Section 3.2. Linear Functions

1. Linear functions and their graphs.
2. Linear regression.

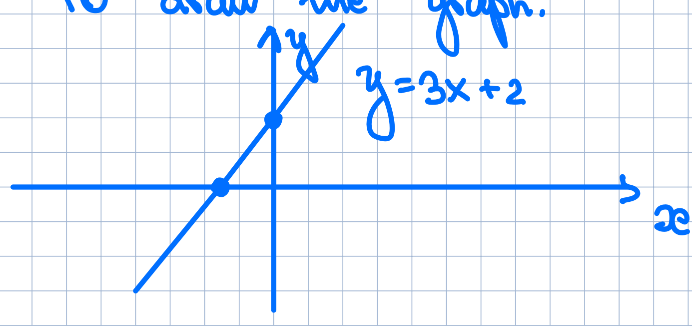
1.

Def. A linear function $f(x)$ is any function that can be written in the form $f(x) = mx + b$, where $m, b \in \mathbb{R}$.
If $m \neq 0$, $f(x) = mx + b$ is also called a first-degree polynomial function.

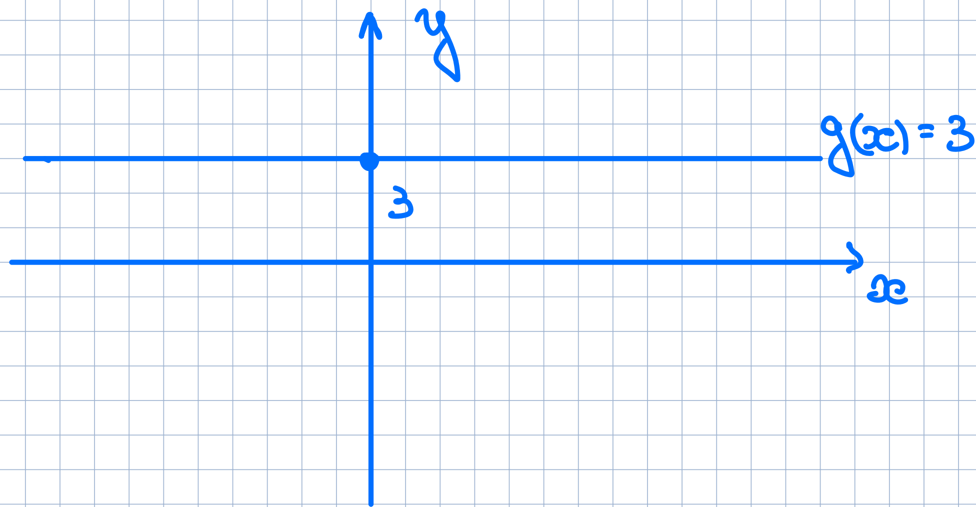
Example

- $f(x) = 3x + 2$

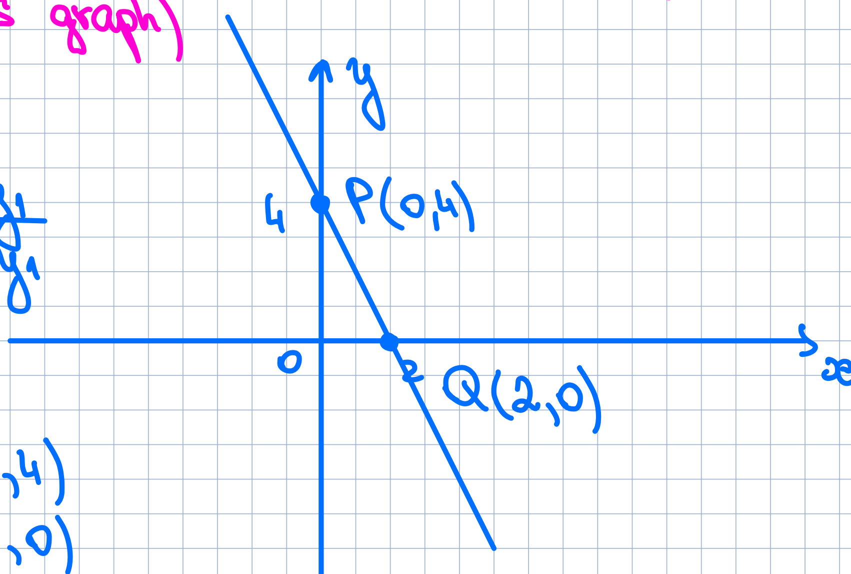
Graphing: we need at least <u>two</u> points to draw the graph.



- $g(x) = 3$ — horizontal line



Example (Finding a linear function given its graph)



l:

$$\frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1}$$

$(x_1, y_1) = (0,4)$
$(x_2, y_2) = (2,0)$

Hence, $\dfrac{x - 0}{2} = \dfrac{y - 4}{-4} \Rightarrow \dfrac{x}{2} = \dfrac{y-4}{-4}$

$$2(y-4) = -4x \Rightarrow \boxed{y = -2x + 4}$$

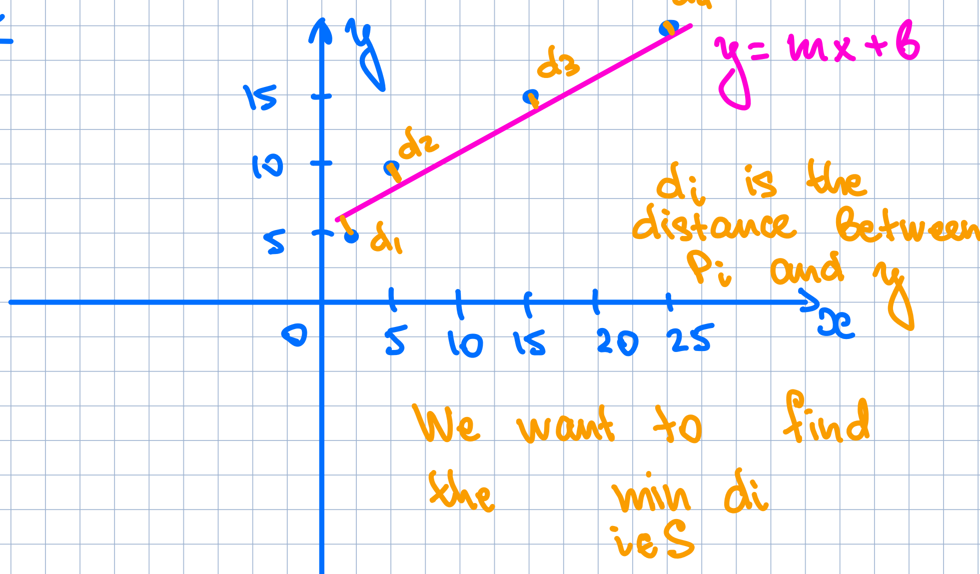
2.

Goal:

given: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$

find: $y = mx + b$ such that its graph comes closest to "fitting" the points

<u>Example</u>



x (weeks after opening)	y (number of clients)
1	14
3	18
6	21
10	25
14	28
18	30
22	35

Linear regression: find the line that most closely fits given data

Least-squares method: minimize the sum of the squares of the deviations between the line and the actual data points.

Step 1

$\bar{x}$ - ?

$\bar{y}$ - ?

$$\bar{x} = \frac{1+3+6+10+14+18+22}{7} \approx$$

$$\approx 10.571429$$

$$\bar{y} \approx 24.428571$$

Step 2

Δx - ?

Δy - ?

$$\Delta x = x - \bar{x}$$

$$\Delta y = y - \bar{y}$$

Δx	Δy
-9.571429	-10.428571
-7.571429	-6.428571
-4.571429	-3.428571
-0.571429	0.571429
3.428571	3.517429
7.428571	5.571429
11.428571	10.571429

Step 3

$$\sum \Delta x \cdot \Delta y = 338.285715$$

$$\sum (\Delta x)^2 = 367.714285$$

Step 4

$$m = \frac{\sum \Delta x \Delta y}{\sum (\Delta x)^2} \approx 0.920$$

$$b = \bar{y} - m\bar{x} \approx 14.7.$$

Therefore,

$$\boxed{y = 0.920x + 14.7}$$

Correlation: with given (x_1,y_1), $(x_2,y_2),\dots,(x_n,y_n)$ find if the collection shows linear dependence of y on x.

We need to calulate the Pearson correlation coefficient r:

$$r = \frac{\sum \Delta x \Delta y}{\sqrt{\sum (\Delta x)^2}\sqrt{\sum (\Delta y)^2}}$$

$$-1 \leq r \leq 1$$

- $r = 0$: no linear dependence
- $|r| \approx 1$: strong linear dependence
- $r > 0$: positive slope in the line of best fit

$r < 0$: negative slope in the line of best fit

Due to our example we have

$r \approx 0.990$. (Strong positive

linear dependence of y on x, we have the line which predicts the future growth very well).