Introduction to Linear Regression

Algorithm Theory - Part Two
OLS Equations

Linear Regression

- Linear Regression OLS Theory
  - We know the equation of a simple straight line:
    - $y = mx + b$
      - m is slope
      - b is intercept with y-axis

Linear Regression

- Linear Regression OLS Theory
  - We can see for **y=mx+b** there is only room for one possible feature x.
  - OLS will allow us to directly solve for the slope **m** and intercept **b.**
  - We will later see we'll need tools like gradient descent to scale this to multiple features.

Linear Regression

- Let's explore how we could translate a real data set into mathematical notation for linear regression.
- Then we'll solve a simple case of one feature to explore OLS in action.
- Afterwards we'll focus on gradient descent for real world data set situations.

Linear Regression

- Linear Regression allows us to build a relationship between multiple **features** to estimate a **target output**.

Area m^2	Bedrooms	Bathrooms	Price
200	3	2	$500,000
190	2	1	$450,000
230	3	3	$650,000
180	1	1	$400,000
210	2	2	$550,000

Linear Regression

- We can translate this data into generalized mathematical notation...

X			y
Area m^2	Bedrooms	Bathrooms	Price
200	3	2	$500,000
190	2	1	$450,000
230	3	3	$650,000
180	1	1	$400,000
210	2	2	$550,000

Linear Regression

- We can translate this data into generalized mathematical notation...

X			y
x_1	x_2	x_3	y
200	3	2	$500,000
190	2	1	$450,000
230	3	3	$650,000
180	1	1	$400,000
210	2	2	$550,000

Linear Regression

- We can translate this data into generalized mathematical notation...

X			y
x_1	x_2	x_3	y
x^1_1	x^1_1	x^1_1	y_1
x^2_1	x^2_1	x^2_1	y_2
x^3_1	x^3_1	x^3_1	y_3
x^4_1	x^4_1	x^4_1	y_4
x^5_1	x^5_1	x^5_1	y_5

Linear Regression

- Now let's build out a linear relationship between the features X and label y.

X			y
x_1	x_2	x_3	y
$x^1{}_1$	$x^1{}_1$	$x^1{}_1$	y_1
$x^2{}_1$	$x^2{}_1$	$x^2{}_1$	y_2
$x^3{}_1$	$x^3{}_1$	$x^3{}_1$	y_3
$x^4{}_1$	$x^4{}_1$	$x^4{}_1$	y_4
$x^5{}_1$	$x^5{}_1$	$x^5{}_1$	y_5

Linear Regression

- Now let's build out a linear relationship between the features X and label y.

X			y
x_1	x_2	x_3	y

Linear Regression

- Reformat for **y = x** equation

y	X		
y	x_1	x_2	x_3

Linear Regression

- Each feature should have some Beta coefficient associated with it.

y	X		
y	x_1	x_2	x_3

$$\hat{y} = \beta_0 x_0 + \cdots + \beta_n x_n$$

Linear Regression

- This is the same as the common notation for a simple line: **y=mx+b**

y	X		
y	x_1	x_2	x_3

$$\hat{y} = \beta_0 x_0 + \cdots + \beta_n x_n$$

Linear Regression

- This is stating there is some Beta coefficient for each feature to minimize error.

y	X		
y	x_1	x_2	x_3

$$\hat{y} = \beta_0 x_0 + \cdots + \beta_n x_n$$

Linear Regression

- We can also express this equation as a sum:

y	X		
y	x_1	x_2	x_3

$$\hat{y} = \beta_0 x_0 + \cdots + \beta_n x_n$$

$$\hat{y} = \sum_{i=0}^{n} \beta_i x_i$$

Linear Regression

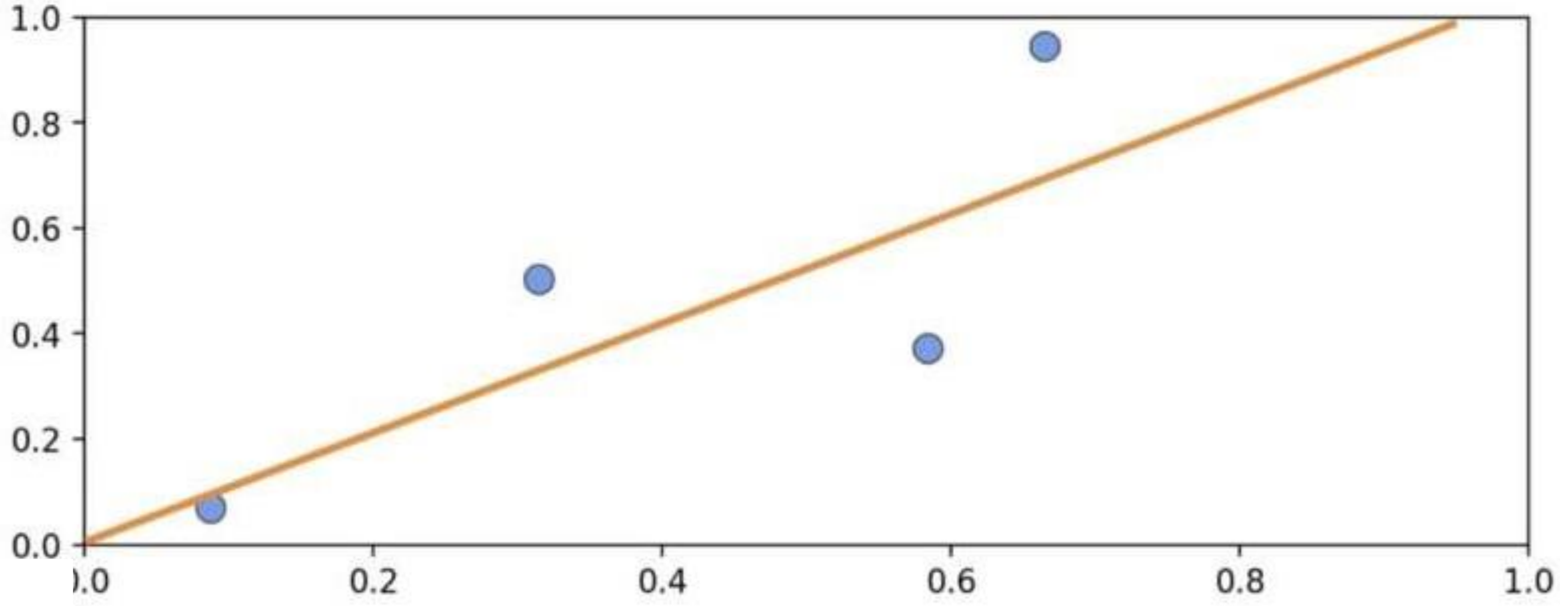
- Note the y hat symbol displays a prediction. There is usually no set of Betas to create a perfect fit to y!

$$\hat{y} = \sum_{i=0}^{n} \beta_i x_i$$

Linear Regression

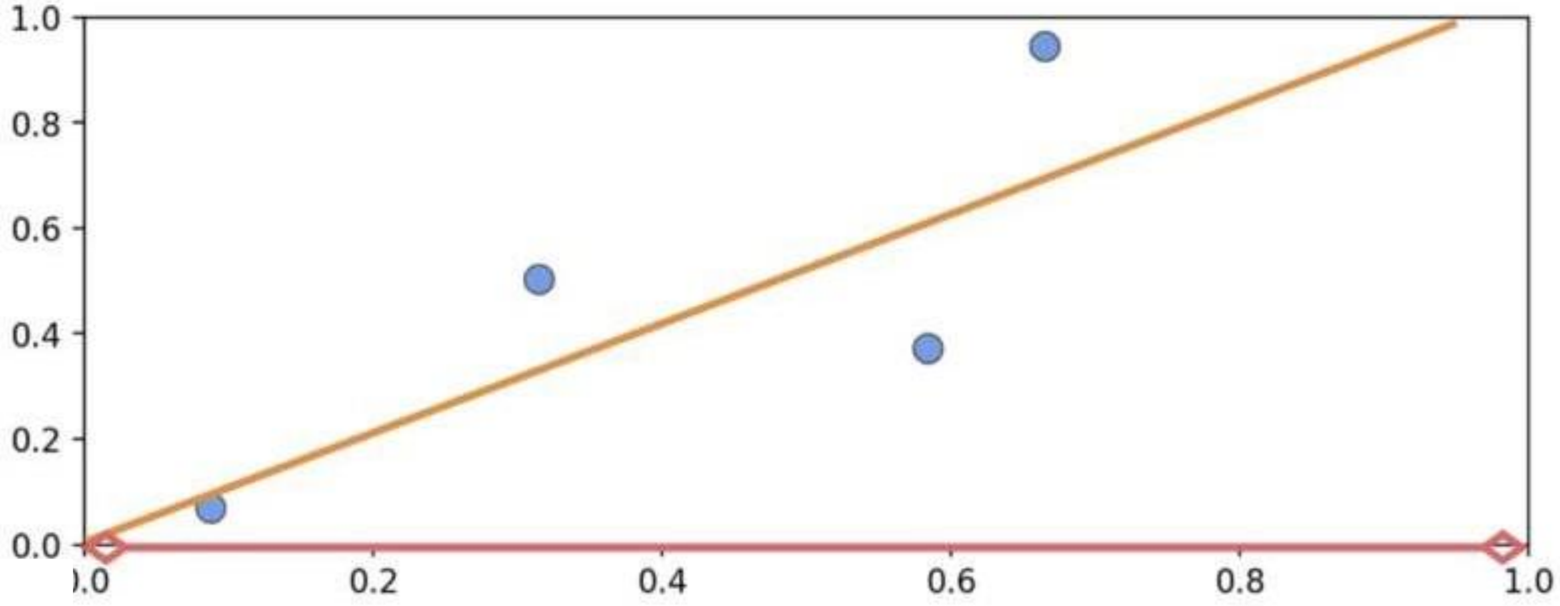
- Line equation:

$$\hat{y} = \sum_{i=0}^{n} \beta_i x_i$$



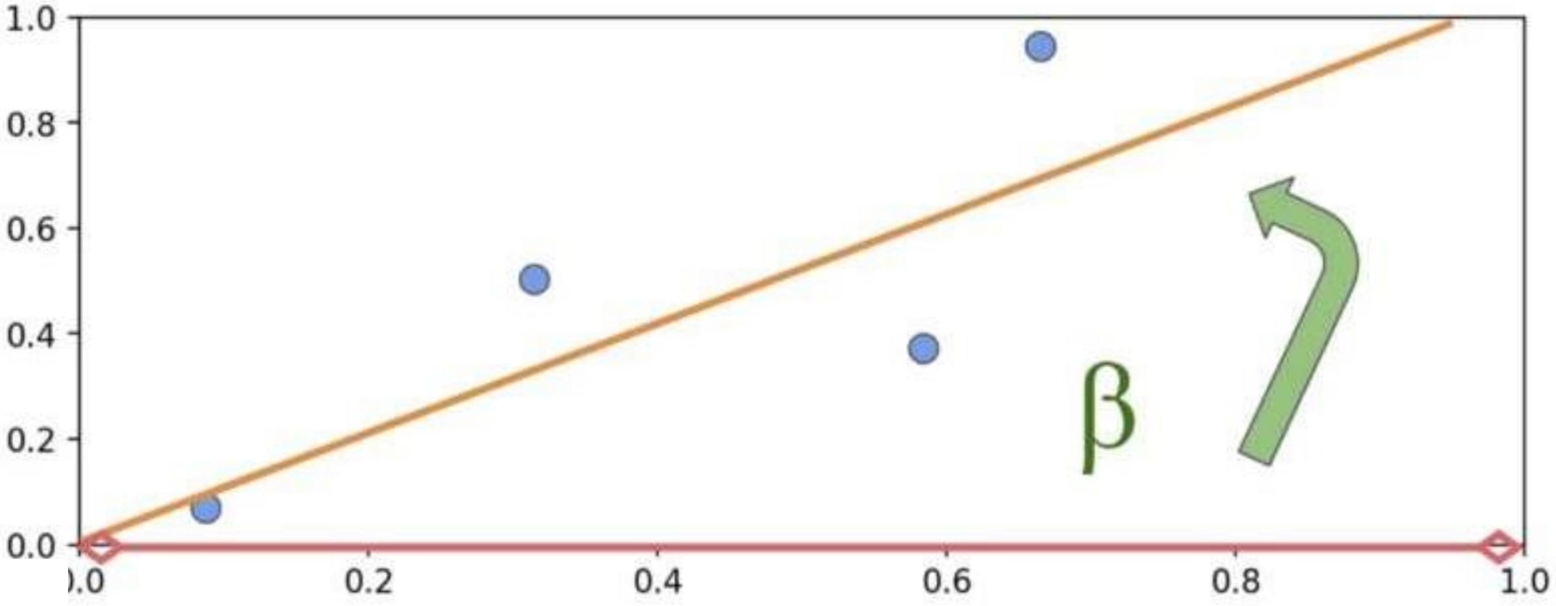
Linear Regression

$$\hat{y} = \sum_{i=0}^{n} \beta_i \boxed{x_i}$$



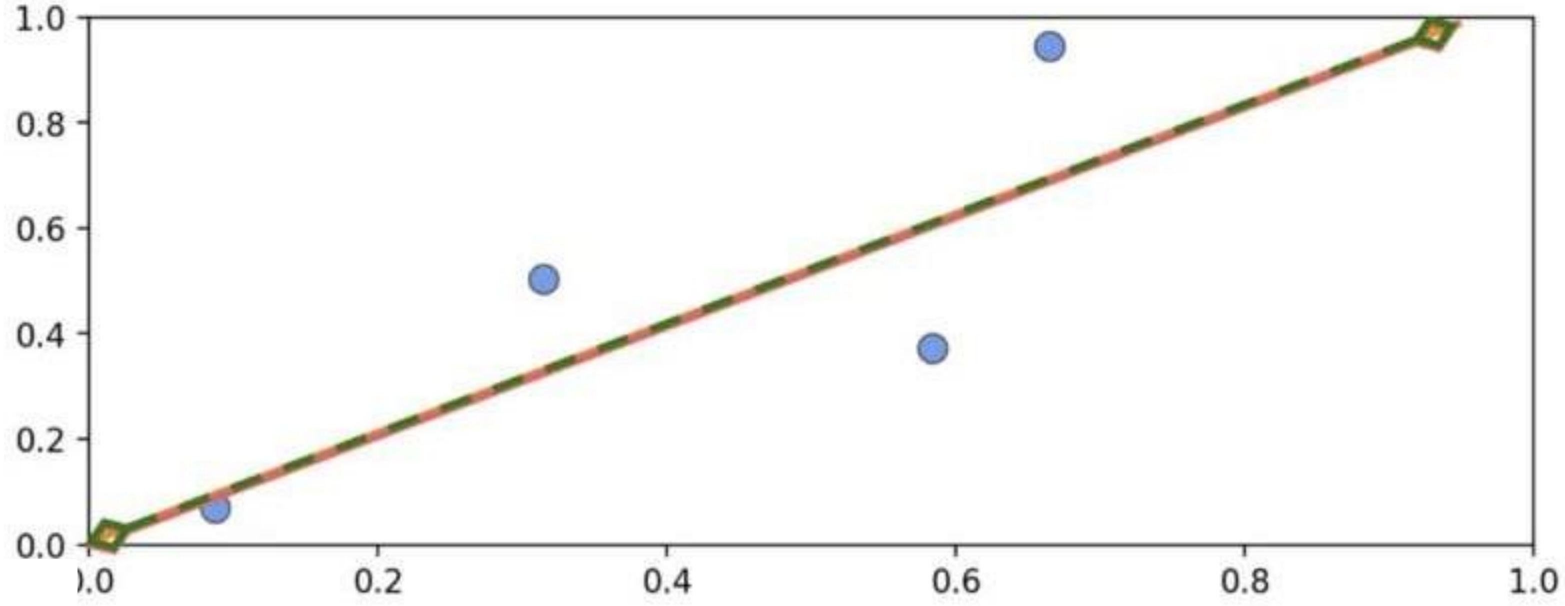
Linear Regression

$$\hat{y} = \sum_{i=0}^{n} \boxed{\beta_i}\boxed{x_i}$$



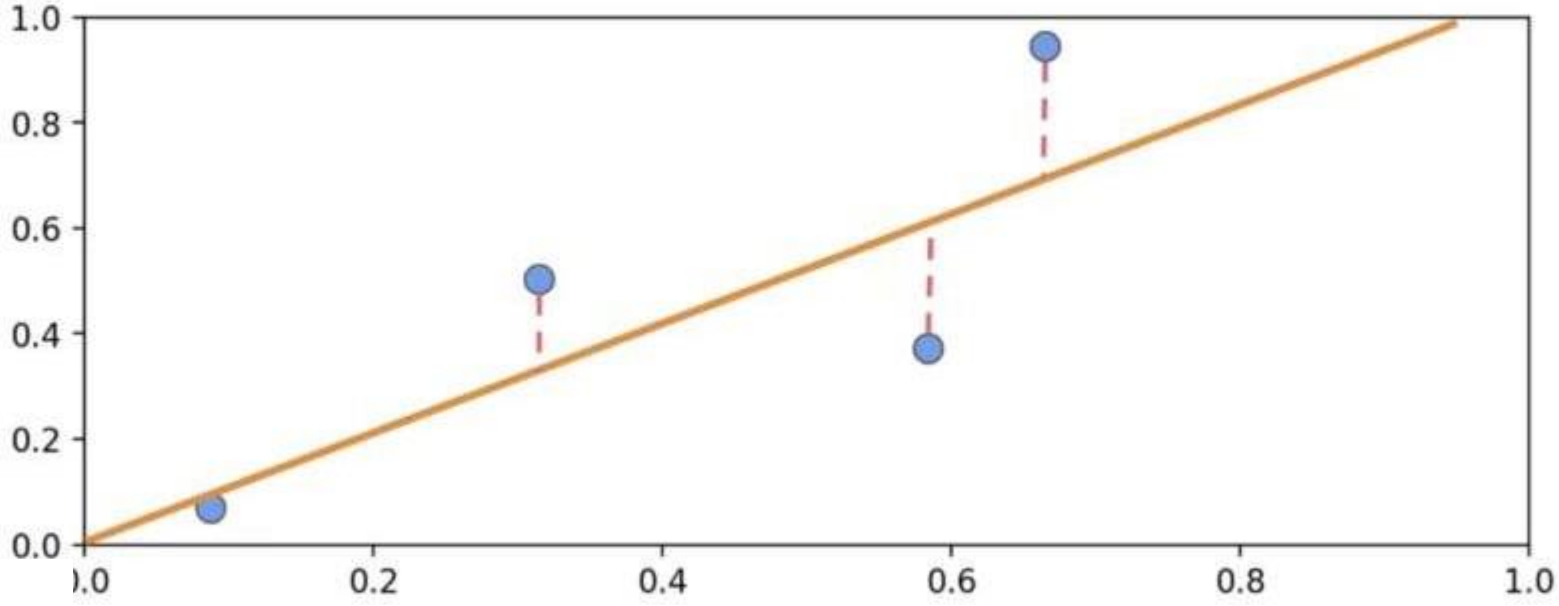
Linear Regression

$$\hat{y} = \sum_{i=0}^{n} \boxed{\beta_i}\boxed{x_i}$$



Linear Regression

$$\boxed{\hat{y}} = \sum_{i=0}^{n} \beta_i x_i$$



Linear Regression

- For simple problems with one X feature we can easily solve for Betas values with an analytical solution.
- Let's quickly solve a simple example problem, then later we will see that for multiple features we will need gradient descent.

Linear Regression

- Recall that the equation of a line follows the form $y = mx + b$ where

Linear Regression

- Recall that the equation of a line follows the form $y = mx + b$ where
  - m is the **slope** of the line, and

Linear Regression

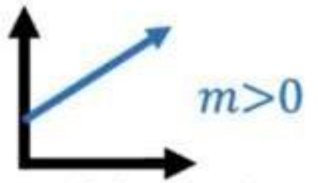
- Recall that the equation of a line follows the form $y = mx + b$ where

  m is the **slope** of the line, and

  b is where the line crosses the y-axis when x=0 (b is the **y-intercept**)

Linear Regression

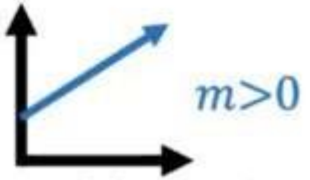
- Recall that the equation of a line follows the form $y = mx + b$ where
  - m is the **slope** of the line, and
  - b is where the line crosses the y-axis when x=0 (b is the **y-intercept**)



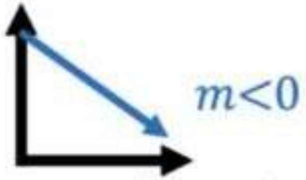
positive slope

Linear Regression

- Recall that the equation of a line follows the form $y = mx + b$ where
  m is the **slope** of the line, and
  b is where the line crosses the y-axis when x=0 (b is the **y-intercept**)



positive slope



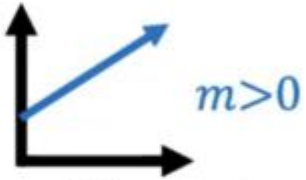
negative slope

Linear Regression

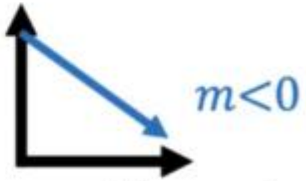
- Recall that the equation of a line follows the form $y = mx + b$ where

  m is the **slope** of the line, and

  b is where the line crosses the y-axis when x=0 (b is the **y-intercept**)



positive slope



negative slope

zero slope

Linear Regression

- In a linear regression, where we try to formulate the relationship between variables, $y = mx + b$ becomes

Linear Regression

- In a linear regression, where we try to formulate the relationship between variables, $y = mx + b$ becomes

$$\hat{y} = b_0 + b_1 x$$

- Our goal is to predict the value of a **dependent variable** (y) based on that of an **independent variable** (x).

Linear Regression

$$\hat{y} = b_0 + b_1 x$$

رواد مصر الرقمية

وزارة الاتصالات وتكنولوجيا المعلومات

- How to derive b_1 and b_0:

Linear Regression

$$\hat{y} = b_0 + b_1 x$$

وزارة الاتصـــــالات
وتكنولوجيا المعلومات

رواد مصر الرقمية

- How to derive b_1 and b_0:

$$b_1 = \rho_{x,y} \frac{\sigma_y}{\sigma_x}$$

$\rho_{x,y} = Pearson\ Correlation\ Coefficient$
$\sigma_x, \sigma_y = Standard\ Deviations$

https://en.wikipedia.org/wiki/Simple_linear_regression

Linear Regression

$$\hat{y} = b_0 + b_1 x$$

- How to derive b_1 and b_0:

$$b_1 = \rho_{x,y} \frac{\sigma_y}{\sigma_x}$$

$\rho_{x,y} = Pearson\ Correlation\ Coefficient$
$\sigma_x, \sigma_y = Standard\ Deviations$

$$= \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\sqrt{\Sigma(x - \bar{x})^2}\sqrt{\Sigma(y - \bar{y})^2}} \cdot \frac{\sqrt{\frac{\Sigma(y - \bar{y})^2}{n}}}{\sqrt{\frac{\Sigma(x - \bar{x})^2}{n}}} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2}$$

Linear Regression

$$\hat{y} = b_0 + b_1 x$$

- How to derive b_1 and b_0:

$$b_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

Linear Regression

$$\hat{y} = b_0 + b_1 x$$

- How to derive b_1 and b_0:

$$b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

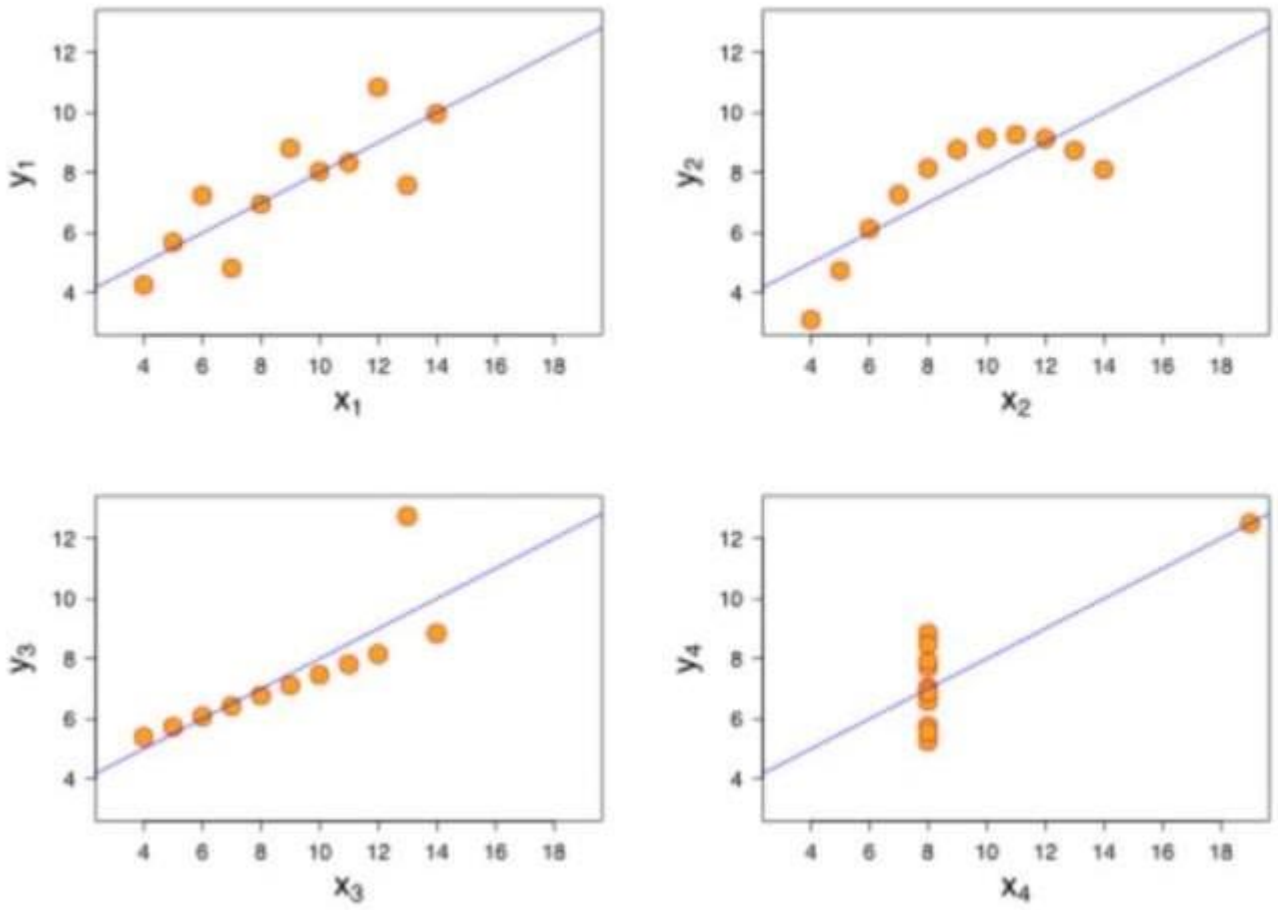
$$b_0 = \bar{y} - b_1 \bar{x}$$

Limitations of Linear Regression

Anscombe's Quartet illustrates the pitfalls of relying on pure calculation.

Limitations of Linear Regression

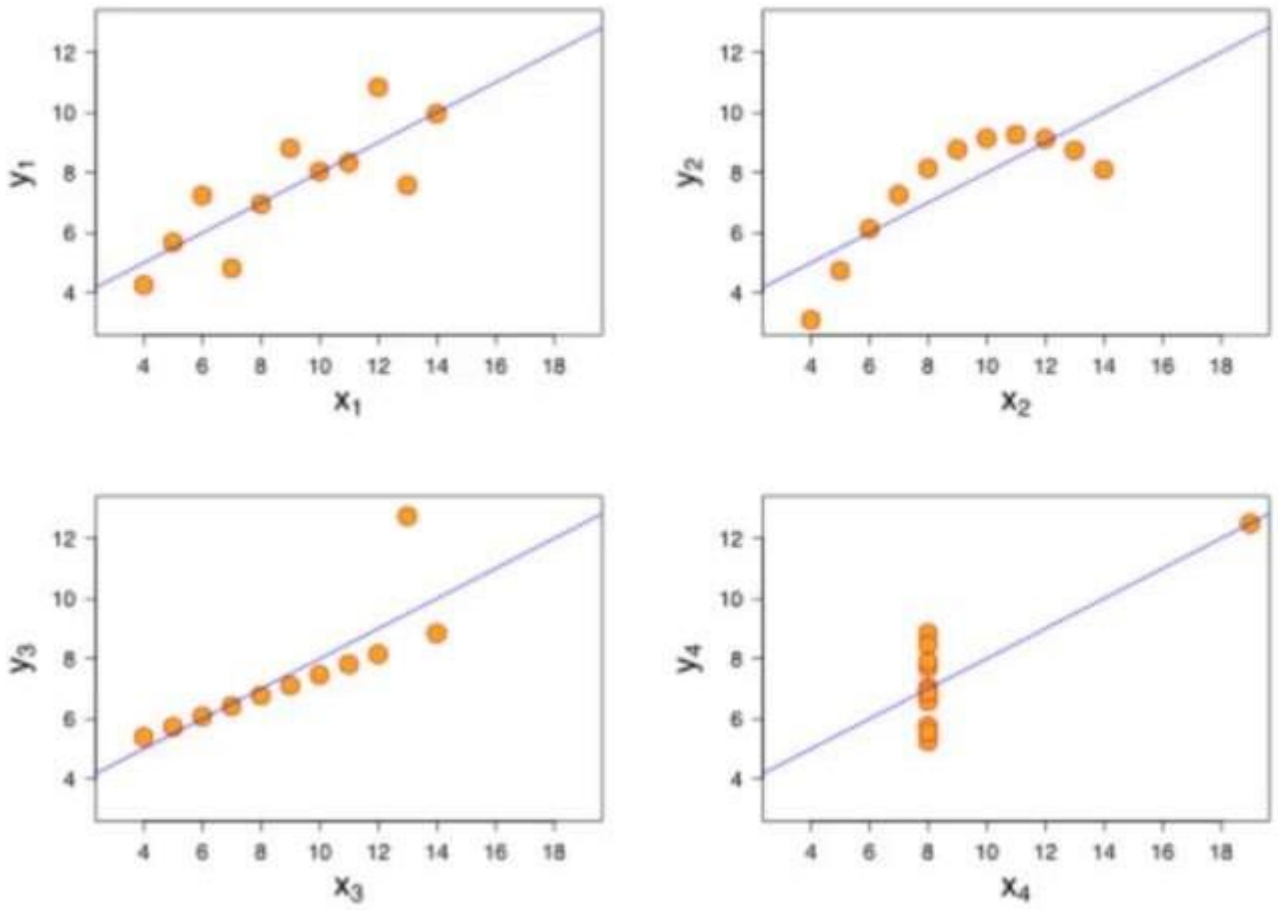
Anscombe's Quartet illustrates the pitfalls of relying on pure calculation.



Limitations of Linear Regression

Anscombe's Quartet illustrates the pitfalls of relying on pure calculation.

Each graph results in the same calculated regression line.



Linear Regression

- A manager wants to find the relationship between the number of hours that a plant is operational in a week and weekly production.

Linear Regression

- Here the **independent variable** x is hours of operation, and the **dependent variable** y is production volume.

Linear Regression

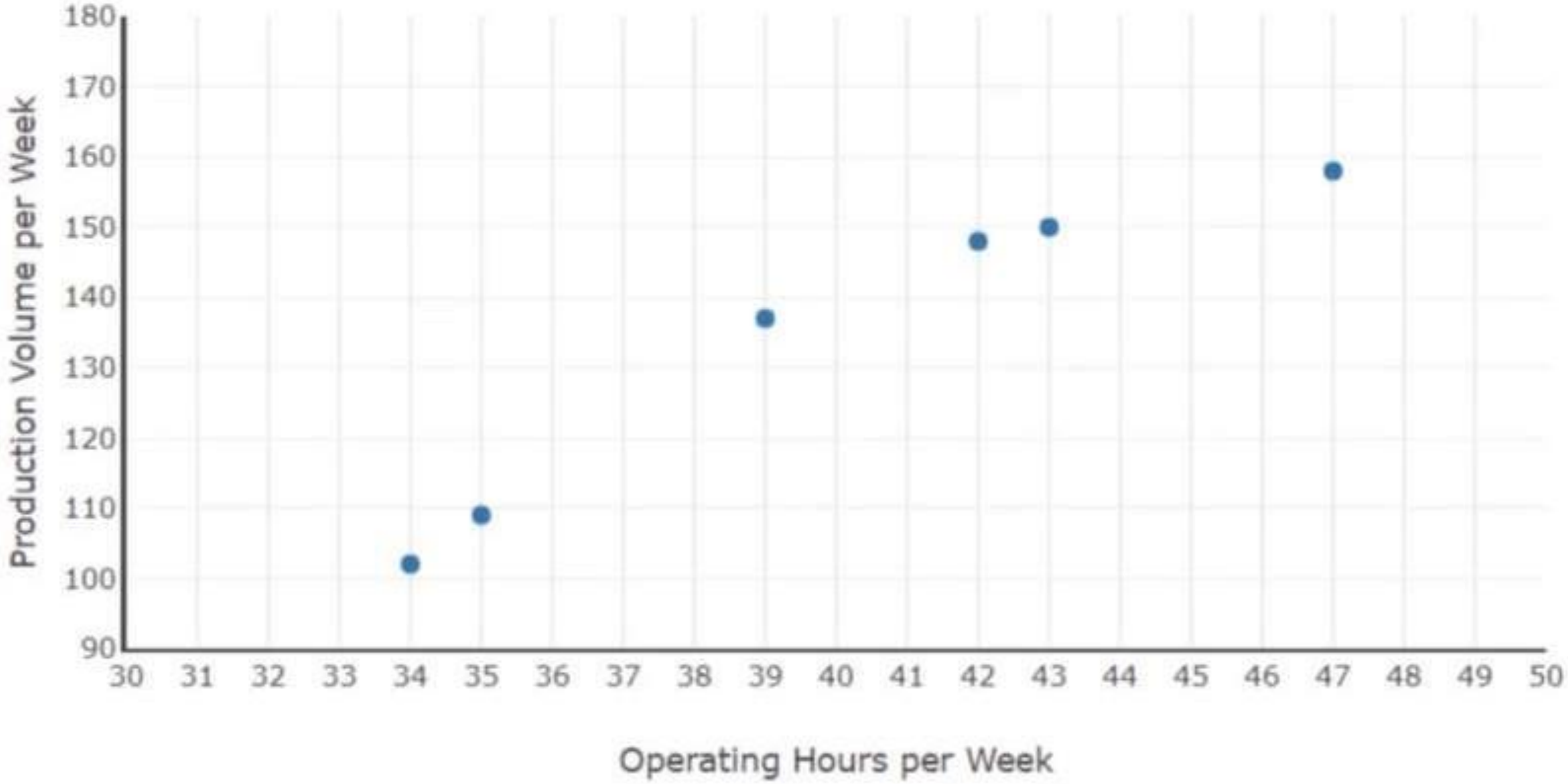
- The manager develops the following table:

Production Hours (x)	Production Volume (y)
34	102
35	109
39	137
42	148
43	150
47	158

Linear Regression

- First, plot the data

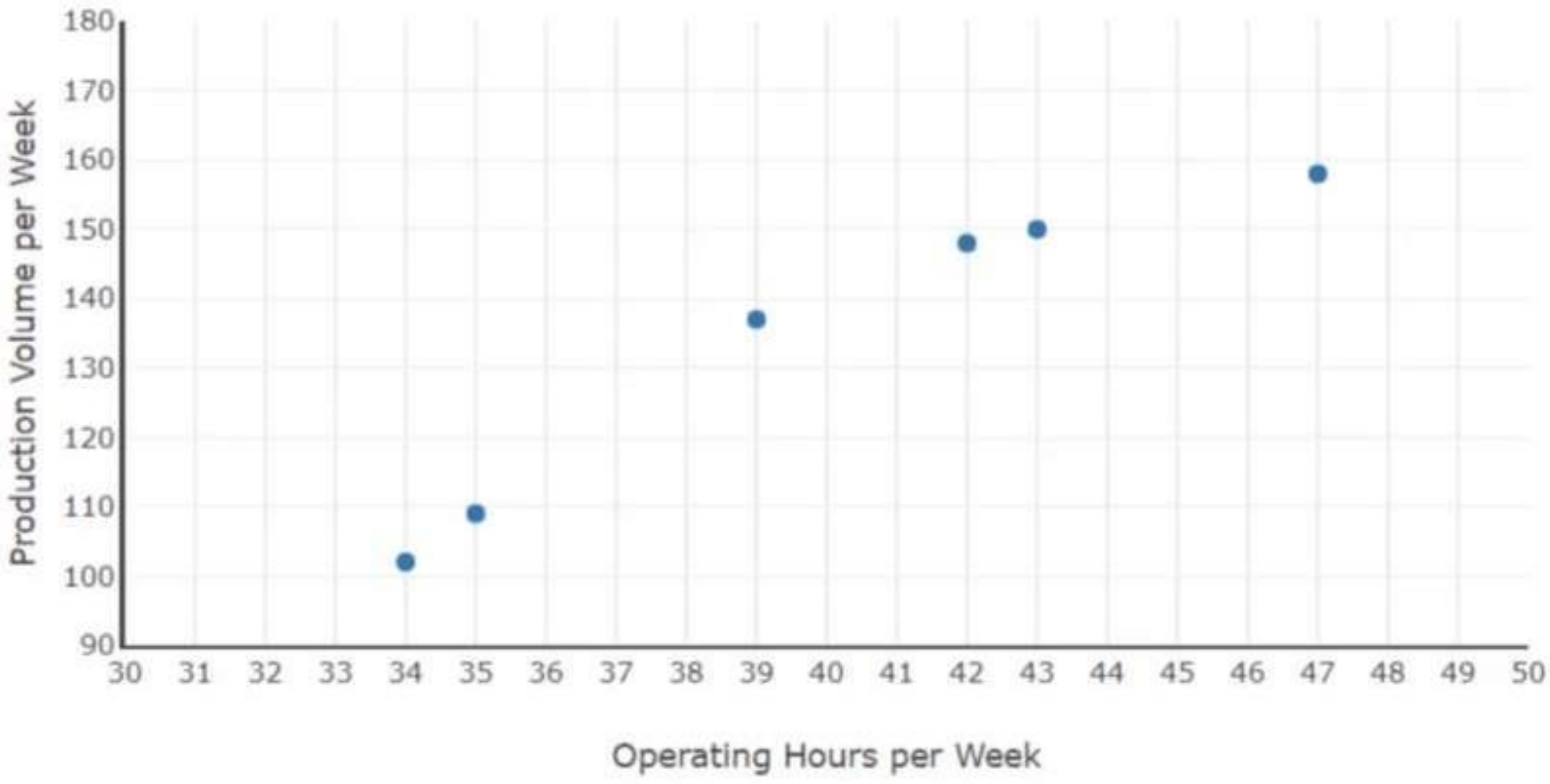
Production Hours (x)	Production Volume (y)
34	102
35	109
39	137
42	148
43	150
47	158



Linear Regression

- First, plot the data   Is there a linear pattern?

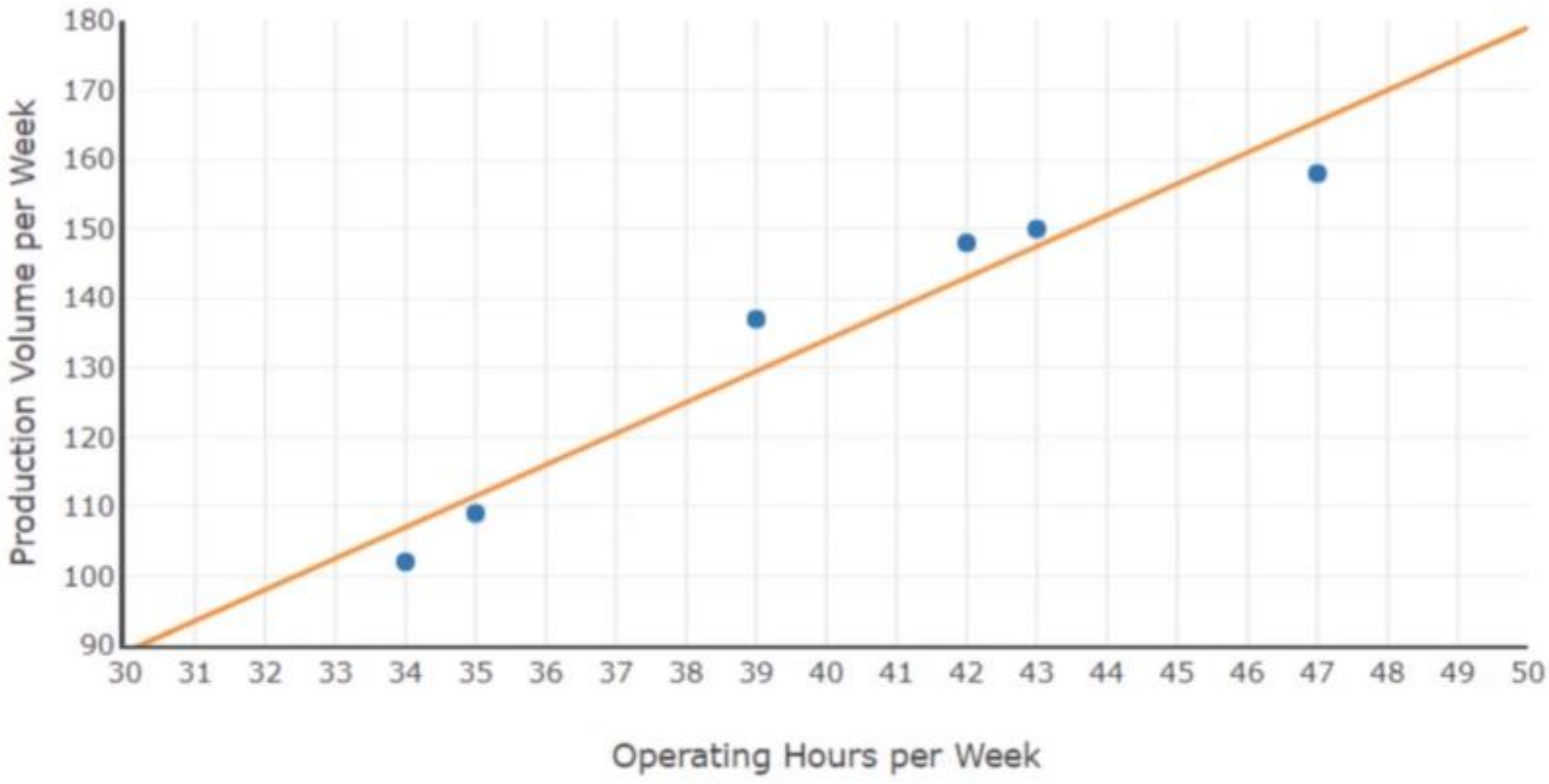
Production Hours (x)	Production Volume (y)
34	102
35	109
39	137
42	148
43	150
47	158



Linear Regression

- First, plot the data   Is there a linear pattern?

Production Hours (x)	Production Volume (y)
34	102
35	109
39	137
42	148
43	150
47	158



Linear Regression

$$\hat{y} = b_0 + b_1 x$$

$$b_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

$$b_0 = \bar{y} - b_1\bar{x}$$

- Run calculations:

Production Hours (x)	Production Volume (y)
34	102
35	109
39	137
42	148
43	150
47	158

Linear Regression

$$\hat{y} = b_0 + b_1 x$$

$$b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

- Run calculations:

	Production Hours (x)	Production Volume (y)
	34	102
	35	109
	39	137
	42	148
	43	150
	47	158
$\bar{x}, \bar{y}$	**40**	**134**

رواد مصر الرقمية

Linear Regression

وزارة الاتصالات
وتكنولوجيا المعلومات

$$\hat{y} = b_0 + b_1 x$$

$$b_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

- Run calculations:

	Production Hours (x)	Production Volume (y)	$(x-\bar{x})$	$(y-\bar{y})$
	34	102	-6	-32
	35	109	-5	-25
	39	137	-1	3
	42	148	2	14
	43	150	3	16
	47	158	7	24
$\bar{x}, \bar{y}$	**40**	**134**		

رواد مصر الرقمية

Linear Regression

وزارة الاتصالات
وتكنولوجيا المعلومات

$$\hat{y} = b_0 + b_1 x$$

$$b_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

$$b_0 = \bar{y} - b_1\bar{x}$$

- Run calculations:

	Production Hours (x)	Production Volume (y)	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})(y-\bar{y})$
	34	102	-6	-32	192
	35	109	-5	-25	125
	39	137	-1	3	-3
	42	148	2	14	28
	43	150	3	16	48
	47	158	7	24	168
$\bar{x}, \bar{y}$	40	134			

رواد مصر الرقمية

وزارة الاتصالات وتكنولوجيا المعلومات

Linear Regression

$$\hat{y} = b_0 + b_1 x$$

$$b_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

- Run calculations:

	Production Hours (x)	Production Volume (y)	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$
	34	102	-6	-32	192	36
	35	109	-5	-25	125	25
	39	137	-1	3	-3	1
	42	148	2	14	28	4
	43	150	3	16	48	9
	47	158	7	24	168	49
$\bar{x}, \bar{y}$	40	134				

Linear Regression

$$\hat{y} = b_0 + b_1 x$$

$$b_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

- Run calculations:

	Production Hours (x)	Production Volume (y)	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$
	34	102	-6	-32	192	36
	35	109	-5	-25	125	25
	39	137	-1	3	-3	1
	42	148	2	14	28	4
	43	150	3	16	48	9
	47	158	7	24	168	49
$\bar{x}, \bar{y}$	40	134		Sum:	558	124

رواد مصر الرقمية

Linear Regression

$$\hat{y} = b_0 + b_1 x$$

$$b_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

وزارة الاتصـــــــالات
وتكنولوجيا المعلومات

- Run calculations:

	Production Hours (x)	Production Volume (y)	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$
	34	102	-6	-32	192	36
	35	109	-5	-25	125	25
	39	137	-1	3	-3	1
	42	148	2	14	28	4
	43	150	3	16	48	9
	47	158	7	24	168	49
$\bar{x}, \bar{y}$	**40**	**134**		Sum:	**558**	**124**
					$\Sigma(x-\bar{x})(y-\bar{y})$	$\Sigma(x-\bar{x})^2$

Linear Regression

$$\hat{y} = b_0 + b_1 x$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

- Run calculations:

	Production Hours (x)	Production Volume (y)
	34	102
	35	109
	39	137
	42	148
	43	150
	47	158
$\bar{x}, \bar{y}$	40	134

$$b_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

Sum:	558	124
	$\Sigma(x-\bar{x})(y-\bar{y})$	$\Sigma(x-\bar{x})^2$

Linear Regression

$$\hat{y} = b_0 + b_1 x$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

- Run calculations:

	Production Hours (x)	Production Volume (y)
	34	102
	35	109
	39	137
	42	148
	43	150
	47	158
$\bar{x}, \bar{y}$	**40**	**134**

$$b_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{558}{124} = \mathbf{4.5}$$

Sum:	**558**	**124**
	$\Sigma(x-\bar{x})(y-\bar{y})$	$\Sigma(x-\bar{x})^2$

Linear Regression

$$\hat{y} = b_0 + b_1 x$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

- Run calculations:

	Production Hours (x)	Production Volume (y)
	34	102
	35	109
	39	137
	42	148
	43	150
	47	158
$\bar{x}, \bar{y}$	40	134

$$b_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{558}{124} = \mathbf{4.5}$$

$$b_0 = \bar{y} - b_1 \bar{x} = 134 - (4.5 \times 40) = \mathbf{-46}$$

Sum:	558	124
	$\Sigma(x-\bar{x})(y-\bar{y})$	$\Sigma(x-\bar{x})^2$

وزارة الاتصـــــــالات وتكنولوجيا المعلومات

رواد مصر الرقمية

Linear Regression

$$\hat{y} = b_0 + b_1 x$$

- Run calculations:

	Production Hours (x)	Production Volume (y)
	34	102
	35	109
	39	137
	42	148
	43	150
	47	158
$\bar{x}, \bar{y}$	40	134

$$b_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{558}{124} = \mathbf{4.5}$$

$$b_0 = \bar{y} - b_1\bar{x} = 134 - (4.5 \times 40) = \mathbf{-46}$$

$$\hat{y} = -46 + 4.5x$$

Sum:	558	124
	$\Sigma(x-\bar{x})(y-\bar{y})$	$\Sigma(x-\bar{x})^2$

رواد مصر الرقمية

وزارة الاتصالات وتكنولوجيا المعلومات

Linear Regression

Based on the formula, if the manager wants to produce 125 units per week, the plant should run for:

Production Hours (x)	Production Volume (y)
34	102
35	109
39	137
42	148
43	150
47	158

Linear Regression

Based on the formula, if the manager wants to produce 125 units per week, the plant should run for:

Production Hours (x)	Production Volume (y)
34	102
35	109
39	137
42	148
43	150
47	158

$$\hat{y} = b_0 + b_1 x$$

$$125 = -46 + 4.5x$$

$$x = \frac{171}{4.5} = \mathbf{38}\ hours\ per\ week$$

Linear Regression

- As we expand to more than a single feature however, an analytical solution quickly becomes unscalable.
- Instead we shift focus on **minimizing** a **cost function** with **gradient descent.**

Linear Regression

- We can use **gradient descent** to solve a **cost function** to calculate Beta values!

$$\hat{y} = \sum_{i=0}^{n} \beta_i x_i$$

Linear Regression

- We'll work on developing a cost function to minimize in the next lectures!

$$\hat{y} = \sum_{i=0}^{n} \beta_i x_i$$