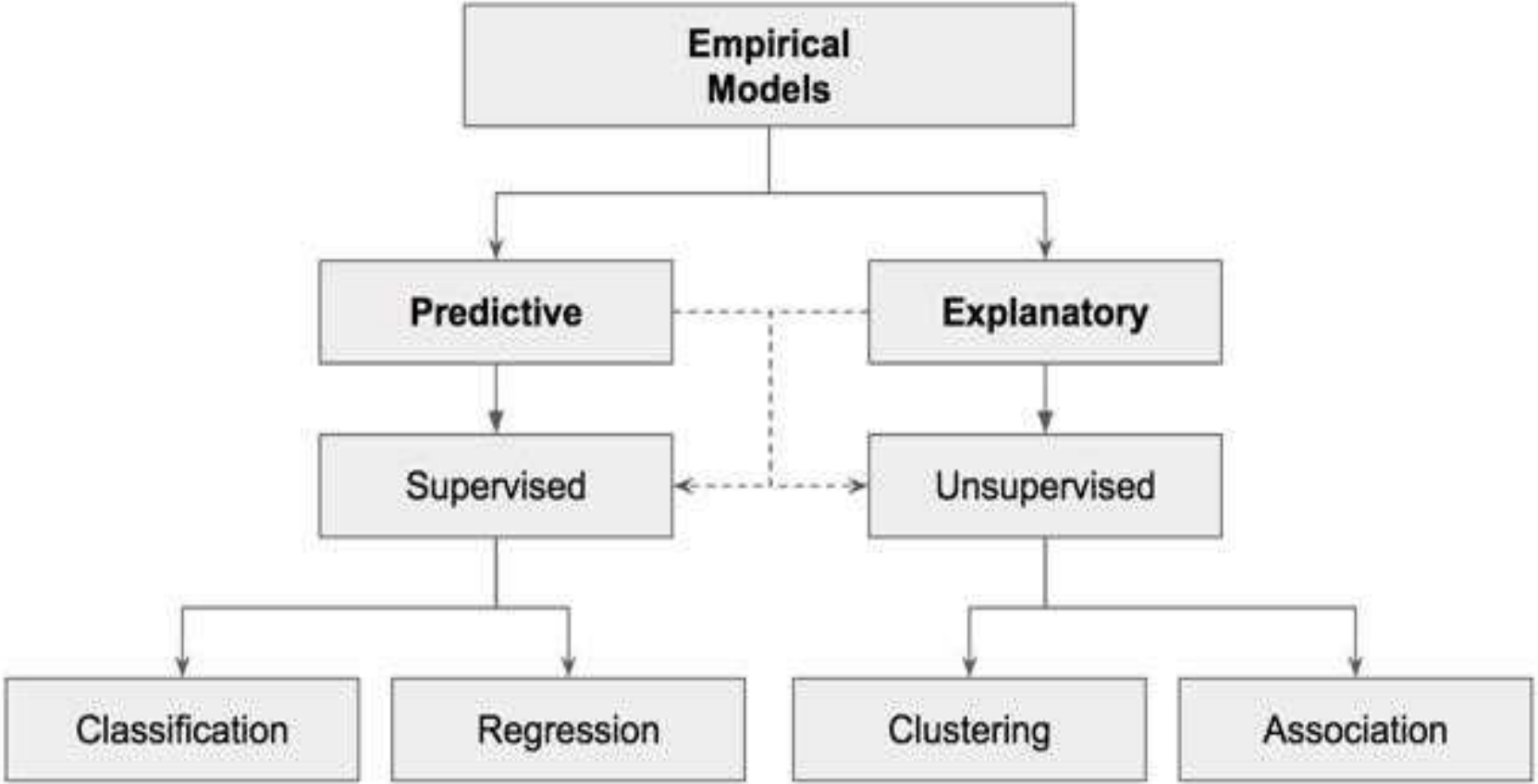
Regerssion

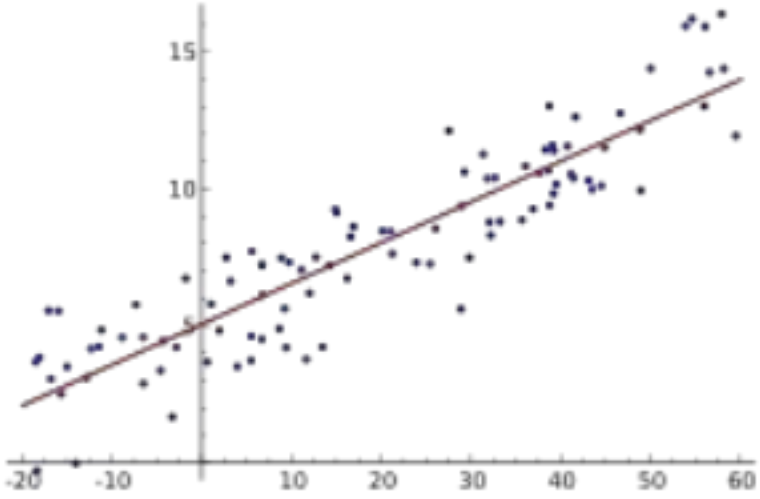

Empirical Models
Predictive
Explanatory
Supervised
Unsupervised
Classification
Regression
Clustering
Association

Linear regression

- $Y=a+bX$

$$a= \frac{(\sum Y)(\sum X^2) - (\sum X)(\sum XY)}{n(\sum x^2) - (\sum x)^2}$$

$$b= \frac{n(\sum XY) - (\sum X)(\sum Y)}{n(\sum x^2) - (\sum x)^2}$$



Solved Example

x	2	4	6	8
y	3	7	5	10

Eq. $y = 1.5 + 0.95x$

x	y	x^2	xy
2	3	4	6
4	7	16	28
6	5	36	30
8	10	64	80
$\Sigma x = 20$	$\Sigma y = 25$	$\Sigma x^2 = 120$	$\Sigma xy = 144$

a= $$\frac{(\sum Y)(\sum X^2) - (\sum X)(\sum XY)}{n(\sum x^2) - (\sum x)^2}$$

Now put the values in the equation

$$a = \frac{25 \times 120 - 20 \times 144}{4 \times 120 - 400}$$

a= $$\frac{120}{80}$$

a=1.5

b= $$\frac{n(\sum XY) - (\sum X)(\sum Y)}{n(\sum x^2) - (\sum x)^2}$$

Put the values in the equation

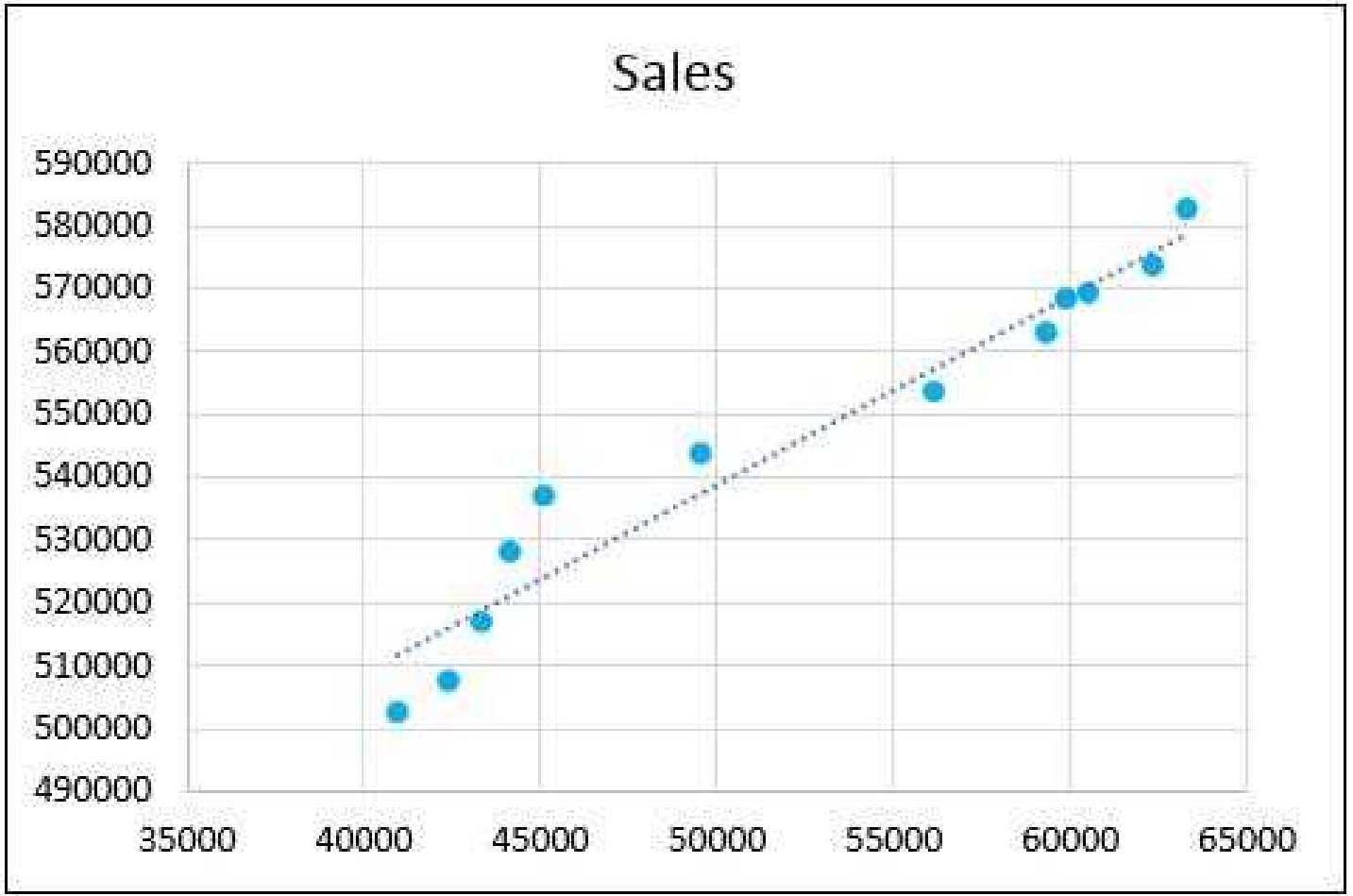
$$b = \frac{4 \times 144 - 20 \times 25}{4 \times 120 - 400}$$

b= $$\frac{76}{80}$$

b=0.95

Sales example

Month	Advertising	Sales
Jan	40937	502729
Feb	42376	507553
Mar	43355	516885
Apr	44126	528347
May	45060	537298
Jun	49546	544066
Jul	56105	553664
Aug	59322	563201
Sep	59877	568657
Oct	60481	569384
Nov	62356	573764
Dec	63246	582746


Sales
590000
580000
570000
560000
550000
540000
530000
520000
510000
500000
490000
35000
40000
45000
50000
55000
60000
65000

Online Store	Monthly E- Commerce Sales (in 1000 s)	Online Advertising Dollars (1000 s)
1	368	1.7
2	340	1.5
3	665	2.8
4	954	5
5	331	1.3
6	556	2.2
7	376	1.3

Example :

You have to study the relationship between the monthly e-commerce sales and the online advertising costs. You have the survey results for 7 online stores for the last year.

Your task is to find the equation of the straight line that fits the data best.

The table on the left represents the survey results from the 7 online stores.

We can see that there is a **positive relationship** between the monthly e-commerce sales (Y) and online advertising costs (X).

Y= 125.8 + 171.5*X

The regression line shows the predicted score on e-commerce sales for each possible value of the online advertising costs.

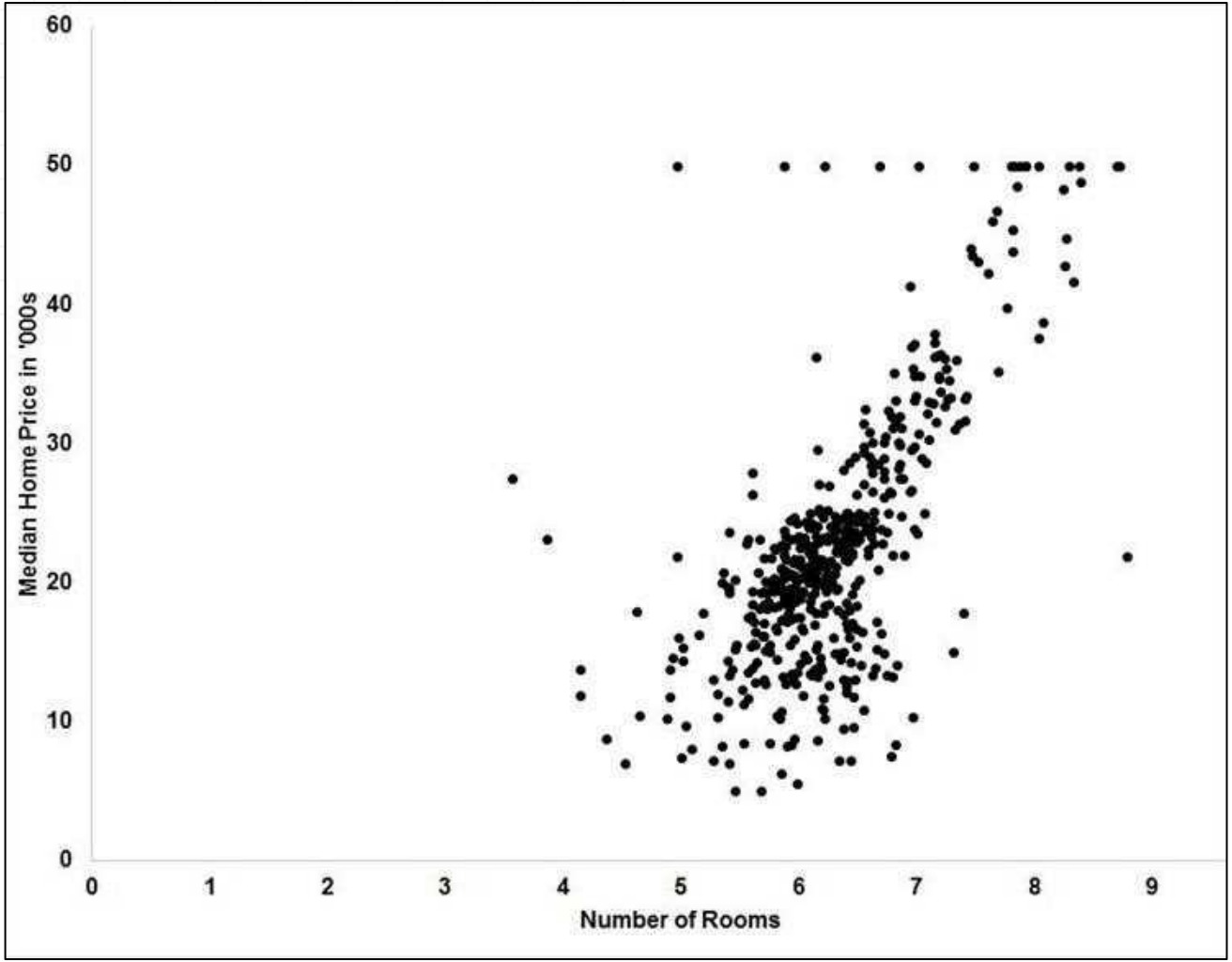
Scatter diagram
With The Regression Line:

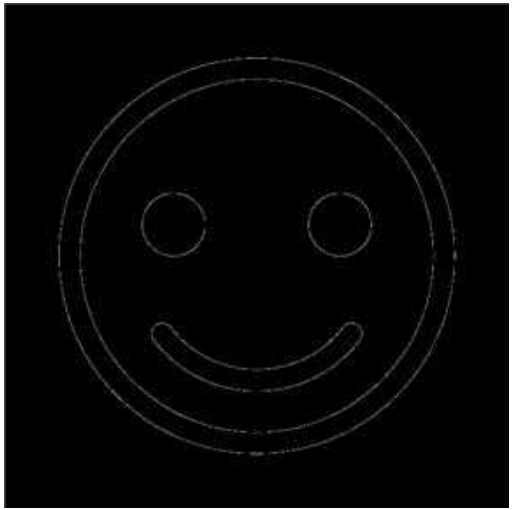
Note: *You can find easily the values for B0 and B1 with the help of paid or free statistical software, online linear regression calculators or Excel.*

Interpretation of the results:

The formula estimates that for each increase of 1 dollar in online advertising costs, the expected monthly e-commerce sales are predicted to increase by $171.5.

House prices example



LinearRegression

```
- 0.119 * CRIM
+ 0.049 * ZN
+ 2.444 * CHAS
- 16.784 * NOX
+ 3.646 * RM
- 1.579 * DIS
+ 0.292 * RAD
- 0.011 * TAX
- 0.916 * PTRATIO
+ 0.009 * B
- 0.556 * LSTAT
+ 37.258
```