```
import pandas as pd
df = pd.read_csv('SYB66_246_202310_Population Growth, Fertility and Mortality Indicators.csv', encoding= 'latin1', skiprows= 1)
df
```

Out[7]:

| | Region/Country/Area | Unnamed: 1 | Year | Series | Value | Footnotes | Source |
|---|---|---|---|---|---|---|---|
| 0 | 1 | Total, all countries or areas | 2010 | Population annual rate of increase (percent) | 1.3 | NaN | United Nations Population Division, New York, ... |
| 1 | 1 | Total, all countries or areas | 2010 | Total fertility rate (children per women) | 2.6 | NaN | United Nations Population Division, New York, ... |
| 2 | 1 | Total, all countries or areas | 2010 | Infant mortality for both sexes (per 1,000 liv... | 37.1 | NaN | United Nations Statistics Division, New York, ... |
| 3 | 1 | Total, all countries or areas | 2010 | Maternal mortality ratio (deaths per 100,000 p... | 254 | NaN | World Health Organization (WHO), the United Na... |
| 4 | 1 | Total, all countries or areas | 2010 | Life expectancy at birth for both sexes (years) | 70.1 | NaN | United Nations Population Division, New York, ... |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 6751 | 722 | SIDS | 2022 | Total fertility rate (children per women) | 2.3 | Projected estimate (medium fertility variant). | United Nations Population Division, New York, ... |
| 6752 | 722 | SIDS | 2022 | Infant mortality for both sexes (per 1,000 liv... | 27.3 | Projected estimate (medium fertility variant). | United Nations Statistics Division, New York, ... |
| 6753 | 722 | SIDS | 2022 | Life expectancy at birth for both sexes (years) | 72.5 | Projected estimate (medium fertility variant). | United Nations Population Division, New York, ... |

In [8]:
```
df['Series'][3]
```

Out[8]:
```
'Maternal mortality ratio (deaths per 100,000 population)'
```

In [9]:
```
# obtenemos las tasa infantil de mortalidad
# primero creamos un nuevo DF con la informacion de la Columna 'Series' filtrando por el campo ('Infant mortality for both sexes

df_mort = df[df['Series']== 'Infant mortality for both sexes (per 1,000 live births)']
df_mort
```

Out[9]:

| | Region/Country/Area | Unnamed: 1 | Year | Series | Value | Footnotes | Source |
|---|---|---|---|---|---|---|---|
| 2 | 1 | Total, all countries or areas | 2010 | Infant mortality for both sexes (per 1,000 liv... | 37.1 | NaN | United Nations Statistics Division, New York, ... |
| 9 | 1 | Total, all countries or areas | 2015 | Infant mortality for both sexes (per 1,000 liv... | 31.5 | NaN | United Nations Statistics Division, New York, ... |
| 16 | 1 | Total, all countries or areas | 2020 | Infant mortality for both sexes (per 1,000 liv... | 28.3 | NaN | United Nations Statistics Division, New York, ... |
| 23 | 1 | Total, all countries or areas | 2022 | Infant mortality for both sexes (per 1,000 liv... | 27.5 | Projected estimate (medium fe... | United Nations Statistics Division, New York, ... |
| 29 | 2 | Africa | 2010 | Infant mortality for both sexes (per 1,000 liv... | 60.9 | | tics Division, New York, ... |

In [10]:
```
# filtramos el DF con los valores que sean igual al año '2020'
df_mort = df[df['Series']== 'Infant mortality for both sexes (per 1,000 live births)']
df_mort = df_mort[df_mort['Year'] == 2020]
df_mort
```

Out[10]:

| | Region/Country/Area | Unnamed: 1 | Year | Series | Value | Footnotes | Source |
|---|---|---|---|---|---|---|---|
| 16 | 1 | Total, all countries or areas | 2020 | Infant mortality for both sexes (per 1,000 liv... | 28.3 | NaN | United Nations Statistics Division, New York, ... |
| 43 | 2 | Africa | 2020 | Infant mortality for both sexes (per 1,000 liv... | 46.4 | NaN | United Nations Statistics Division, New York, ... |
| 70 | 15 | Northern Africa | 2020 | Infant mortality for both sexes (per 1,000 liv... | 22.6 | NaN | United Nations Statistics Division, New York, ... |
| 97 | 202 | Sub-Saharan Africa | 2020 | Infant mortality for both sexes (per 1,000 liv... | 50.0 | NaN | United Nations Statistics Division, New York, ... |
| 124 | 14 | Eastern Africa | 2020 | Infant mortality for both sexes (per 1,000 liv... | 38.3 | NaN | United Nations Statistics Division, New York, ... |

In [11]:
```
df_mort = df[df['Series']== 'Infant mortality for both sexes (per 1,000 live births)']
df_mort = df_mort[df_mort['Year'] == 2020]
df_mort = df_mort[['Unnamed: 1', 'Value']]
df_mort.rename({'Unnamed: 1': 'Country', 'Value': 'Infant Mortality'}, axis = 1, inplace= True) # renombramos las columnas 'Unn
# axis=1 especifica que se están cambiando los nombres de las columnas (en lugar de los índices de fila)
# inplace=True indica que los cambios se aplican directamente al DataFrame df_tasa_mort sin necesidad de crear uno nuevo.
df_mort = df_mort.set_index('Country')
df_mort
```

Out[11]:

| Country | Infant Mortality |
|---|---|
| Total, all countries or areas | 28.3 |
| Africa | 46.4 |
| Northern Africa | 22.6 |
| Sub-Saharan Africa | 50.0 |
| Eastern Africa | 38.3 |

In [12]:
```
# Obtenemos informacion de la esperanza de vida al nacer
# Filtramos de la columna 'Series' el campo 'Life expectancy at birth for both sexes (years)''
df_esp = df[df['Series']== 'Life expectancy at birth for both sexes (years)']
df_esp
```

Out[12]:

| | Region/Country/Area | Unnamed: 1 | Year | Series | Value | Footnotes | Source |
|---|---|---|---|---|---|---|---|
| 4 | 1 | Total, all countries or areas | 2010 | Life expectancy at birth for both sexes (years) | 70.1 | NaN | United Nations Population Division, New York, ... |
| 11 | 1 | Total, all countries or areas | 2015 | Life expectancy at birth for both sexes (years) | 71.8 | NaN | United Nations Population Division, New York, ... |
| 18 | 1 | Total, all countries or areas | 2020 | Life expectancy at birth for both sexes (years) | 72.0 | NaN | United Nations Population Division, New York, ... |
| 24 | 1 | Total, all countries or areas | 2022 | Life expectancy at birth for both sexes (years) | 71.7 | Projected estimate (medium fertility variant). | United Nations Population Division, New York, ... |
| 31 | 2 | Africa | 2010 | Life expectancy at birth for both sexes (years) | 58.6 | NaN | United Nations Population Division, New York, ... |

In [13]:
```
df_esp = df[df['Series']== 'Life expectancy at birth for both sexes (years)']
df_esp = df_esp[df_esp['Year']== 2020]
df_esp = df_esp[['Unnamed: 1','Value']]
df_esp.rename({'Unnamed: 1': 'Country', 'Value': 'Life expectancy'}, axis= 1, inplace = True)
df_esp= df_esp.set_index('Country')
df_esp
```

Out[13]:

| Country | Life expectancy |
|---|---|
| Total, all countries or areas | 72.0 |
| Africa | 62.2 |
| Northern Africa | 71.2 |
| Sub-Saharan Africa | 60.3 |
| Eastern Africa | 63.7 |
| ... | ... |

In [42]:
```
# Tasa de crecimiento anual
df_creci = df[df['Series']== 'Population annual rate of increase (percent)']
df_creci = df_creci[df_creci['Year']== 2020]
df_creci = df_creci[['Unnamed: 1','Value']]
df_creci.rename({'Unnamed: 1': 'Country','Value': 'Population Increase'}, axis = 1, inplace= True)
df_creci = df_creci.set_index('Country')
df_creci
```

Out[42]:

| Country | Population Increase |
|---|---|
| Total, all countries or areas | 0.9 |
| Africa | 2.4 |
| Northern Africa | 1.8 |
| Sub-Saharan Africa | 2.6 |
| Eastern Africa | 2.6 |

In [44]:
```
# tasa global fecundidad
df_fecu = df[df['Series']== 'Total fertility rate (children per women)']
df_fecu = df_fecu[df_fecu['Year']== 2020]
df_fecu = df_fecu[['Unnamed: 1','Value']]
df_fecu.rename({'Unnamed: 1': 'Country','Value': 'Fertility rate'}, axis = 1, inplace= True)
df_fecu = df_fecu.set_index('Country')
df_fecu
```

Out[44]:

| Country | Fertility rate |
|---|---|
| Total, all countries or areas | 2.3 |
| Africa | 4.4 |
| Northern Africa | 3.1 |
| Sub-Saharan Africa | 4.7 |
| Eastern Africa | 4.3 |

In [45]:
```
# esperanza de vida al nacer de hombres
df_vida_H = df[df['Series']== 'Life expectancy at birth for males (years)']
df_vida_H = df_vida_H[df_vida_H['Year']== 2020]
df_vida_H = df_vida_H[['Unnamed: 1','Value']]
df_vida_H.rename({'Unnamed: 1': 'Country','Value': 'Life expectancy (males)'}, axis = 1, inplace= True)
df_vida_H = df_vida_H.set_index('Country')
df_vida_H
```

Out[45]:

| Country | Life expectancy (males) |
|---|---|
| Total, all countries or areas | 69.4 |
| Africa | 60.3 |
| Northern Africa | 69.0 |
| Sub-Saharan Africa | 58.4 |
| Eastern Africa | 61.2 |

In [46]:
```
# esperanza de vida al nacer de mujeres
df_vida_M = df[df['Series']== 'Life expectancy at birth for females (years)']
df_vida_M = df_vida_M[df_vida_M['Year']== 2020]
df_vida_M = df_vida_M[['Unnamed: 1','Value']]
df_vida_M.rename({'Unnamed: 1': 'Country','Value': 'Life expectancy (females)'}, axis = 1, inplace= True)
df_vida_M = df_vida_M.set_index('Country')
df_vida_M
```

Out[46]:

| Country | Life expectancy (females) |
|---|---|
| **Total, all countries or areas** | 74.8 |
| **Africa** | 64.2 |
| **Northern Africa** | 73.6 |
| **Sub-Saharan Africa** | 62.2 |
| **Eastern Africa** | 66.3 |

In [47]:
```
# Reunimos en un solo DF la informacion Obtenida anteriormente
# join='outer' indica que se realizará una unión externa, lo que significa que todas las filas y columnas de ambos DataFrames
# se conservarán, y las celdas faltantes se llenarán con valores NaN.

cont_inf = pd.concat([df_mort, df_esp, df_creci, df_fecu, df_vida_H, df_vida_M], join = 'outer', axis=1)
cont_inf = cont_inf[1:] # excluimos la primer fila 'Total, all countries or areas'
cont_inf = cont_inf.dropna() # Eliminamos todas las filas que contienen valores faltantes (NaN)
cont_inf
```

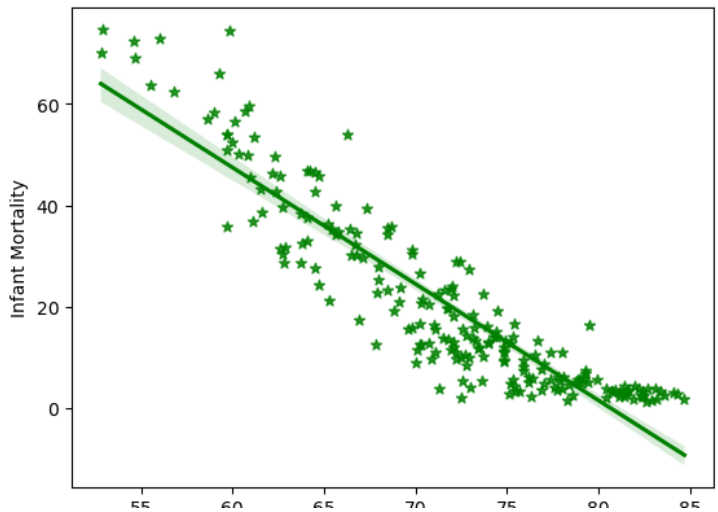
Out[47]:

| Country | Infant Mortality | Life expectancy | Population Increase | Fertility rate | Life expectancy (males) | Life expectancy (females) |
|---|---|---|---|---|---|---|
| **Africa** | 46.4 | 62.2 | 2.4 | 4.4 | 60.3 | 64.2 |
| **Northern Africa** | 22.6 | 71.2 | 1.8 | 3.1 | 69.0 | 73.6 |
| **Sub-Saharan Africa** | 50.0 | 60.3 | 2.6 | 4.7 | 58.4 | 62.2 |
| **Eastern Africa** | 38.3 | 63.7 | 2.6 | 4.3 | 61.2 | 66.3 |
| **Middle Africa** | 54.0 | 59.7 | 3.1 | 5.7 | 57.7 | 61.7 |

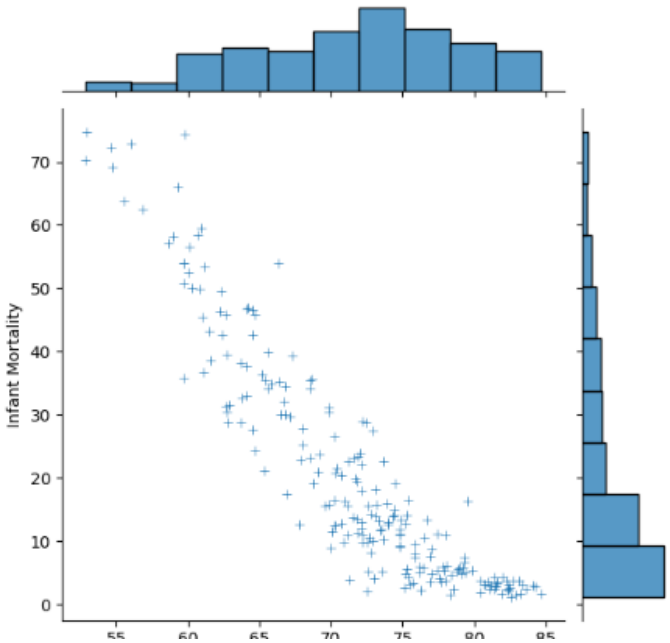
In [17]:
```
# Cambiamos el tipo de datos de Variable categorica a Variable numerica para poseterior graficar

cont_inf['Infant Mortality']= pd.to_numeric(cont_inf['Infant Mortality'], downcast= 'float')
cont_inf['Life expectancy']= pd.to_numeric(cont_inf['Life expectancy'], downcast= 'float')
```

```
import seaborn as sns
sns.regplot(data= cont_inf, y= 'Infant Mortality', x= 'Life expectancy', marker= '*', color = 'green')
```

Out[23]: <Axes: xlabel='Life expectancy', ylabel='Infant Mortality'>




In [34]:
```
graf = sns.jointplot(data = cont_inf, x= 'Life expectancy', y= 'Infant Mortality', kind = 'scatter', marker= '+')
```



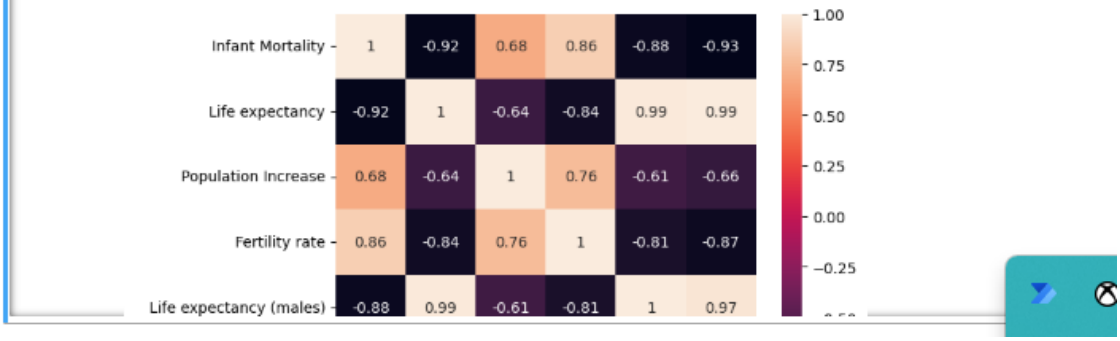
In [48]:
```
#  determinamos qué variables de esta base de datos se encuentran más correlacionadas en dicho año. Interprete sus resulta

cont_inf.corr()
```

Out[48]:

| | Infant Mortality | Life expectancy | Population Increase | Fertility rate | Life expectancy (males) | Life expectancy (females) |
|---|---|---|---|---|---|---|
| **Infant Mortality** | 1.000000 | -0.915268 | 0.681825 | 0.855781 | -0.881275 | -0.934610 |
| **Life expectancy** | -0.915268 | 1.000000 | -0.639729 | -0.844138 | 0.991792 | 0.992530 |
| **Population Increase** | 0.681825 | -0.639729 | 1.000000 | 0.762612 | -0.610315 | -0.663781 |
| **Fertility rate** | 0.855781 | -0.844138 | 0.762612 | 1.000000 | -0.805313 | -0.868746 |
| **Life expectancy (males)** | -0.881275 | 0.991792 | -0.610315 | -0.805313 | 1.000000 | 0.969268 |
| **Life expectancy (females)** | -0.934610 | 0.992530 | -0.663781 | -0.868746 | 0.969268 | 1.000000 |

In [51]:
```
sns.heatmap(cont_inf.corr(), annot= True)
```

Out[51]: <Axes: >




**Si la correlación es positiva, significa que ambas variables se mueven en la misma dirección.**

**Si es negativa, quiere decir que, cuando el valor de una variable aumenta, los valores de las otras variables disminuyen.**

**Ejemplo para Infant mortality inicia en (positivo) si la comparamos con Life Expectancy (males) vemos que el valor es negativo(-0.88) lo que nos indica que La esperanza de vida para los Hombres en la infancia es menor en relacion a la esperanza de vida para las mujeres (-0.93)**

In [ ]: