (como las correlaciones de Pearson y Spearman, y la prueba t) funcionan mejor con datos continuos y proporcionan resultados más significativos.

Justificación de las pruebas estadísticas utilizadas

1. Correlación de Pearson: La correlación de Pearson mide la relación lineal entre dos variables continuas. Esta prueba permite determinar si existe una relación directa y significativa entre el tiempo de estudio y las notas.

- Ha sido usada para las siguentes variables: **'StudyTimeWeekly', 'ParentalSupport'**

2. Correlación de Spearman: La correlación de Spearman mide la relación monotónica entre dos variables, que no necesariamente tienen que se lineales. Es útil cuando la relación entre las variables no es estrictamente lineal.

- Ha sido usada en las siguientes variables: **'Absences'**

3. Prueba t de muestras independientes: La prueba t de muestras independientes compara las medias de dos grupos para ver si hay una diferencia significativa entre ellos. Es adecuada para comparar las variables categóricas binarias con una variable continua.

- Ha sido usada en las siguientes variables: **'Gender', 'Tutoring', 'Extracurricular', 'Sports', 'Music', 'Volunteering'**

4. ANOVA (Análisis de varianza) de un factor: ANOVA es apropiada para comparar las medias de tres o más grupos para determinar si al menos una de las medias es significativamente diferente de las demás. Es adecuada para comparar variables categóricas politómicas con una variable continua.

- Ha sido usada en las siguientes variables: **'Ethnicity', 'ParentalEducation'**

En resumen:

- **Pearson:** Relaciones lineales entre variables continuas.
- **Spearman:** Relaciones monotónicas sin necesidad que sean lineales.
- **t:** variable categórica binómica y variable continua.
- **ANOVA:** variable categórica politómica y variable continua.

In [ ]:
```
# Variables a correlacionar con 'GPA'
variables = ['Age', 'StudyTimeWeekly', 'Absences', 'Gender', 'Ethnicity', 'ParentalEducation', 'Tutoring', 'Par
```

```
# Calcular la correlación de Spearman
spearman_results = {}
for var in variables:
  correlation, p_value = spearmanr(df[var], df['GPA'])
  spearman_results[var] = {'Spearman Correlation': correlation, 'p_value': p_value}

# Mostrar resultados
spearman_results_df = pd.DataFrame(spearman_results).transpose().sort_values(by='Spearman Correlation')
spearman_results_df
```

Planteamiento de hipotesis

1. En relación a la variable StudyTimeWeekly

- **H0**: No hay una relación de proporcionalidad directa entre el tiempo semanal de estudio y las notas de los alumnos (cuanto mas estudio, mejores notas)
- **H1**: Hay una relación de proporcionalidad directa entre el tiempo semanal de estudio y las notas de los alumnos (cuanto mas estudio, mejores notas)

In [119…
```
# Prueba de correlación de Pearson para StudyTimeWeekly

corr, p_value = pearsonr(df['StudyTimeWeekly'], df['GPA'])

print(f'Correlación de Pearson: {corr:.3f}, p_value: {p_value:.3f}')

if p_value < 0.05:
    print("H1 <> Hay una relación significativa entre el tiempo semanal de estudio y las notas.")
else:
    print("H0 <> No hay una relación significativa entre el tiempo semanal de estudio y las notas.")
```

```
Correlación de Pearson: 0.179, p_value: 0.000
H1 <> Hay una relación significativa entre el tiempo semanal de estudio y las notas.
```

2. En relación a la variable Absences

- **H0**: No hay una relación de proporcionalidad inversa entre el numero de ausencias las notas de los alumnos (cuanto mas absencias, peores notas)
- **H1**: Hay una relación de proporcionalidad inversa entre el numero de ausencias las notas de los alumnos (cuanto mas absencias, peores notas)

In [120…
```
# Prueba de correlación de Spearman para Absences

corr, p_value = spearmanr(df['Absences'], df['GPA'])

print(f'Correlación de Spearman: {corr:.3f}, p_value: {p_value:.3f}')
if p_value < 0.05:
    print("H1 <> Hay una relación significativa entre el número de ausencias y las notas.")
else:
    print("H0 <> No hay una relación significativa entre el número de ausencias y las notas.")
```

```
Correlación de Spearman: -0.925, p_value: 0.000
H1 <> Hay una relación significativa entre el número de ausencias y las notas.
```

3. En relación a la variable Gender

- **H0**: No hay diferencias significativas en las notas de hombres y mujeres
- **H1**: Hay diferencias significativas en las notas de hombres y mujeres

In [121…
```
# Prueba t de muestras independientes para Gender

# Dividir en grupos por género
df_male = df[df['Gender'] == 0]['GPA']
df_female = df[df['Gender'] == 1]['GPA']

# Prueba t de muestras independientes
t_stat, p_value = ttest_ind(df_male, df_female, equal_var=False)

print(f't_stat: {t_stat:.3f}, p_value: {p_value:.3f}')
if p_value < 0.05:
    print("H1 <> Hay diferencias significativas en las notas entre hombres y mujeres.")
else:
    print("H0 <> No hay diferencias significativas en las notas entre hombres y mujeres.")
```

```
t_stat: 0.653, p_value: 0.514
H0 <> No hay diferencias significativas en las notas entre hombres y mujeres.
```

4. En relación a la variable Ethnicity

- **H0**: No hay diferencias significativas en las notas alumnos de diferentes etnias
- **H1**: Hay diferencias significativas en las notas alumnos de diferentes etnias

In [122…
```
# Prueba ANOVA de una vía para Ethnicity

# Dividir en grupos por etnicidad
ethnic_groups = df['Ethnicity'].unique()
groups = [df[df['Ethnicity'] == group]['GPA'] for group in ethnic_groups]

# ANOVA de una vía
f_stat, p_value = f_oneway(*groups)

print(f'f_stat: {f_stat:.3f}, p_value: {p_value:.3f}')
if p_value < 0.05:
    print("H1 <> Hay diferencias significativas en las notas entre diferentes etnias.")
else:
    print("H0 <> No hay diferencias significativas en las notas entre diferentes etnias.")
```

```
f_stat: 0.958, p_value: 0.412
H0 <> No hay diferencias significativas en las notas entre diferentes etnias.
```

5. En relación a la variable ParentalEducation

- **H0**: No hay diferencias significativas en las notas de los alumnos según el nuvel de educación parental
- **H1**: Hay diferencias significativas en las notas de los alumnos según el nuvel de educación parental

In [123…
```
# Prueba ANOVA de una vía para ParentalEducation

# Dividir en grupos por educación parental
education_levels = df['ParentalEducation'].unique()
groups = [df[df['ParentalEducation'] == level]['GPA'] for level in education_levels]

# ANOVA de una vía
f_stat, p_value = f_oneway(*groups)
print(f'f_stat: {f_stat:.3f}, p_value: {p_value:.3f}')

if p_value < 0.05:
    print("H1 <> Existen diferencias significativas en las notas según el nivel de educación parental.")
else:
    print("H0 <> No hay evidencia suficiente de diferencias significativas en las notas según el nivel de educa
```

```
f_stat: 1.808, p_value: 0.124
H0 <> No hay evidencia suficiente de diferencias significativas en las notas según el nivel de educación parenta
l.
```

6. En relación a la variable Tutoring

- **H0**: No hay diferencias significativas en las notas de los alumnos que tienen tutor y los que no
- **H1**: Hay diferencias significativas en las notas de los alumnos que tienen tutor y los que no

In [124…
```
# Prueba t de muestras independientes para Tutoring

# Dividir en grupos por tutoría
df_tutored = df[df['Tutoring'] == 1]['GPA']
df_no_tutored = df[df['Tutoring'] == 0]['GPA']

# Prueba t de muestras independientes
t_stat, p_value = ttest_ind(df_tutored, df_no_tutored, equal_var=False)

print(f't_stat: {t_stat:.3f}, p_value: {p_value:.3f}')
if p_value < 0.05:
    print("H1 <> Los alumnos con tutor tienen significativamente mejores notas.")
else:
    print("H0 <> No hay diferencias significativas en las notas entre los alumnos con tutoría y los que no.")
```

```
t_stat: 7.172, p_value: 0.000
H1 <> Los alumnos con tutor tienen significativamente mejores notas.
```

7. En relación a la variable ParentalSupport

- **H0**: No hay una relación significativa entre el soporte parental y las notas
- **H1**: Hay una relación significativa entre el soporte parental y las notas

In [125…
```
# Prueba de correlación de Pearson para ParentalSupport

corr, p_value = pearsonr(df['ParentalSupport'], df['GPA'])
print(f'Correlación de Pearson: {corr:.3f}, p_value: {p_value:.3f}')

if p_value < 0.05:
```

```
    print("H1 <> Existe una relación significativa entre el soporte parental y las notas.")
else:
    print("H0 <> No hay evidencia suficiente de una relación significativa entre el soporte parental y las notas
```

```
Correlación de Pearson: 0.191, p_value: 0.000
H1 <> Existe una relación significativa entre el soporte parental y las notas.
```

8. En relación a la variable Extracurricular

- **H0**: No hay diferencias significativas en las notas de los alumnos que hacen actividades extracurriculares y los que no
- **H1**: Hay diferencias significativas en las notas de los alumnos que hacen actividades extracurriculares y los que no

In [126…
```
# Prueba t de muestras independientes para Extracurricular

extracurricular_grades = df[df['Extracurricular'] == 1]['GPA']
no_extracurricular_grades = df[df['Extracurricular'] == 0]['GPA']

t_stat, p_value = ttest_ind(extracurricular_grades, no_extracurricular_grades, equal_var=False)
print(f't_stat: {t_stat:.3f}, p_value: {p_value:.3f}')

if p_value < 0.05:
    print("H1 <> Existen diferencias significativas en las notas entre los alumnos que hacen actividades extracu
else:
    print("H0 <> No hay evidencia suficiente de diferencias significativas en las notas entre los alumnos que ha
```

```
t_stat: 4.609, p_value: 0.000
H1 <> Existen diferencias significativas en las notas entre los alumnos que hacen actividades extracurriculares
y los que no.
```

9. En relación a la variable Sports

- **H0**: No hay diferencias significativas en las notas de los alumnos que practican deporte y los que no
- **H1**: Hay diferencias significativas en las notas de los alumnos que practican deporte y los que no

In [127…
```
# Prueba t de muestras independientes para Sports

sports_grades = df[df['Sports'] == 1]['GPA']
no_sports_grades = df[df['Sports'] == 0]['GPA']

t_stat, p_value = ttest_ind(sports_grades, no_sports_grades, equal_var=False)
print(f't_stat: {t_stat:.3f}, p_value: {p_value:.3f}')

if p_value < 0.05:
    print("H1 <> Existen diferencias significativas en las notas entre los alumnos que practican deporte y los 
else:
    print("H0 <> No hay evidencia suficiente de diferencias significativas en las notas entre los alumnos que p
```

```
t_stat: 2.851, p_value: 0.004
H1 <> Existen diferencias significativas en las notas entre los alumnos que practican deporte y los que no.
```

10. En relación a la variable Music

- **H0**: No hay diferencias significativas en las notas de los alumnos que practican musica y los que no
- **H1**: Hay diferencias significativas en las notas de los alumnos que practican musica y los que no

In [128…
```
# Prueba t de muestras independientes para Music

music_grades = df[df['Music'] == 1]['GPA']
no_music_grades = df[df['Music'] == 0]['GPA']

t_stat, p_value = ttest_ind(music_grades, no_music_grades, equal_var=False)
print(f't_stat: {t_stat:.3f}, p_value: {p_value:.3f}')

if p_value < 0.05:
    print("H1 <> Existen diferencias significativas en las notas entre los alumnos que practican música y los q
else:
    print("H0 <> No hay evidencia suficiente de diferencias significativas en las notas entre los alumnos que p
```

```
t_stat: 3.597, p_value: 0.000
H1 <> Existen diferencias significativas en las notas entre los alumnos que practican música y los que no.
```

11. En relación a la variable Volunteering

- **H0**: No hay diferencias significativas en las notas de los alumnos que hacen voluntariado y los que no
- **H1**: Hay diferencias significativas en las notas de los alumnos que hacen voluntariado y los que no

In [129…
```
# Prueba t de muestras independientes para Volunteering

volunteering_grades = df[df['Volunteering'] == 1]['GPA']
```

```
no_volunteering_grades = df[df['Volunteering'] == 0]['GPA']

t_stat, p_value = ttest_ind(volunteering_grades, no_volunteering_grades, equal_var=False)
print(f't_stat: {t_stat:.3f}, p_value: {p_value:.3f}')

if p_value < 0.05:
    print("H1 <> Existen diferencias significativas en las notas entre los alumnos que hacen voluntariado y los
else:
    print("H0 <> No hay evidencia suficiente de diferencias significativas en las notas entre los alumnos que ha
```

```
t_stat: 0.161, p_value: 0.872
H0 <> No hay evidencia suficiente de diferencias significativas en las notas entre los alumnos que hacen volunta
riado y los que no.
```

Conclusiones

Variables con relación

- Existe una relación significativa entre el tiempo de estudio y las notas.
- Existe una relación significativa inversa entre las ausencias y las notas.
- Existen diferencias significativas en las notas entre los alumnos que tienen tutor y los que no.
- Existe una relación significativa entre el soporte parental y las notas.
- Existen diferencias significativas en las notas entre los alumnos que hacen actividades extracurriculares y los que no.

Variables sin relación

- No hay evidencia suficiente de diferencias significativas en las notas entre hombres y mujeres.
- No hay evidencia suficiente de diferencias significativas en las notas entre las diferentes etnias.
- No hay evidencia suficiente de diferencias significativas en las notas según el nivel de educación parental.