

## Homework 1 - Question #2

### Part 1

Draw the datapoints and regression curve.

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

points = {'x_values': [0, 2, 2, 5],
          'y_values': [2, 10, 12, 20],
          'estimation': [3, 11, 11, 23]}
points_dataset = pd.DataFrame(data=points)

points_dataset
```

	x_values	y_values	estimation
0	0	2	3
1	2	10	11
2	2	12	11
3	5	20	23

```
reg_points = {'reg_x': [0, 2, 5],
              'reg_y': [3, 11, 23]}
reg_dataset = pd.DataFrame(data=reg_points)

reg_dataset
```

	reg_x	reg_y
0	0	3
1	2	11
2	5	23

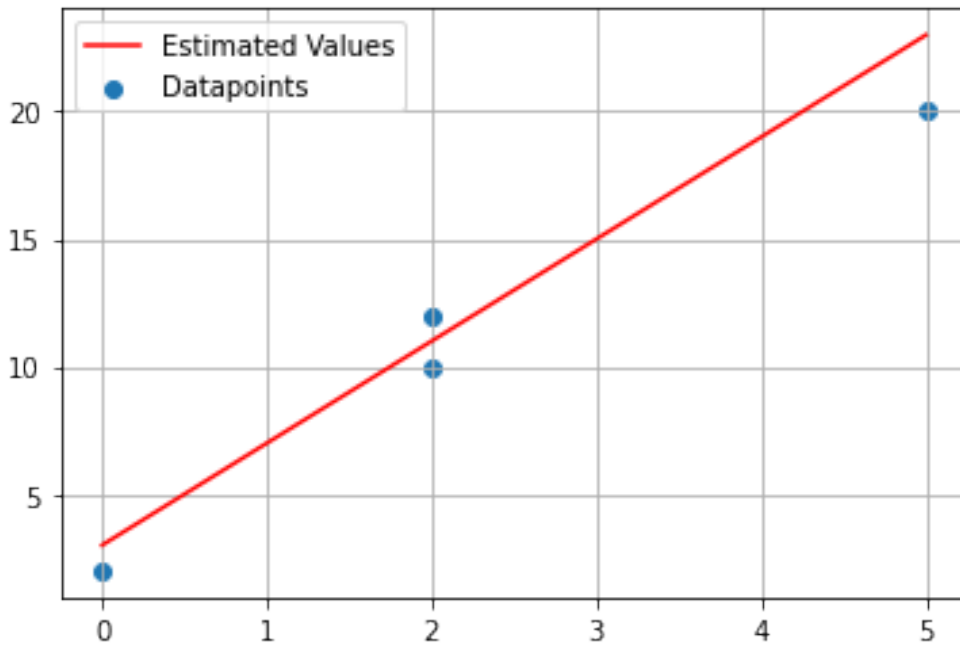
*#SOLUTION APPROACH 1: Regression Line by Enterin Points*

```
x = points_dataset.x_values
y = points_dataset.y_values

regg_x = reg_dataset.reg_x
regg_y = reg_dataset.reg_y

plt.scatter(x,y)
plt.plot(regg_x, regg_y, color = 'red')

plt.grid()
plt.legend(["Estimated Values", "Datapoints"])
plt.show()
```



*#SOLUTION APPROACH 2: Regression Line by Formula*

```
points = np.array([(0,3), (2,11), (5,23)])
```

```
# get x and y vectors
```

```
x = points[:,0]
```

```
y = points[:,1]
```

```
# calculate polynomial
```

```
z = np.polyfit(x, y, 1)
```

```
print(z)
```

```
a = z[0]
```

```
b = z[1]
```

```
y_pred = a*x + b
```

```
plt.plot(x, y_pred, color = 'red')
```

```
x = points_dataset.x_values
```

```
y = points_dataset.y_values
```

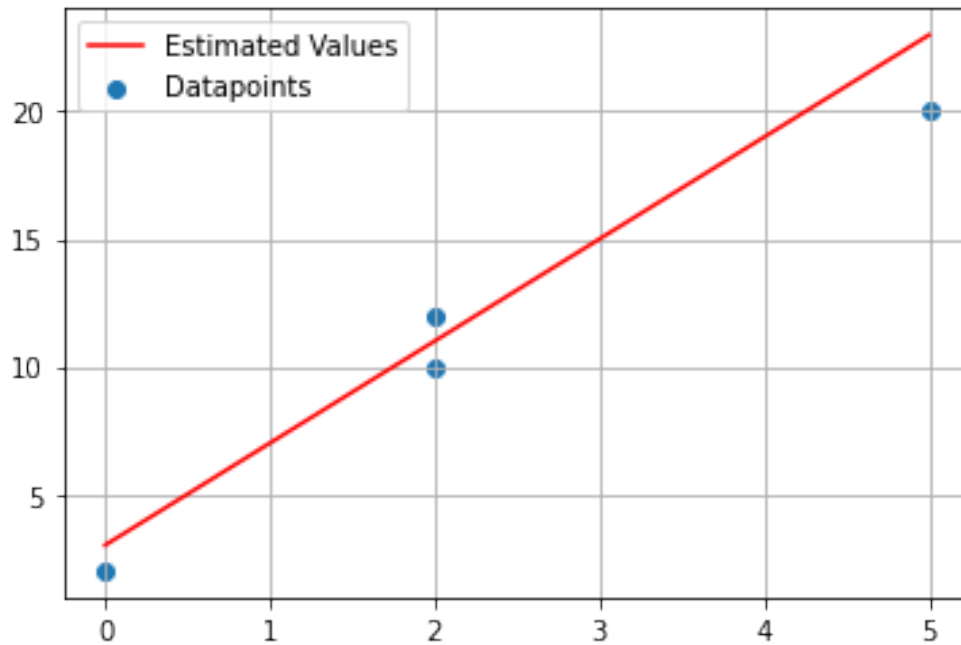
```
plt.scatter(x,y)
```

```
plt.grid()
```

```
plt.legend(["Estimated Values", "Datapoints"])
```

```
plt.show()
```

```
[4. 3.]
```



## Part 2

What is the MSE? Calculate error for each instance.

- Instance 1 --->  $(2 - 3)^2 = 1$
- Instance 2 --->  $(10 - 11)^2 = 1$
- Instance 3 --->  $(12 - 11)^2 = 1$
- Instance 4 --->  $(20 - 23)^2 = 9$

MSE (Mean Squared Error) =  $(1 + 1 + 1 + 9) / 4 = 3$