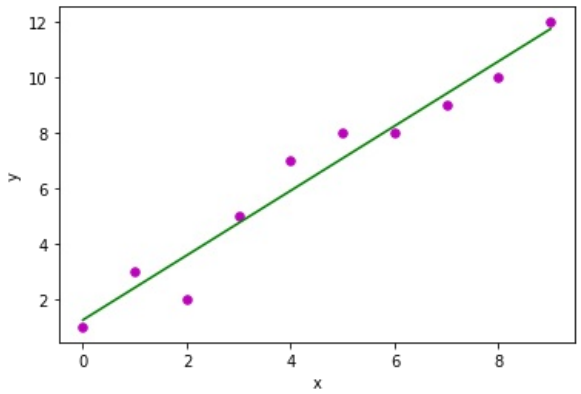
In [2]:

```
import numpy as np
import matplotlib.pyplot as plt

def estimate_coef(x, y):
        # number of observations/points
        n = np.size(x)

        # mean of x and y vector
        m_x, m_y = np.mean(x), np.mean(y)

        # calculating cross-deviation and deviation about x
        SS_xy = np.sum(y*x) - n*m_y*m_x
        SS_xx = np.sum(x*x) - n*m_x*m_x

        # calculating regression coefficients
        b_1 = SS_xy / SS_xx
        b_0 = m_y - b_1*m_x

        return(b_0, b_1)

def plot_regression_line(x, y, b):
        # plotting the actual points as scatter plot
        plt.scatter(x, y, color = "m",
                        marker = "o", s = 30)

        # predicted response vector
        y_pred = b[0] + b[1]*x

        # plotting the regression line
        plt.plot(x, y_pred, color = "g")

        # putting labels
        plt.xlabel('x')
        plt.ylabel('y')

        # function to show plot
        plt.show()

def main():
        # observations
        x = np.array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9])
        y = np.array([1, 3, 2, 5, 7, 8, 8, 9, 10, 12])

        # estimating coefficients
        b = estimate_coef(x, y)
        print("Estimated coefficients:\n b_0 = {} \ \n b_1 = {}".format(b[0], b[1]))

        # plotting regression line
        plot_regression_line(x, y, b)

if __name__ == "__main__":
        main()
```

```
Estimated coefficients:
 b_0 = 1.2363636363636363 \
 b_1 = 1.1696969696969697
```



In [2]:

```
import pandas as pd
import numpy as np
df=pd.read_csv('/home/anudeep/Desktop/Mall_Customers.csv')
print(df.head())
```

```
   CustomerID   Genre  Age  Annual Income (k$)  Spending Score (1-100)
0           1    Male   19                  15                      39
1           2    Male   21                  15                      81
2           3  Female   20                  16                       6
3           4  Female   23                  16                      77
4           5  Female   31                  17                      40
```

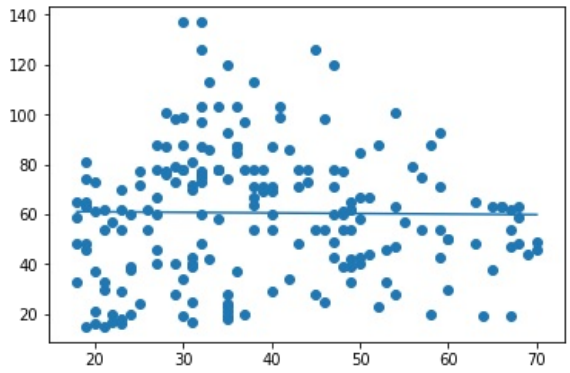
In [3]:

```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

df=pd.read_csv('/home/anudeep/Desktop/Mall_Customers.csv')
X=df['Age']
Y=df['Annual Income (k$)']
x_mean=np.mean(X)
y_mean=np.mean(Y)
n=len(X)
N=0
Z=0
for i in range(n):
    N+=(X[i]-x_mean)*(Y[i]-y_mean)
    Z+=(X[i]-x_mean)**2
M=(N/Z)
C=y_mean-(M*x_mean)
print(M,C)
min_x=np.min(X)
max_x=np.max(X)
x=np.linspace(min_x,max_x,1000)
y=M*x+C
plt.plot(x,y,label='Regression line')
plt.scatter(X,Y,label='scatter plot')
```

```
-0.02331097176261534 61.46563125297761
```

Out[3]:

```
<matplotlib.collections.PathCollection at 0x7f12dc3b4350>
```



In [14]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
anu={'x(year)':[2005,2006,2007,2008,2009],'y(sales)':[12,19,29,37,45]}
df=pd.DataFrame(anu)
X=df[['x(year)']]
y=df[['y(sales)']]
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.4,random_state=101)
from sklearn.linear_model import LinearRegression
lm=LinearRegression()
lm.fit(X_train,y_train)
print(lm.intercept_)
print(lm.coef_)
cdf=pd.DataFrame(lm.coef_,X.columns,columns=['coef'])
print(cdf)
```

```
[-18034.66666667]
[[9.]]
         coef
x(year)   9.0
```

In [8]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
iris=sns.load_dataset('iris')
iris.head()
X=iris[['sepal_length','petal_width']]
y=iris['petal_length']
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.4,random_state=101)
from sklearn.linear_model import LinearRegression
lm=LinearRegression()
lm.fit(X_train,y_train)
print(lm.intercept_)
print(lm.coef_)
cdf=pd.DataFrame(lm.coef_,X.columns,columns=['coef'])
print(cdf)
```

```
-1.5648268692704077
[0.55623269 1.70331289]
                  coef
sepal_length  0.556233
petal_width   1.703313
```

In [5]:

```
import pandas as pd
anu={'year':[2013,2014,2015,2016,2017,2018],'GDP':[6.2,6.5,5.48,6.54,7.18,7.93],'Car sales':[26.3,26.65,25.03,26.
01,27.9,30.47]}
df=pd.DataFrame(anu)
print(df)
```

```
   year   GDP  Car sales
0  2013  6.20      26.30
1  2014  6.50      26.65
2  2015  5.48      25.03
3  2016  6.54      26.01
4  2017  7.18      27.90
5  2018  7.93      30.47
```

In [9]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
anu={'x(year)':[2005,2006,2007,2008,2009],'y(sales)':[37,12,19,29,45]}
df=pd.DataFrame(anu)
X=df[['x(year)']]
y=df[['y(sales)']]
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.4,random_state=101)
from sklearn.linear_model import LinearRegression
lm=LinearRegression()
lm.fit(X_train,y_train)
print(lm.intercept_)
print(lm.coef_)
cdf=pd.DataFrame(lm.coef_,X.columns,columns=['coef'])
print(cdf)
```

```
[-17039.5]
[[8.5]]
         coef
x(year)   8.5
```

In [7]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
anu={'x(year)':[2005,2006,2007,2008,2009],'y(sales)':[37,12,19,29,45]}
df=pd.DataFrame(anu)
```

In [ ]:

```
X = np.array([6.2,6.5,5.48,6.54,7.18])
Y = np.array([26.65,25.03,26.01,27.9,30.47])
```