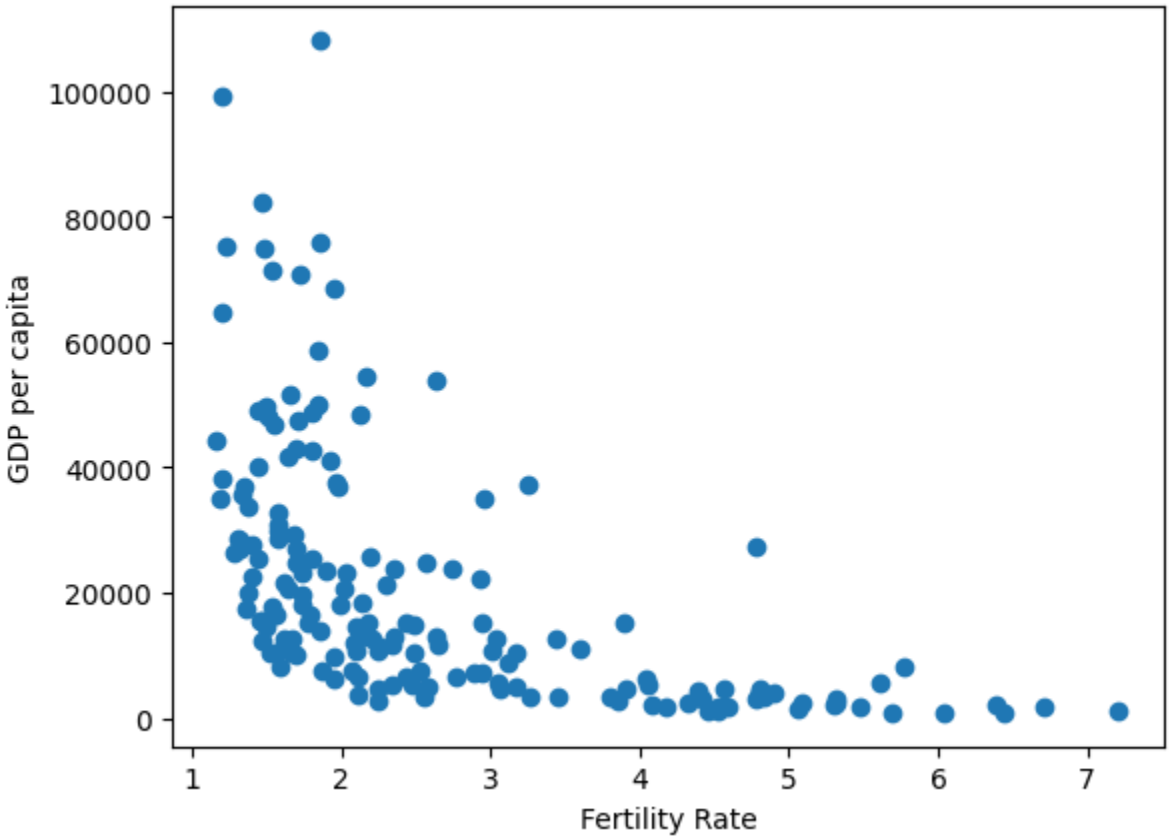
In [70]:
```
import pandas as pd
import matplotlib.pyplot as plt
from scipy.stats import linregress
df = pd.read_csv("MTA7DataSet.csv").set_index('Entity')
```

1. Create a scatter plot with Fertility as the Response variable and GDP as the explanatory variable. Then paste the plot below and answer the questions below.

In [71]:
```
fig, ax = plt.subplots()

ax.set_xlabel('Fertility Rate')
ax.set_ylabel('GDP per capita')
plt.scatter(df['Fertility rate'], df['GDP per capita'])
plt.show()
```



Q:Do you think there is a relationship between Fertility Rates and GDP?

A: Yes

Q: Is this a linear relationship or a non-linear relationship?

A: Non-linear

Q: What is the direction of the relationship?

A: Negative

Q: Are there any Outliers?

A: Yes

2. Now use Minitab to find the Covariance, Standard Deviation for GDP, and Standard Deviation for Fertility and GDP. Paste the output below.

In [72]:
```
variableDf = df.drop(columns=['Code','Continent', 'Year', 'Population (histo
covDF = variableDf.cov()
fertStd = df['Fertility rate'].std()
gdpStd = df['GDP per capita'].std()
cov = covDF['Fertility rate'].iloc[1]
print(f"Covariance Fertility vs GDP Per Capita: {round(cov, 3)}")
print(f'Standard Devation Fertility: {round(fertStd,3)}, GDP: {round(gdpStd,
```

```
Covariance Fertility vs GDP Per Capita: -15378.761
Standard Devation Fertility: 1.379, GDP: 20570.05
```

Use the values from Minitab and the formula r=s_{xy}/(s_{x} s_{y}) to find the correlation coefficient.

3. Use Minitab to find the Correlation Coefficient and paste the output below. Does this value match the one you got from your calculation above? (If it does not then you did something wrong, try to fix it.)

In [73]:
```
corrCoef = cov/(fertStd * gdpStd)
print(f"correlation coefficient: {round(corrCoef, 3)}")
```

```
correlation coefficient: -0.542
```

4. Using the value for the correlation coefficient, find the coefficient of determination.

In [74]:
```
print(f'Coefficient of determination: {round(corrCoef**2, 3)}' )
```

```
Coefficient of determination: 0.294
```

5. Use Minitab to find the least squares linear regression line and paste the output below. Make sure to include the coefficient of determination. Does the value for the coefficient of determination match your calculation? (It should, fix it if it does not.)

In [75]:
```
slope, intercept, r, p, se = linregress(df['Fertility rate'], df['GDP per ca

print(f"Model: ŷ={round(intercept,3)}+({round(slope, 3)})x")
print(f'Coefficient of determination: {round(r**2,3)}')
```

```
Model: ŷ=42430.135+(-8081.94)x
Coefficient of determination: 0.294
```

6. Use Minitab to construct the scatterplot with the linear regression line and paste the output below.

In [76]:
```
fig, ax= plt.subplots()
ax.set_xlabel('Fertility Rate')
ax.set_ylabel('GDP per capita')

ax.axline((0,intercept), slope = slope)
scatter = plt.scatter(df['Fertility rate'], df['GDP per capita'])
ax.set_xlim(df['Fertility rate'].min()-df['Fertility rate'].std()/2)
plt.show()
```