

FIND

- ALPHA
- BETA
- MEAN VALUES
- MARK'S Mark if he plans to spend 75 hours studying and scored 24 for his quiz

What do ALPHA & BETA values represent in this model?

```
import pandas as pd
from sklearn import linear_model
import numpy as np
from google.colab import drive
```

```
drive.mount("/content/gdrive", force_remount=True)
df = pd.read_csv("/content/gdrive/My Drive/Colab Notebooks/scoreTable.csv")
```

```
Mounted at /content/gdrive
```

OUTPUT

CODE

EXTRA COMMENT

```
# print table  
df
```

	Student	Hours	Score	Mark
0	Ming	130	15	78
1	Delphine	50	20	69
2	Amir	115	23	80
3	Bibek	80	12	65
4	Alesso	65	20	70
5	Farah	150	25	87
6	Pham	78	30	80
7	Bruce	67	12	82
8	Dalisay	60	10	70
9	Susilo	100	35	90

We can use multiple linear regression, which find a function of the form:

$$y = a + bx_1 + bx_2 + \dots + bx_n$$

$$\text{Mark} = a(\text{intercept}) + \text{Hours}x_1 + \text{Score}x_2$$

```
# Create Linear Regression class model
reg = linear_model.LinearRegression()
# fit model
# Independent Variables = 'Hours','Score' & Dependent variable = Mark
```

```
reg.fit(df[['Hours','Score']],df.Mark)
```

```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

```
# find coefficient 1st number is BETA(x1) and 2nd is BETA(x2)
reg.coef_
```

```
array([0.11409458, 0.51822625])
```

```
BETA(x1) = 0.11409458
```

```
BETA(x2) = 0.51822625
```

```
# find intercept
reg.intercept_
```

```
56.420365122476255
```

```
ALPHA = 56.420365122476255
```

```
# predict MARK'S Mark if he studies 75h & gets a 24 at Quiz
reg.predict([[75,24]])
```

```
array([77.41488841])
```

```
# how did it calculate that
```

```
# {[coefficient(BETA(x1))] * 75 + [coefficient(BETA(x2)) * 24] + Intercept (ALPHA)}
```

```
0.11409458*75 + 0.51822625*24 + 56.420365122476255
```

```
77.41488862247625
```

- MEAN X1 = 89.5
- MEAN X2 = 20.2
- MEAN Y = 77.1

HOW CAN WE CALCULATE ALPHA & BETA?

Regression Equation = $\hat{y} = b_1X_1 + b_2X_2 + a$

$$b_1 = ((SPX_1Y) * (SSX_2) - (SPX_1X_2) * (SPX_2Y)) / ((SSX_1) * (SSX_2) - (SPX_1X_2) * (SPX_1X_2)) = 617630.2 / 5413317.8 = 0.11409$$

$$b_2 = ((SPX_2Y) * (SSX_1) - (SPX_1X_2) * (SPX_1Y)) / ((SSX_1) * (SSX_2) - (SPX_1X_2) * (SPX_1X_2)) = 2805323.4 / 5413317.8 = 0.51823$$

$$a = MY - b_1MX_1 - b_2MX_2 = 77.1 - (0.11 * 89.5) - (0.52 * 20.2) = 56.42037$$

$$\hat{y} = 0.11409x_1 + 0.51823x_2 + 56.42037 = 77.4148884$$

Sum of $X_1 = 895$

Sum of $X_2 = 202$

Sum of $Y = 771$

Mean $X_1 = 89.5$

Mean $X_2 = 20.2$

Mean $Y = 77.1$

Sum of squares (SSX_1) = 9820.5

Sum of squares (SSX_2) = 611.6

Sum of products (SPX_1Y) = 1519.5

Sum of products (SPX_2Y) = 404.8

Sum of products (SPX_1X_2) = 770

a = intercept

ANSWERS:

- MEAN X1 = 89.5
- MEAN X2 = 20.2
- MEAN Y = 77.1
- BETA(x1) = 0.11409458
- BETA(x2) = 0.51822625
- ALPHA (intercept) = 56.420365122476255
- MARK'S MARK = 77.41488841