In [1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn import preprocessing,svm
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import StandardScaler
```

In [3]:
```
df=pd.read_csv(r"C:\Users\LENOVO\Downloads\Advertising.csv")
```

In [4]:
```
df
```

Out[4]:

| | TV | Radio | Newspaper | Sales |
|---|---|---|---|---|
| **0** | 230.1 | 37.8 | 69.2 | 22.1 |
| **1** | 44.5 | 39.3 | 45.1 | 10.4 |
| **2** | 17.2 | 45.9 | 69.3 | 12.0 |
| **3** | 151.5 | 41.3 | 58.5 | 16.5 |
| **4** | 180.8 | 10.8 | 58.4 | 17.9 |
| **...** | ... | ... | ... | ... |
| **195** | 38.2 | 3.7 | 13.8 | 7.6 |
| **196** | 94.2 | 4.9 | 8.1 | 14.0 |
| **197** | 177.0 | 9.3 | 6.4 | 14.8 |
| **198** | 283.6 | 42.0 | 66.2 | 25.5 |
| **199** | 232.1 | 8.6 | 8.7 | 18.4 |

200 rows × 4 columns

In [5]:
```
df.head(18)
```

Out[5]:

| | TV | Radio | Newspaper | Sales |
|---|---|---|---|---|
| **0** | 230.1 | 37.8 | 69.2 | 22.1 |
| **1** | 44.5 | 39.3 | 45.1 | 10.4 |
| **2** | 17.2 | 45.9 | 69.3 | 12.0 |
| **3** | 151.5 | 41.3 | 58.5 | 16.5 |
| **4** | 180.8 | 10.8 | 58.4 | 17.9 |
| **5** | 8.7 | 48.9 | 75.0 | 7.2 |
| **6** | 57.5 | 32.8 | 23.5 | 11.8 |
| **7** | 120.2 | 19.6 | 11.6 | 13.2 |
| **8** | 8.6 | 2.1 | 1.0 | 4.8 |
| **9** | 199.8 | 2.6 | 21.2 | 15.6 |
| **10** | 66.1 | 5.8 | 24.2 | 12.6 |
| **11** | 214.7 | 24.0 | 4.0 | 17.4 |
| **12** | 23.8 | 35.1 | 65.9 | 9.2 |
| **13** | 97.5 | 7.6 | 7.2 | 13.7 |
| **14** | 204.1 | 32.9 | 46.0 | 19.0 |
| **15** | 195.4 | 47.7 | 52.9 | 22.4 |
| **16** | 67.8 | 36.6 | 114.0 | 12.5 |
| **17** | 281.4 | 39.6 | 55.8 | 24.4 |

In [6]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 4 columns):
 #   Column     Non-Null Count  Dtype  
---  ------     --------------  -----  
 0   TV         200 non-null    float64
 1   Radio      200 non-null    float64
 2   Newspaper  200 non-null    float64
 3   Sales      200 non-null    float64
dtypes: float64(4)
memory usage: 6.4 KB
```

In [7]:
```
df.describe()
```

Out[7]:

| | TV | Radio | Newspaper | Sales |
|---|---|---|---|---|
| **count** | 200.000000 | 200.000000 | 200.000000 | 200.000000 |
| **mean** | 147.042500 | 23.264000 | 30.554000 | 15.130500 |
| **std** | 85.854236 | 14.846809 | 21.778621 | 5.283892 |
| **min** | 0.700000 | 0.000000 | 0.300000 | 1.600000 |
| **25%** | 74.375000 | 9.975000 | 12.750000 | 11.000000 |
| **50%** | 149.750000 | 22.900000 | 25.750000 | 16.000000 |
| **75%** | 218.825000 | 36.525000 | 45.100000 | 19.050000 |
| **max** | 296.400000 | 49.600000 | 114.000000 | 27.000000 |

In [8]:
```
df=df[['Sales','TV','Radio','Newspaper']]
df.columns=['sales','tv','radio','newspaper']
```

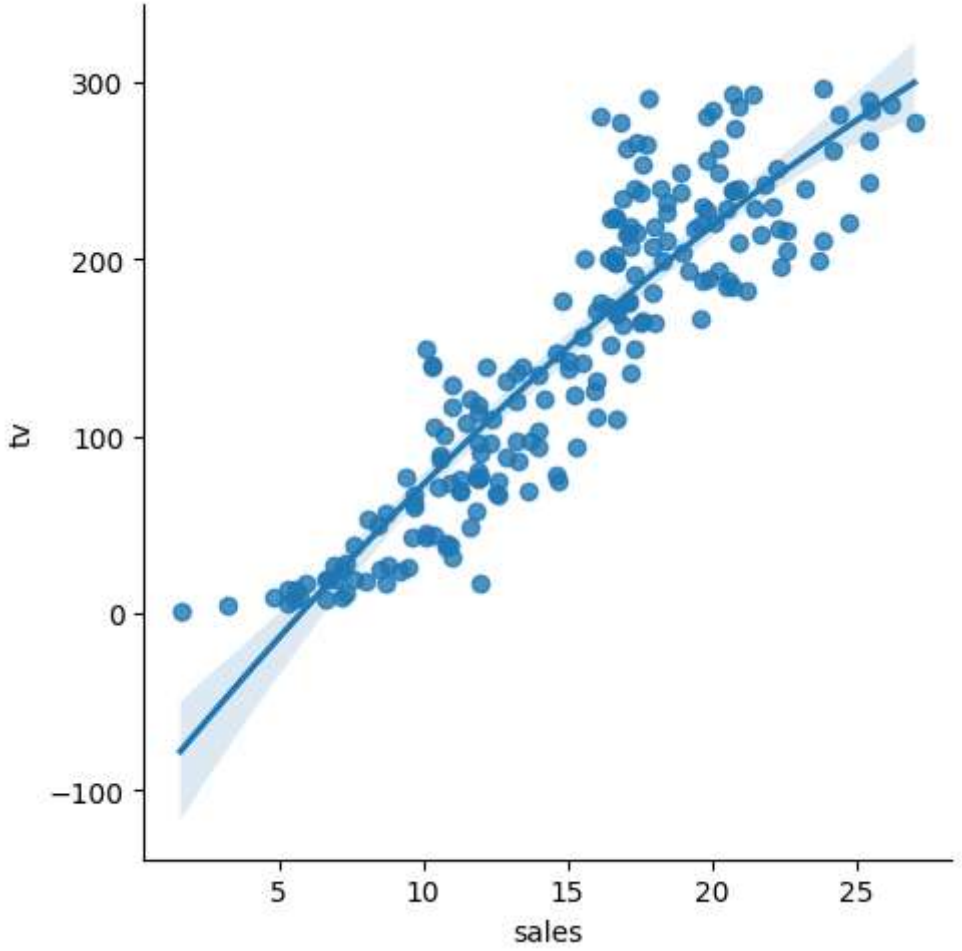
In [9]:
```
df.head(23)
```

Out[9]:

| | sales | tv | radio | newspaper |
|---|---|---|---|---|
| **0** | 22.1 | 230.1 | 37.8 | 69.2 |
| **1** | 10.4 | 44.5 | 39.3 | 45.1 |
| **2** | 12.0 | 17.2 | 45.9 | 69.3 |
| **3** | 16.5 | 151.5 | 41.3 | 58.5 |
| **4** | 17.9 | 180.8 | 10.8 | 58.4 |
| **5** | 7.2 | 8.7 | 48.9 | 75.0 |
| **6** | 11.8 | 57.5 | 32.8 | 23.5 |
| **7** | 13.2 | 120.2 | 19.6 | 11.6 |
| **8** | 4.8 | 8.6 | 2.1 | 1.0 |
| **9** | 15.6 | 199.8 | 2.6 | 21.2 |
| **10** | 12.6 | 66.1 | 5.8 | 24.2 |
| **11** | 17.4 | 214.7 | 24.0 | 4.0 |
| **12** | 9.2 | 23.8 | 35.1 | 65.9 |
| **13** | 13.7 | 97.5 | 7.6 | 7.2 |
| **14** | 19.0 | 204.1 | 32.9 | 46.0 |
| **15** | 22.4 | 195.4 | 47.7 | 52.9 |
| **16** | 12.5 | 67.8 | 36.6 | 114.0 |
| **17** | 24.4 | 281.4 | 39.6 | 55.8 |
| **18** | 11.3 | 69.2 | 20.5 | 18.3 |
| **19** | 14.6 | 147.3 | 23.9 | 19.1 |
| **20** | 18.0 | 218.4 | 27.7 | 53.4 |
| **21** | 17.5 | 237.4 | 5.1 | 23.5 |
| **22** | 5.6 | 13.2 | 15.9 | 49.6 |

In [10]:
```
sns.lmplot(x='sales',y='tv',data=df,order=2)
```

Out[10]:
```
<seaborn.axisgrid.FacetGrid at 0x1bbc58d4ed0>
```



In [11]:
```
df.describe()
```

Out[11]:

| | sales | tv | radio | newspaper |
|---|---|---|---|---|
| **count** | 200.000000 | 200.000000 | 200.000000 | 200.000000 |
| **mean** | 15.130500 | 147.042500 | 23.264000 | 30.554000 |
| **std** | 5.283892 | 85.854236 | 14.846809 | 21.778621 |
| **min** | 1.600000 | 0.700000 | 0.000000 | 0.300000 |
| **25%** | 11.000000 | 74.375000 | 9.975000 | 12.750000 |
| **50%** | 16.000000 | 149.750000 | 22.900000 | 25.750000 |
| **75%** | 19.050000 | 218.825000 | 36.525000 | 45.100000 |
| **max** | 27.000000 | 296.400000 | 49.600000 | 114.000000 |

In [12]:
```
df.fillna(method='ffill',inplace=True)
```

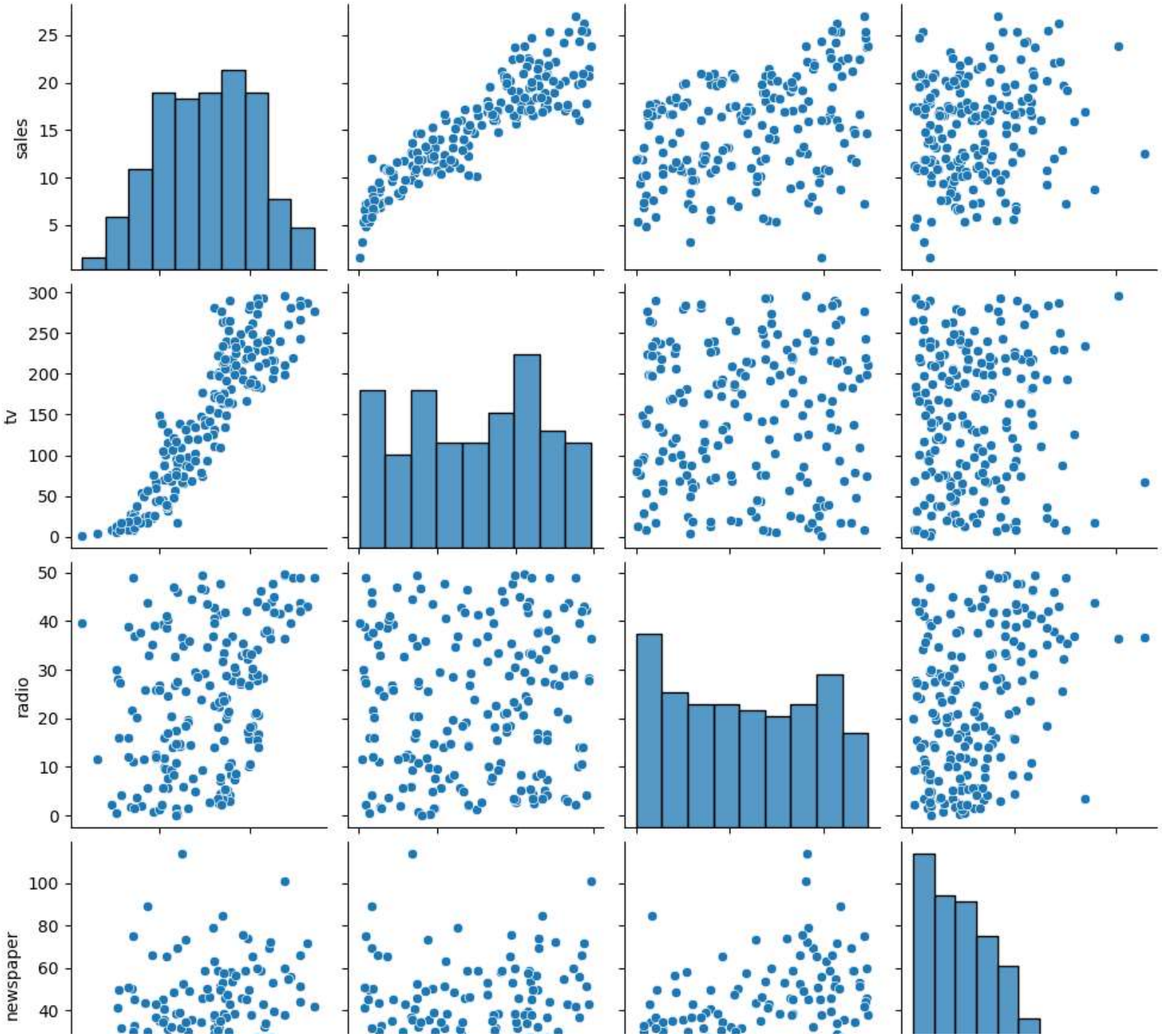
In [13]:
```
x=np.array(df['sales']).reshape(-1,1)
y=np.array(df['tv']).reshape(-1,1)
```

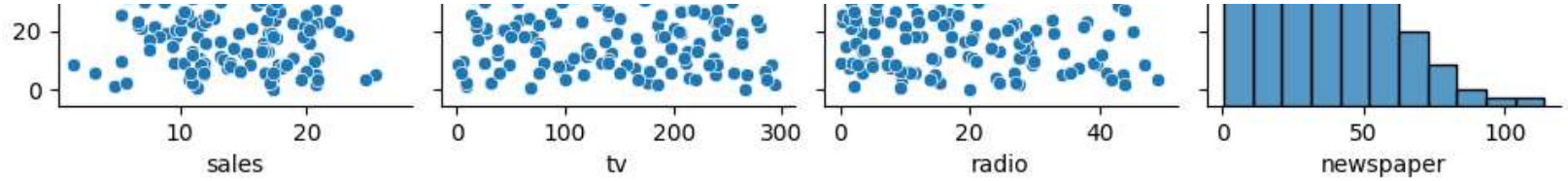
In [14]:
```
df.dropna(inplace=True)
```

In [15]:
```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.5)
```

In [16]:
```
sns.pairplot(df)
```

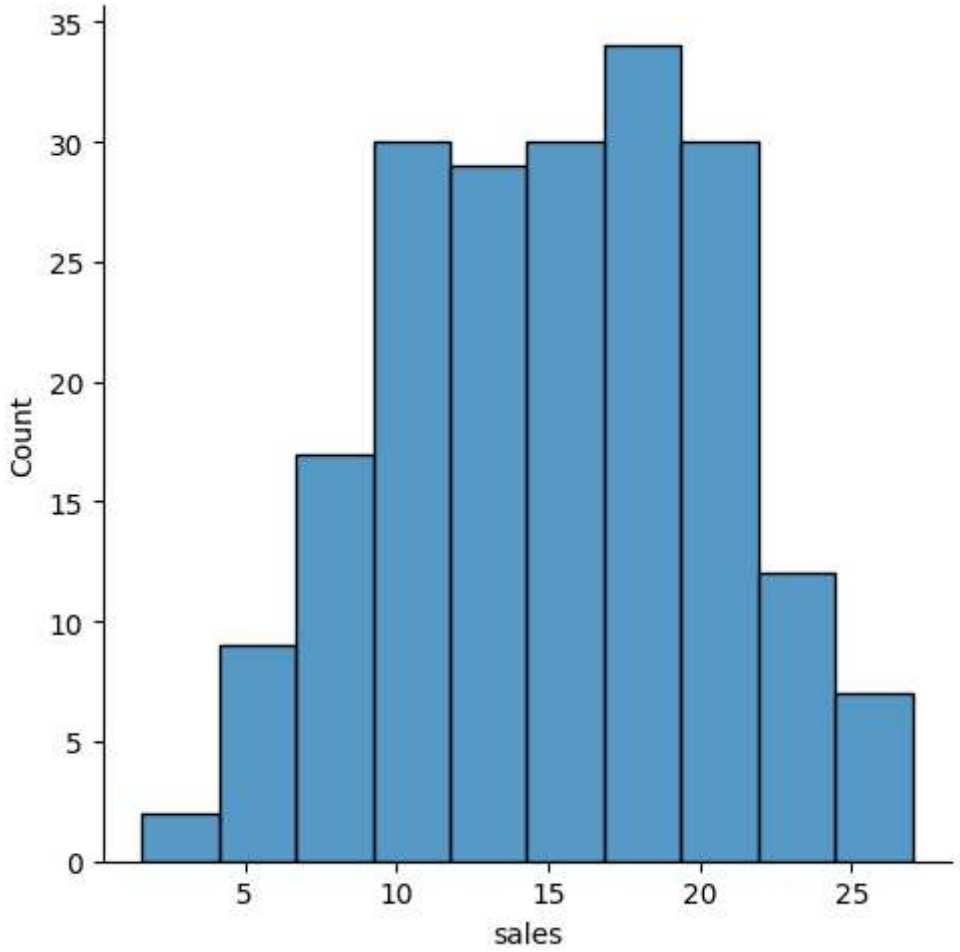
Out[16]:
```
<seaborn.axisgrid.PairGrid at 0x1bbd8879590>
```


sales
tv
radio
newspaper
25
20
15
10
5
300
250
200
150
100
50
0
50
40
30
20
10
0
100
80
60
40



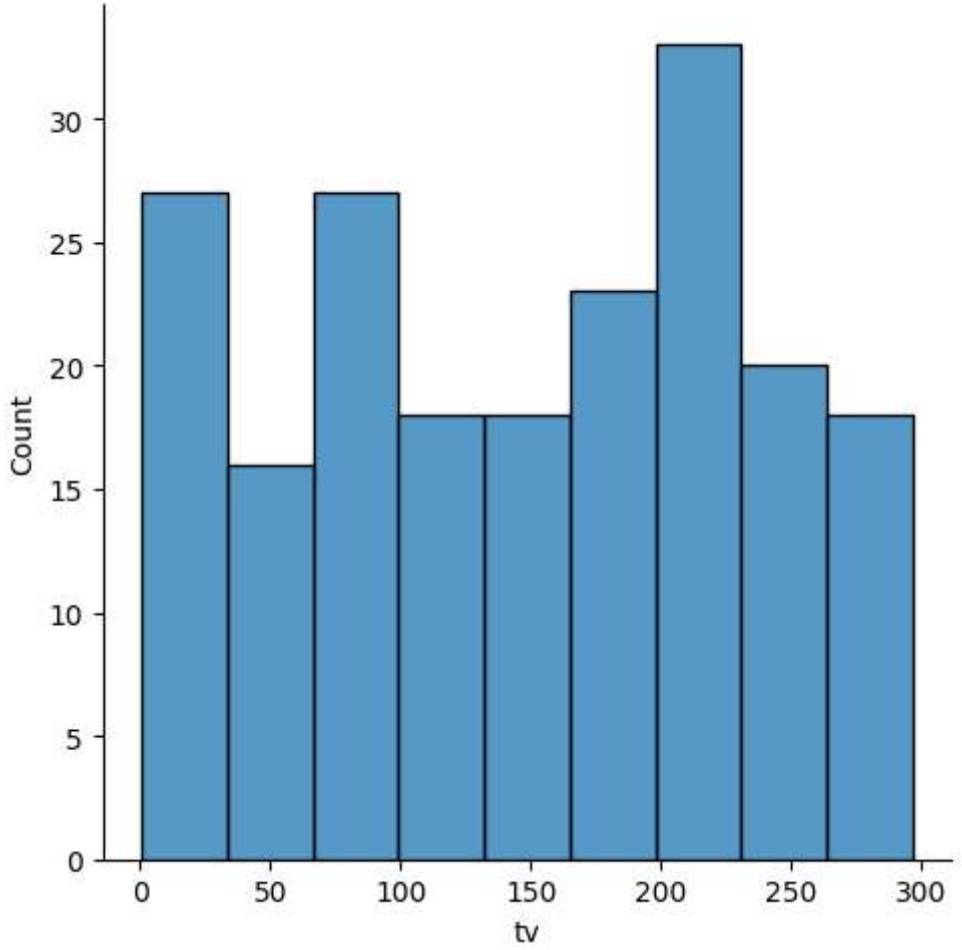

In [17]:
```
sns.displot(df['sales'])
```

Out[17]: <seaborn.axisgrid.FacetGrid at 0x1bbdbdeb810>



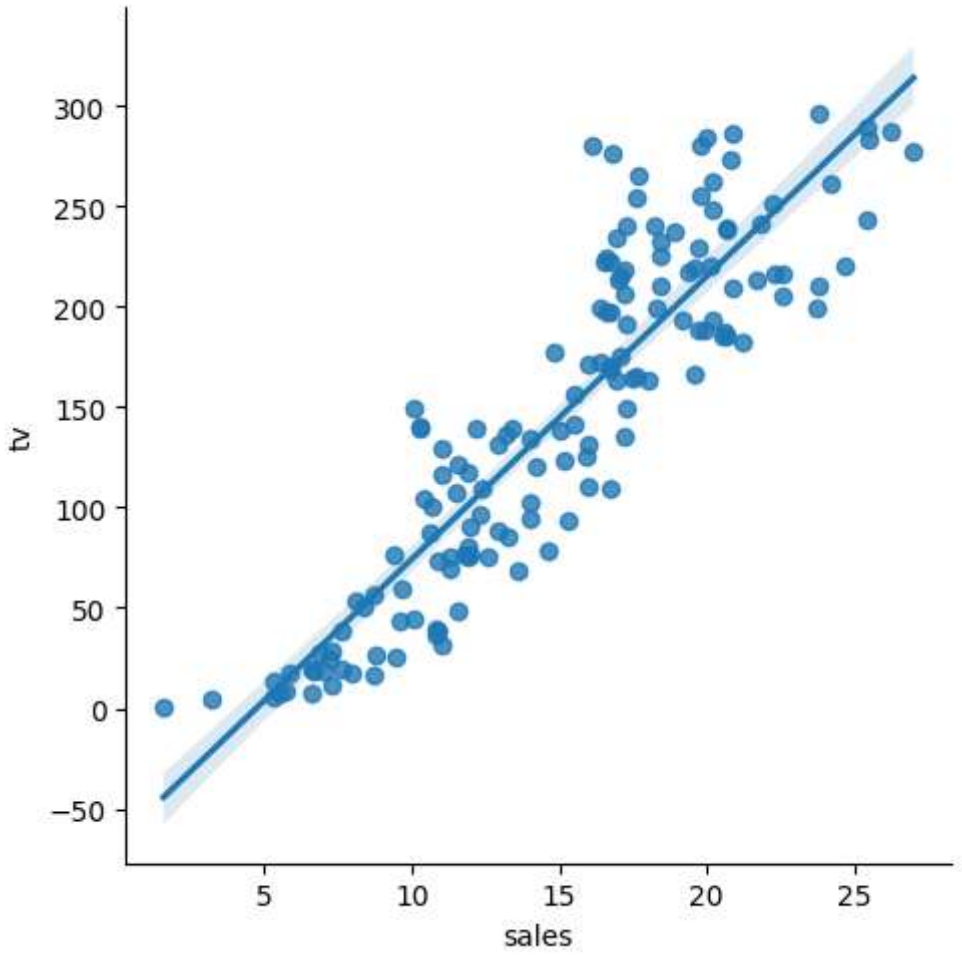
In [18]:
```
sns.displot(df['tv'])
```

Out[18]:
```
<seaborn.axisgrid.FacetGrid at 0x1bbdbec6cd0>
```



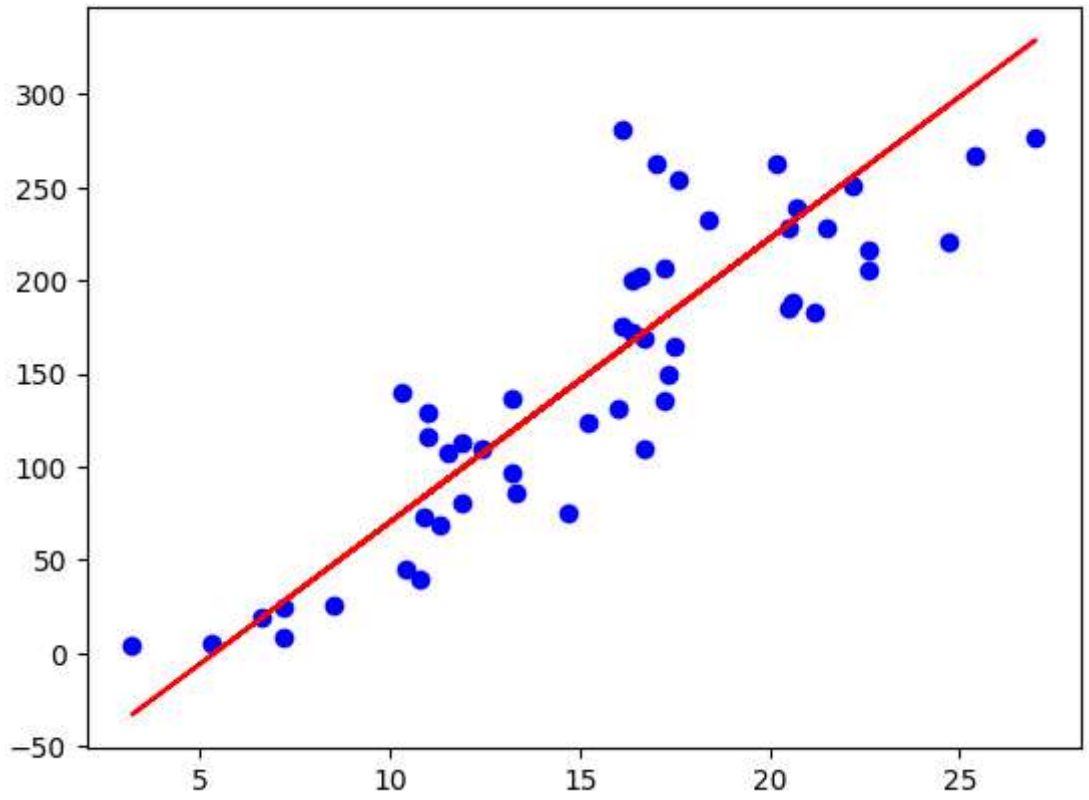
In [19]:
```
df500=df[:][50:500]
sns.lmplot(x='sales',y='tv',data=df500,order=1)
```

Out[19]: <seaborn.axisgrid.FacetGrid at 0x1bbdbc72210>



In [20]:
```
df500.fillna(method='ffill',inplace=True)
x=np.array(df['sales']).reshape(-1,1)
y=np.array(df['tv']).reshape(-1,1)
df.dropna(inplace=True)
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.25)
regr=LinearRegression()
regr.fit(x_train,y_train)
print("Regression:",regr.score(x_test,y_test))
y_pred=regr.predict(x_test)
plt.scatter(x_test,y_test,color='b')
plt.plot(x_test,y_pred,color='r')
plt.show()
```

```
Regression: 0.7405370859996265
```



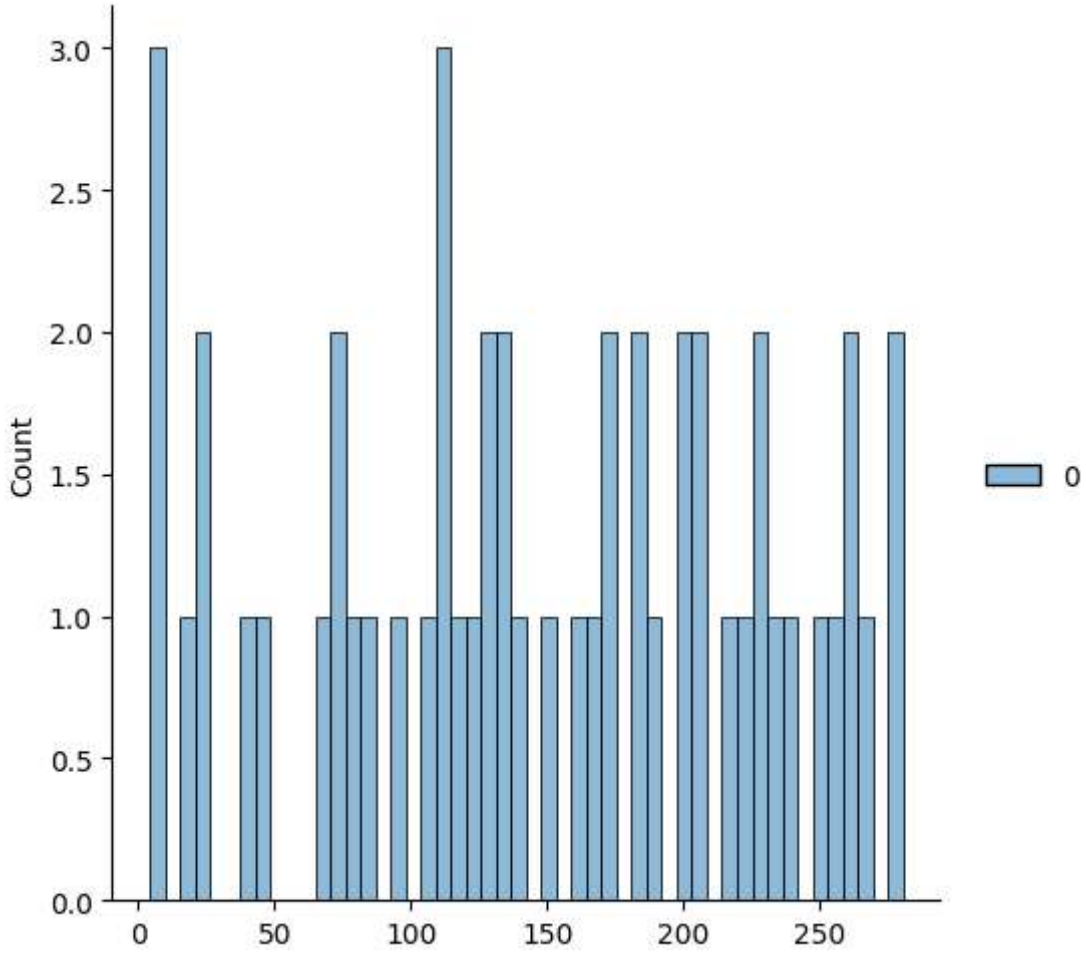
In [21]:
```
df.shape
```

Out[21]:
```
(200, 4)
```

In [22]:
```
df.isnull().sum()
```

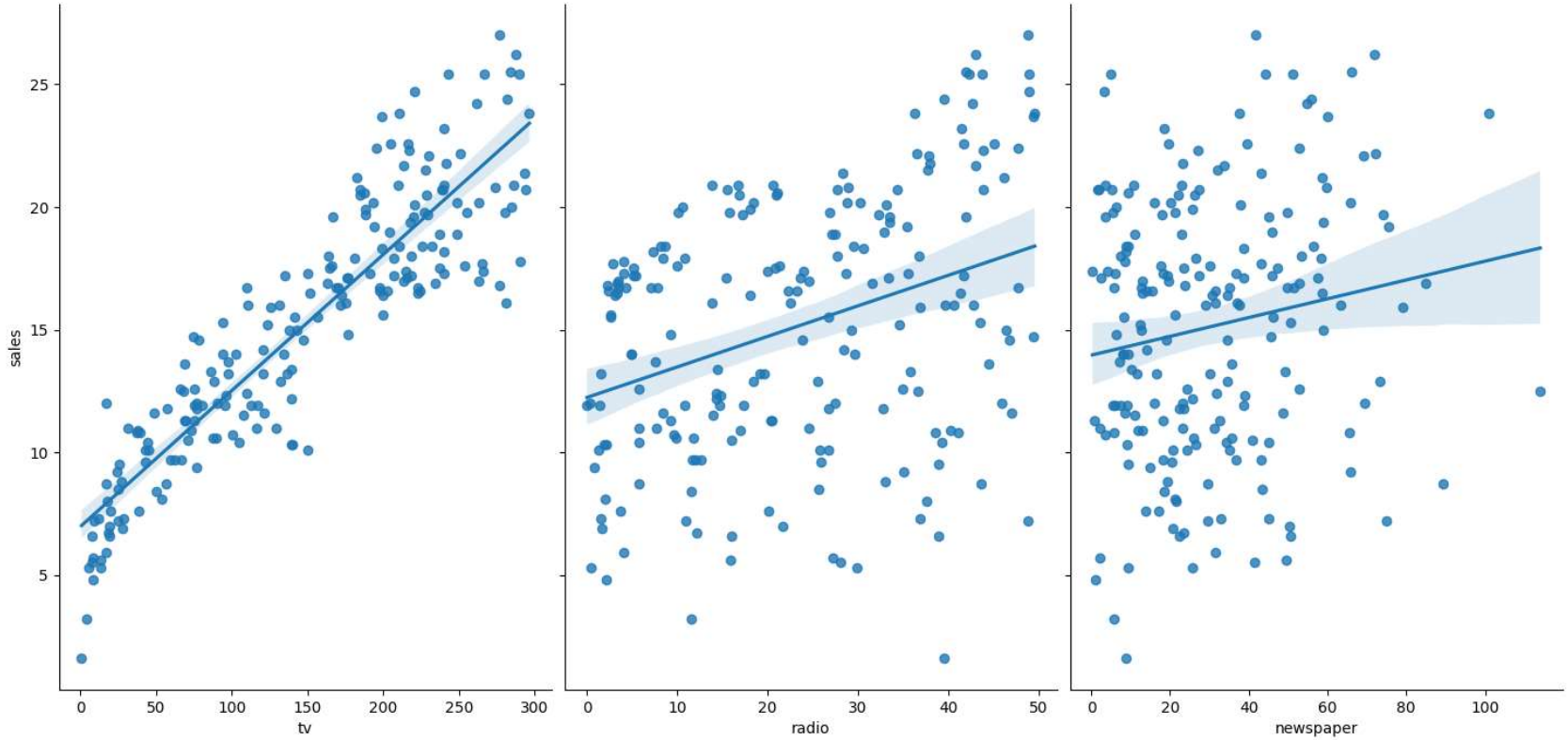
Out[22]:
```
sales        0
tv           0
radio        0
newspaper    0
dtype: int64
```

In [23]:
```
sns.displot((y_test),bins=50);
```



In [24]:
```
sns.pairplot(df,x_vars=['tv','radio','newspaper'],y_vars='sales',height=7,aspect=0.7,kind='reg')
```

Out[24]:
```
<seaborn.axisgrid.PairGrid at 0x1bbdc489590>
```




In [ ]:
```

```