In [44]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

In [45]:
```
df = pd.read_csv(r"C:\Users\neetu27\Downloads\archive (3)\iris.data.csv")
```

In [46]:
```
df.shape
```

Out[46]:
```
(149, 5)
```

In [47]:
```
df.head()
```

Out[47]:

| | 5.1 | 3.5 | 1.4 | 0.2 | Iris-setosa |
|---|---|---|---|---|---|
| **0** | 4.9 | 3.0 | 1.4 | 0.2 | Iris-setosa |
| **1** | 4.7 | 3.2 | 1.3 | 0.2 | Iris-setosa |
| **2** | 4.6 | 3.1 | 1.5 | 0.2 | Iris-setosa |
| **3** | 5.0 | 3.6 | 1.4 | 0.2 | Iris-setosa |
| **4** | 5.4 | 3.9 | 1.7 | 0.4 | Iris-setosa |

In [48]:
```
df.tail()
```

Out[48]:

| | 5.1 | 3.5 | 1.4 | 0.2 | Iris-setosa |
|---|---|---|---|---|---|
| **144** | 6.7 | 3.0 | 5.2 | 2.3 | Iris-virginica |
| **145** | 6.3 | 2.5 | 5.0 | 1.9 | Iris-virginica |
| **146** | 6.5 | 3.0 | 5.2 | 2.0 | Iris-virginica |
| **147** | 6.2 | 3.4 | 5.4 | 2.3 | Iris-virginica |
| **148** | 5.9 | 3.0 | 5.1 | 1.8 | Iris-virginica |

In [49]:
```
df.columns
```

Out[49]:
```
Index(['5.1', '3.5', '1.4', '0.2', 'Iris-setosa'], dtype='object')
```

In [50]:
```
df.isna().sum()
```

Out[50]:
```
5.1            0
3.5            0
1.4            0
0.2            0
Iris-setosa    0
dtype: int64
```

In [51]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 149 entries, 0 to 148
Data columns (total 5 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   5.1          149 non-null    float64
 1   3.5          149 non-null    float64
 2   1.4          149 non-null    float64
 3   0.2          149 non-null    float64
 4   Iris-setosa  149 non-null    object 
dtypes: float64(4), object(1)
memory usage: 5.9+ KB
```

In [52]:
```
df.dtypes
```

Out[52]:
```
5.1            float64
3.5            float64
1.4            float64
0.2            float64
Iris-setosa     object
dtype: object
```

In [53]:
```
df.memory_usage()
```

Out[53]:
```
Index           128
5.1            1192
3.5            1192
1.4            1192
0.2            1192
Iris-setosa    1192
dtype: int64
```

In [54]:
```
df.describe(include='all')
```

Out[54]:

| | 5.1 | 3.5 | 1.4 | 0.2 | Iris-setosa |
|---|---|---|---|---|---|
| **count** | 149.000000 | 149.000000 | 149.000000 | 149.000000 | 149 |
| **unique** | NaN | NaN | NaN | NaN | 3 |
| **top** | NaN | NaN | NaN | NaN | Iris-versicolor |
| **freq** | NaN | NaN | NaN | NaN | 50 |
| **mean** | 5.848322 | 3.051007 | 3.774497 | 1.205369 | NaN |
| **std** | 0.828594 | 0.433499 | 1.759651 | 0.761292 | NaN |
| **min** | 4.300000 | 2.000000 | 1.000000 | 0.100000 | NaN |
| **25%** | 5.100000 | 2.800000 | 1.600000 | 0.300000 | NaN |
| **50%** | 5.800000 | 3.000000 | 4.400000 | 1.300000 | NaN |
| **75%** | 6.400000 | 3.300000 | 5.100000 | 1.800000 | NaN |
| **max** | 7.900000 | 4.400000 | 6.900000 | 2.500000 | NaN |

In [55]:
```
print(df[10:21])
# it will print the rows from 10 to 20.

# you can also save it in a variable for further use in analysis
sliced_data=df[10:21]
print(sliced_data)
```

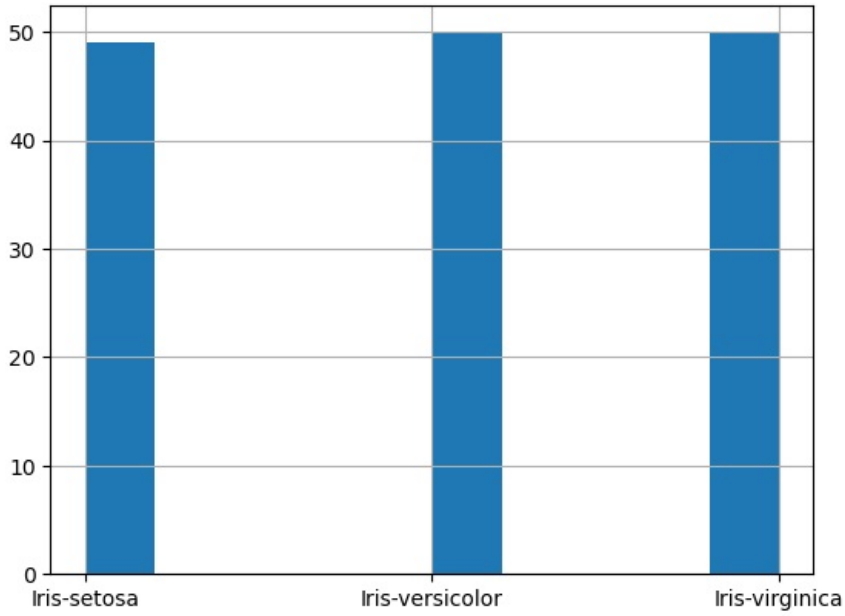
```
    5.1  3.5  1.4  0.2  Iris-setosa
10  4.8  3.4  1.6  0.2  Iris-setosa
11  4.8  3.0  1.4  0.1  Iris-setosa
12  4.3  3.0  1.1  0.1  Iris-setosa
13  5.8  4.0  1.2  0.2  Iris-setosa
14  5.7  4.4  1.5  0.4  Iris-setosa
15  5.4  3.9  1.3  0.4  Iris-setosa
16  5.1  3.5  1.4  0.3  Iris-setosa
17  5.7  3.8  1.7  0.3  Iris-setosa
18  5.1  3.8  1.5  0.3  Iris-setosa
19  5.4  3.4  1.7  0.2  Iris-setosa
20  5.1  3.7  1.5  0.4  Iris-setosa
    5.1  3.5  1.4  0.2  Iris-setosa
10  4.8  3.4  1.6  0.2  Iris-setosa
11  4.8  3.0  1.4  0.1  Iris-setosa
12  4.3  3.0  1.1  0.1  Iris-setosa
13  5.8  4.0  1.2  0.2  Iris-setosa
14  5.7  4.4  1.5  0.4  Iris-setosa
15  5.4  3.9  1.3  0.4  Iris-setosa
16  5.1  3.5  1.4  0.3  Iris-setosa
17  5.7  3.8  1.7  0.3  Iris-setosa
18  5.1  3.8  1.5  0.3  Iris-setosa
19  5.4  3.4  1.7  0.2  Iris-setosa
20  5.1  3.7  1.5  0.4  Iris-setosa
```

In [59]:
```
df.isnull().sum()
```

Out[59]:
```
5.1            0
3.5            0
1.4            0
0.2            0
Iris-setosa    0
dtype: int64
```

In [61]:
```
df['Iris-setosa'].hist()
```

Out[61]:
```
<Axes: >
```




In [67]:
```
df.corr()
```

```
C:\Users\neetu27\AppData\Local\Temp\ipykernel_3068\1134722465.py:1: FutureWarning: The default value of numeric
_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns
or specify the value of numeric_only to silence this warning.
  df.corr()
```

Out[67]:

| | 5.1 | 3.5 | 1.4 | 0.2 |
|---|---|---|---|---|
| **5.1** | 1.000000 | -0.103784 | 0.871283 | 0.816971 |
| **3.5** | -0.103784 | 1.000000 | -0.415218 | -0.350733 |
| **1.4** | 0.871283 | -0.415218 | 1.000000 | 0.962314 |
| **0.2** | 0.816971 | -0.350733 | 0.962314 | 1.000000 |

In [70]:
```
corr=df.corr()
fig,ax = plt.subplots(figsize=(5,4))
sns.heatmap(corr,annot=True,ax=ax)
```

```
C:\Users\neetu27\AppData\Local\Temp\ipykernel_3068\2376665244.py:1: FutureWarning: The default value of numeric
_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns
or specify the value of numeric_only to silence this warning.
  corr=df.corr()
```

Out[70]:
```
<Axes: >
```