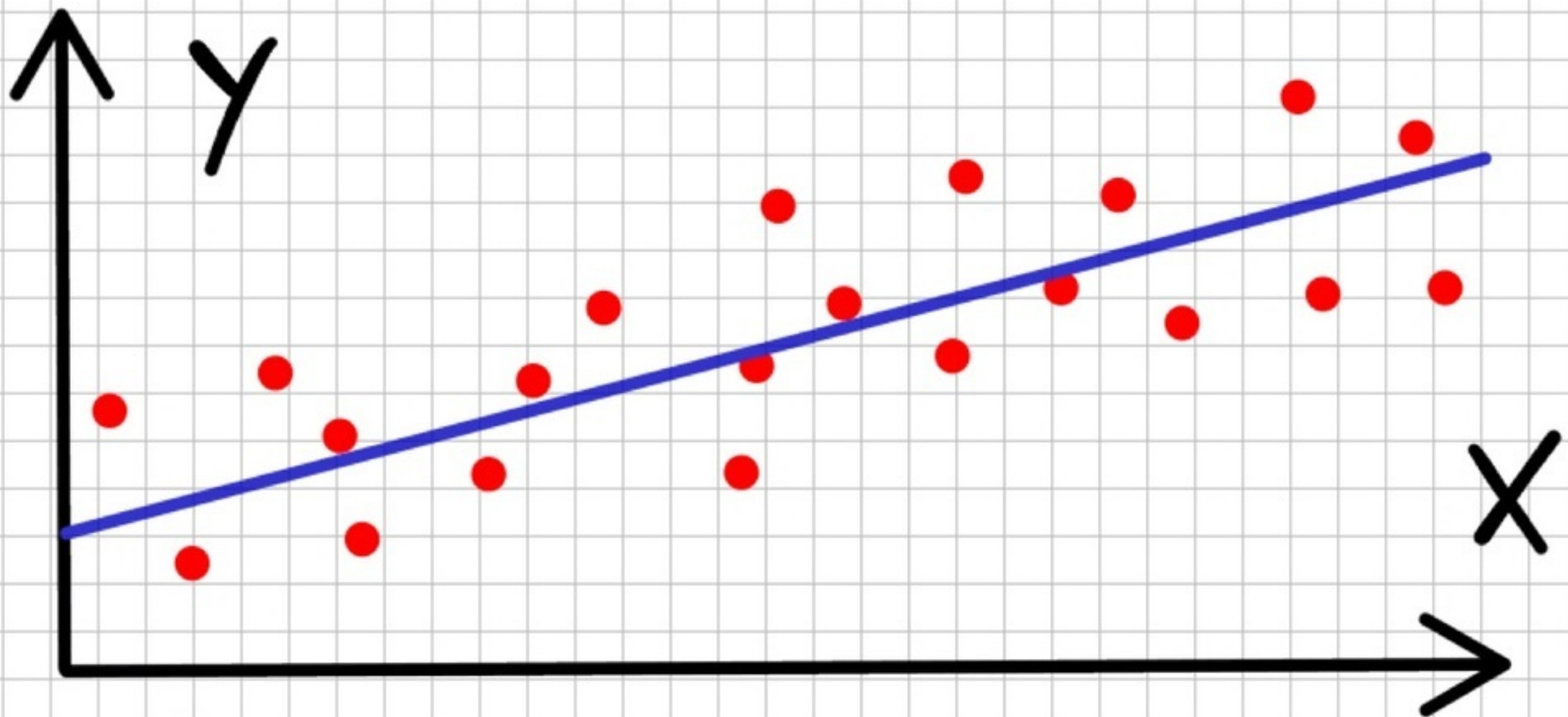
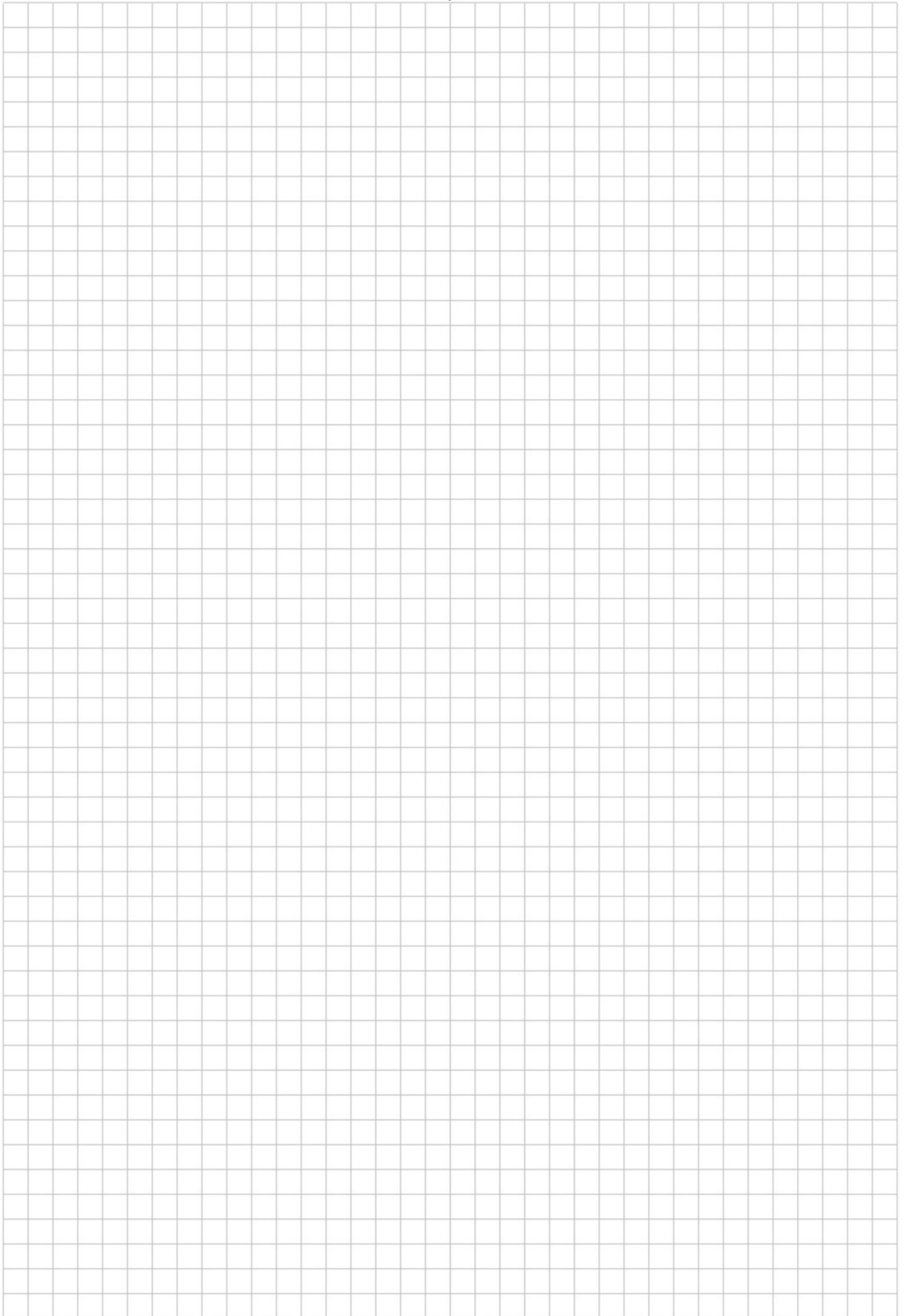


PrivateTeacher
Cours Privés de Science

Régression Linéaire







Introduction

Les droites de régression linéaires sont des modèles dont on se sert pour représenter la réalité.

En tant que modèle, une droite possède plusieurs avantages par rapport aux observations. Une droite en effet, permet de :

- 1) Résumer une série d'observations
- 2) Expliquer les observations.
- 3) Prédire les prochaines observations.

Un modèle cependant n'est pas la réalité mais seulement une approximation de celle-ci. Pour cette raison, une erreur est toujours associée aux droites de régression, comme nous allons le voir dans un instant.





La place du marché

Un marchand de pomme vend le fruit de sa récolte sur la place du marché.



Voici les prix affichés sur son étalage :

0.5 kilo de pomme coûte 1.5 CHF

1 kilo de pomme coûte 3 CHF

10 kilo de pomme coûte 30 CHF

On l'a compris, pour savoir quel est le prix des pommes, il suffit de multiplier le poids des pommes par 3.





Si x représente le poids des pommes,
et y le prix, alors on peut écrire
la relation sous forme d'une équation :

$$y = 3x$$

On dit alors que le prix des pommes
est proportionnel au poids, car le
rapport entre le prix et le poids
est toujours le même quel que soit
le nombre de pomme : $\frac{30}{10} = 3$ $\frac{9}{3} = 3$

On appelle 3 le facteur de proportionnalité

En mathématiques, deux valeurs sont dites
proportionnelles si leur rapport reste
identique, même quand ces valeurs changent.

si	1	kg	coûte	3	CHF
alors	2	kg	coûtent	$3 \cdot 2 =$	6
et	4	kg	coûtent	$3 \cdot 4 =$	12
et	10	kg	coûtent	$3 \cdot 10 =$	30

$$3x = y$$





Cette relation simple permet de résumer toute la liste des prix.

Elle permet également d'expliquer le prix : si je paie cher, c'est parce que j'ai mis beaucoup de pommes.

Cette relation enfin permet de prédire un prix qui ne serait pas affiché par le marchand : si je choisis 3.45 kilos de pommes, alors je paie $3.45 \cdot 3 = 10.35$ CHF





Représentation graphique

La droite est une manière visuelle de représenter une relation de proportionnalité.

On la construit à l'aide d'un tableau :

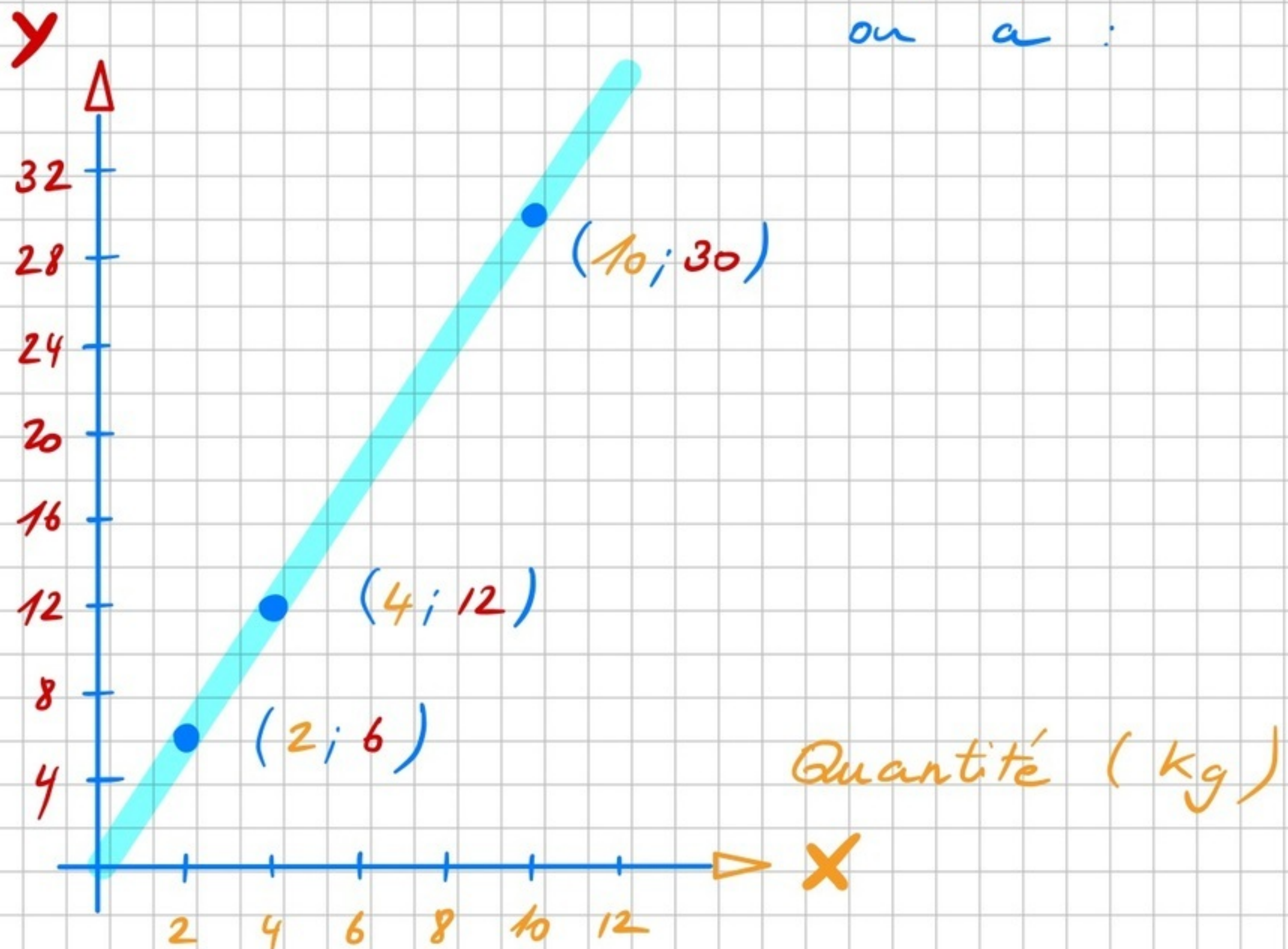
X	Y
2	6
4	12
10	30

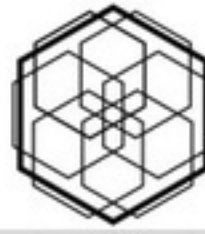
Il existe une relation entre la quantité de pommes X et le prix des pommes Y

Mathématiquement elle prend la forme $y = 3x$

Prix (CHF)

Graphiquement on a :

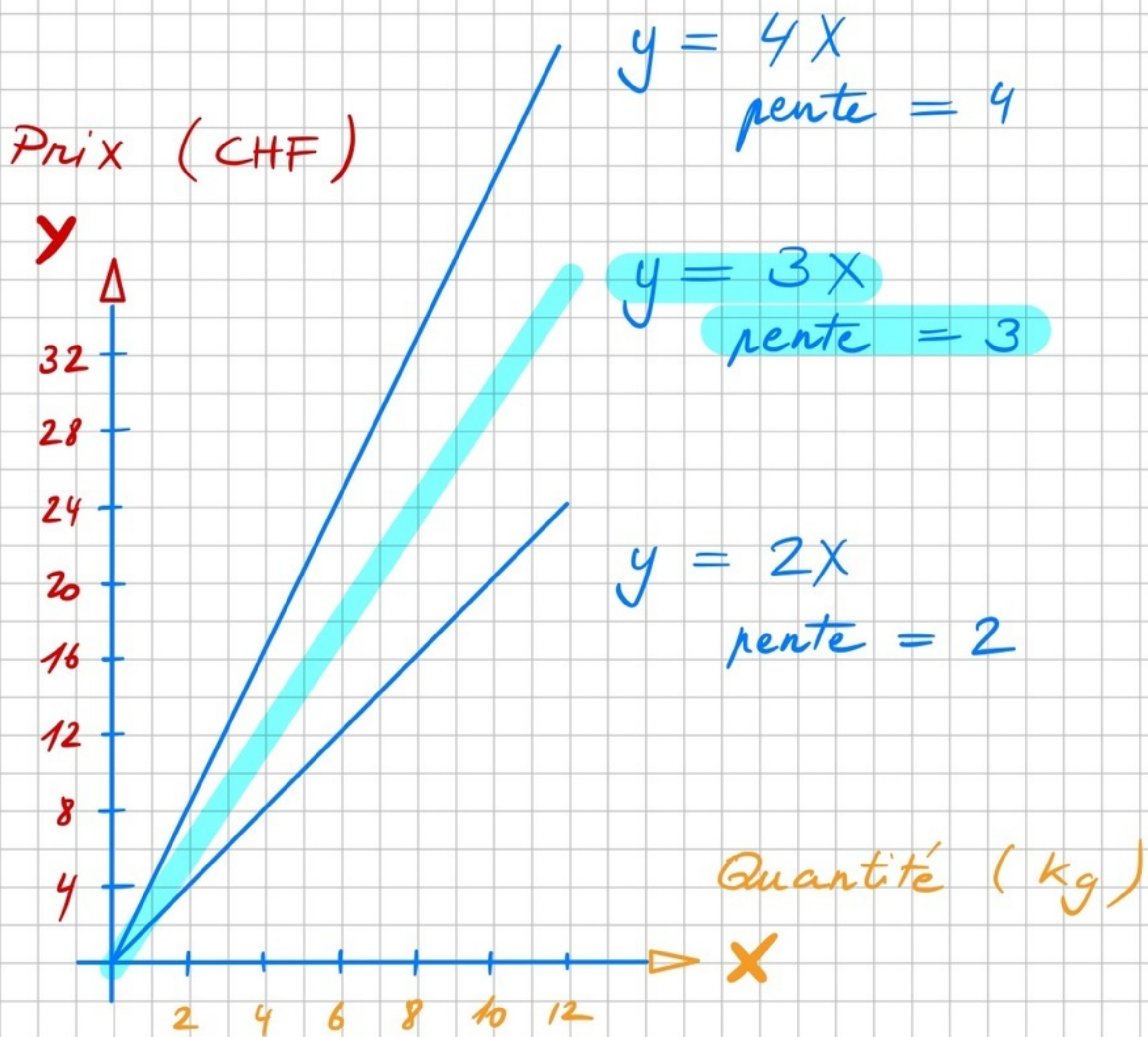




La pente

Si le prix du kilo de pomme avait été plus élevé, le prix des pomme augmenterait plus rapidement et la pente de la droite serait plus élevée

À l'inverse si le prix avait été plus bas, la pente aurait été plus faible





On peut lire la pente directement sur l'équation.

Elle est le facteur qui multiplie "x"

La pente représente le coefficient de proportionnalité entre le prix et la quantité.

À l'inverse, on peut lire la pente directement sur le graphique
voici comment s'y prendre :





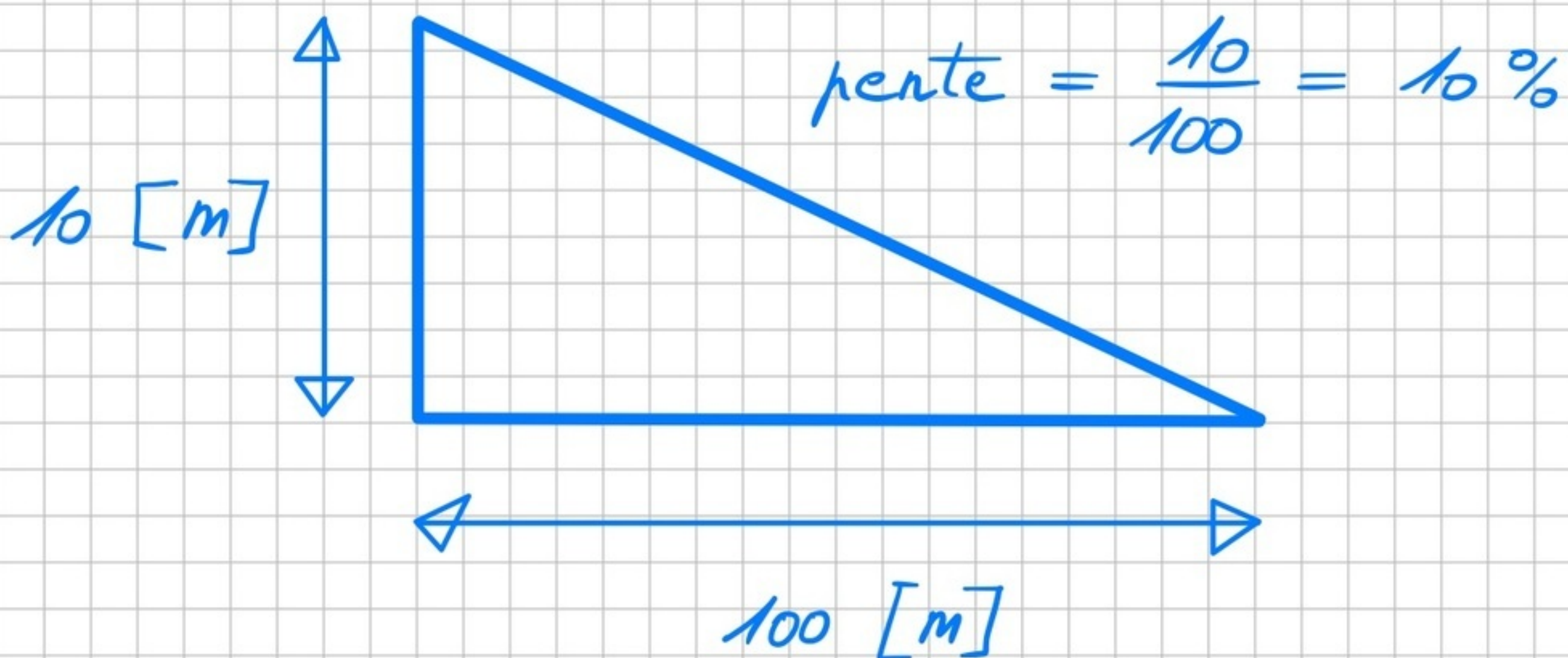
Calcul de la pente

La pente indiquée sur ce panneau est de 10 %.

Cela signifie que pour 100 [m] parcouru horizontalement, la voiture sera descendue de 10 [m] verticalement.



Une pente est donc le rapport entre une distance verticale et une distance horizontale.





Ordonnée à l'origine

Imaginons maintenant que l'entrée à la place du marché soit payante.

Avant de pouvoir acheter des pommes au prix indiqué, il faut s'acquitter d'un montant de 8 CHF.

Le prix des pommes "y" est toujours proportionnel au nombre de pommes ($y = 3x$) mais cette fois, le montant de l'entrée s'y ajoute systématiquement si bien que le prix final devient :

$$y = 3x + 8$$

2 kg coûtent	$3 \cdot 2 + 8 = 14$
4 kg coûtent	$3 \cdot 4 + 8 = 20$
10 kg coûtent	$3 \cdot 10 + 8 = 38$

$$3x + 8 = y$$

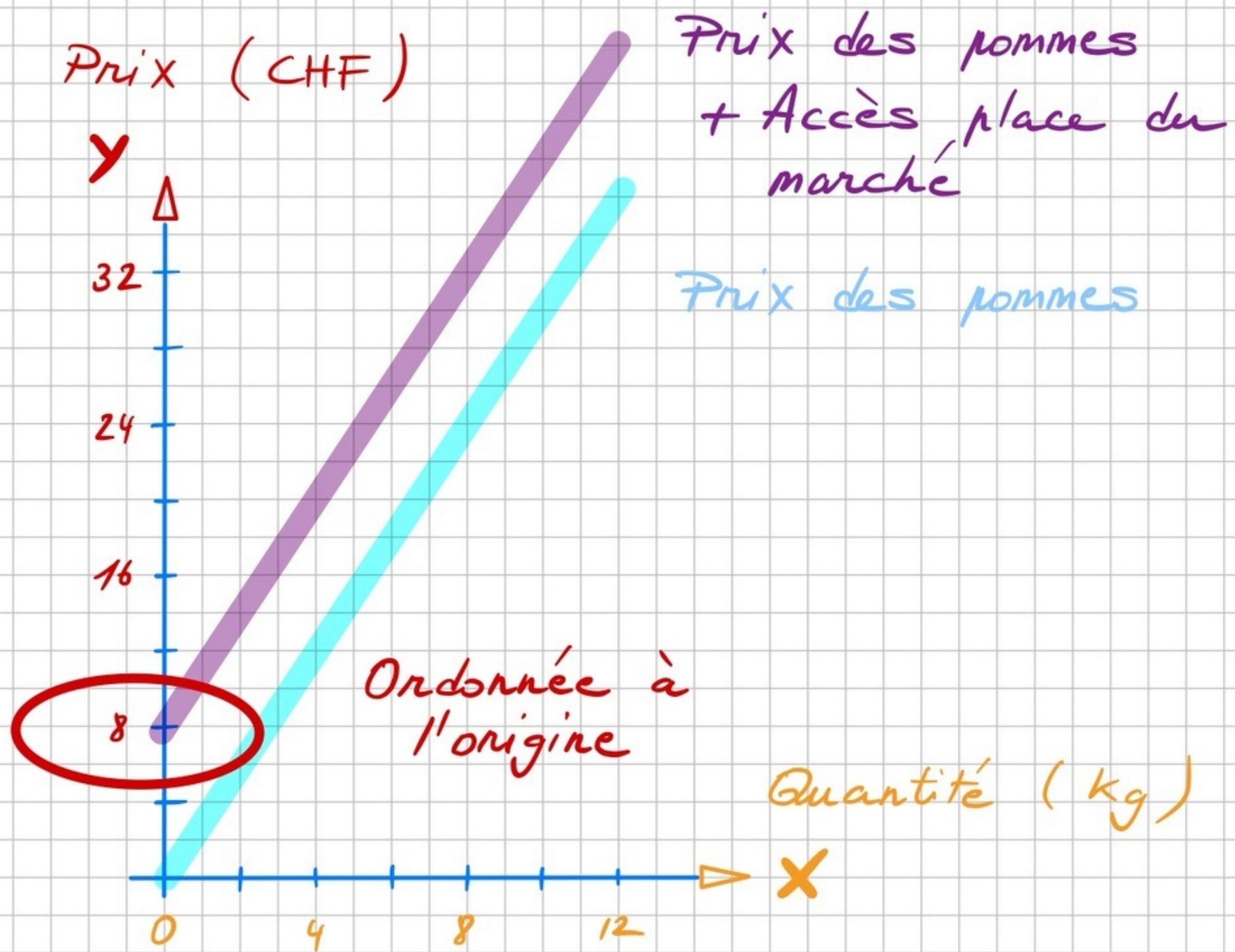
Pour dessiner la relation entre le prix des pommes et le nombre de pommes, on refait notre tableau

x	y
2	14
4	20
10	38





Graphiquement, cela se traduit par une droite parallèle au dessus de la précédente :



On le voit, quelque soit le nombre de pommes, le prix final est toujours supérieur de 8 CHF

Sur le graphique en $x = 0$ (0 pomme) on lit que le prix est de 8 CHF

Cette valeur de y en $x = 0$ se nomme l'ordonnée à l'origine





Equation de la droite

Les droites nous l'avons vu, sont déterminées par deux valeurs :
sa pente et son ordonnée à l'origine.

Les droites en générale possède toutes une pente que l'on note "a" et une ordonnée à l'origine que l'on note "b"

$$y = ax + b$$

Il s'agit là de la forme générale de l'équation de la droite.

Cette forme désigne la famille de toutes les droites d'équation :

$$y = 4x + 2$$

désigne une droite en particulier.

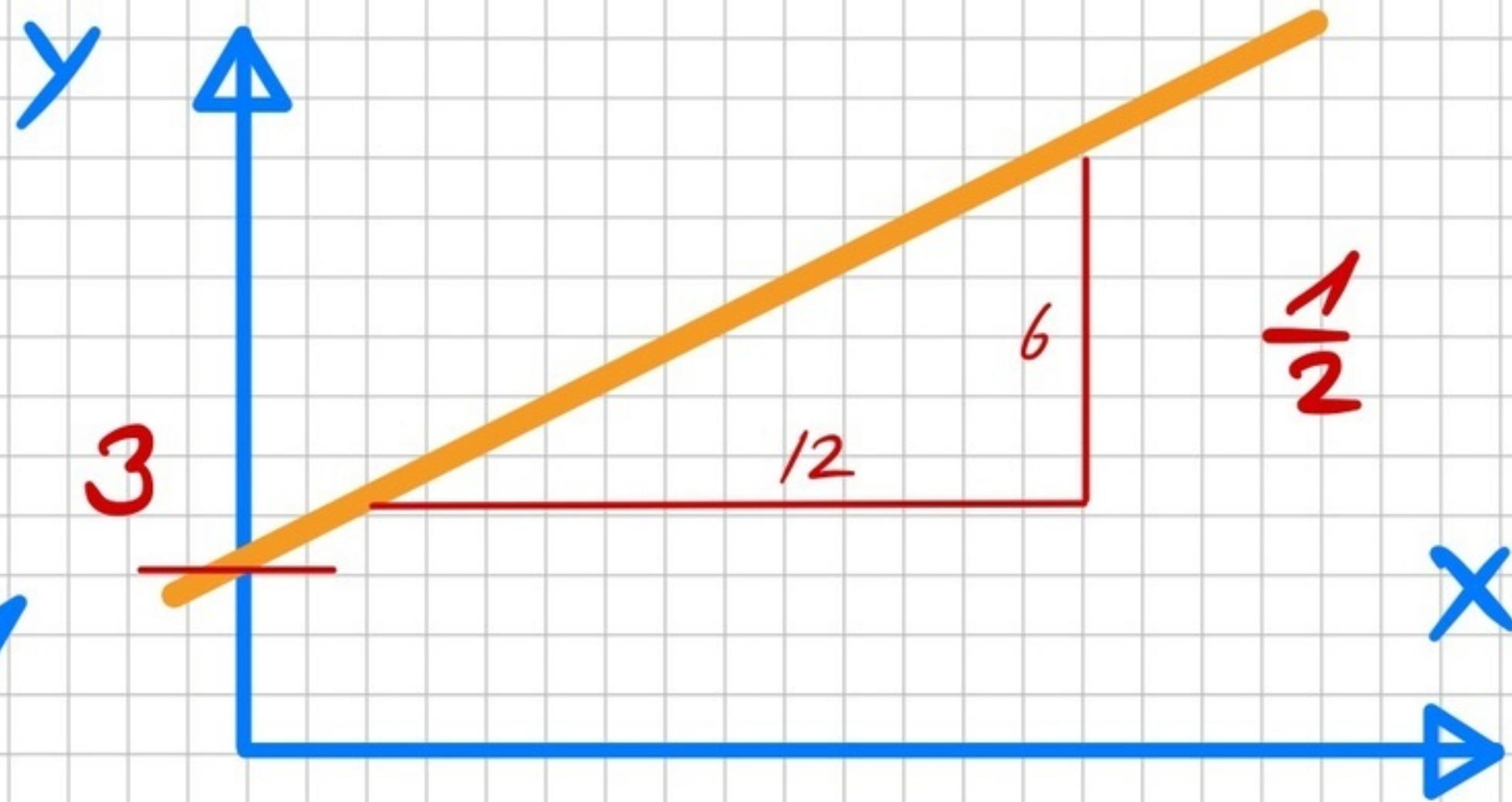
On peut choisir "a" et "b" librement et obtenir n'importe quelle droite en particulier.

Pour cette raison, "a" et "b" se nomment les paramètres de la droite.





La droite en bref



On peut dessiner
le graphique
à partir de
l'équation.

On peut lire
l'équation
sur le
graphique

$$y = \frac{1}{2}x + 3$$





Régression linéaire

Lorsque l'on passe de l'équation générale de la droite $y = ax + b$ à une équation particulière p. exple $y = 4x + 7$, dans le but d'ajuster la droite à des observations on effectue ce qu'on appelle une régression linéaire

La régression linéaire est un processus qui permet de trouver l'équation d'une droite particulière : la droite qui se superpose à nos observations.

De telles équations représentent donc un modèle de la réalité dont on peut dès lors se servir pour faire des prédictions.

Dans une certaine mesure on peut substituer le modèle à la réalité mais il ne faut jamais oublier que ceux-ci ne sont que des approximations





Nous avons vu quels étaient les avantages d'utiliser une droite pour représenter une relation de proportionnalité.

Voyons à présent comment trouver une telle relation au sein de nos données.

Nos observations sont réunies dans le tableau ci-contre.

Voici comment réaliser une régression linéaire à la main.

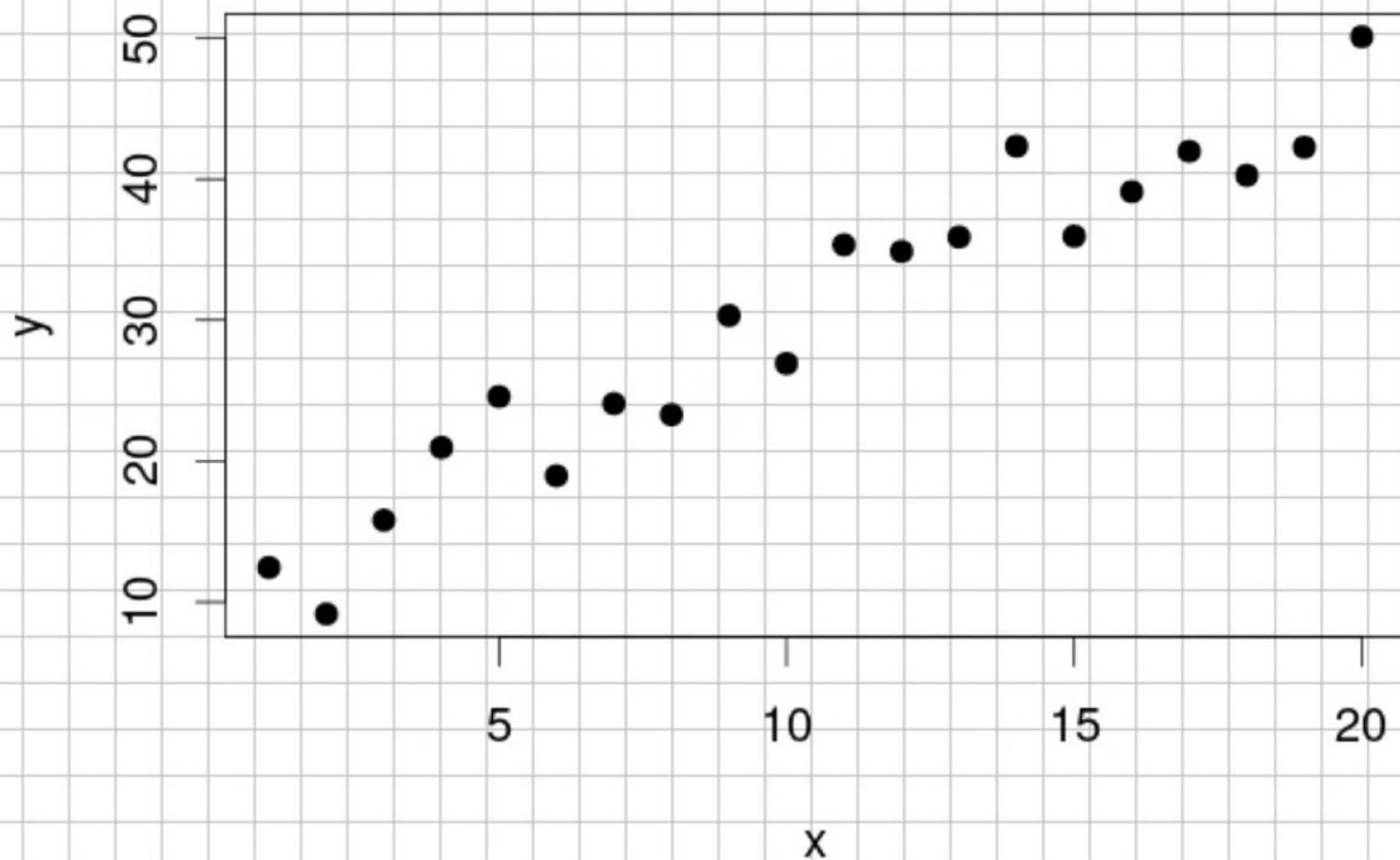
X	Y
1	12.45
2	9.14
3	15.80
4	20.96
5	24.57
6	18.95
7	24.07
8	23.28
9	30.31
10	26.92
11	35.34
12	34.86
13	35.89
14	42.34
15	35.95
16	39.10
17	41.97
18	40.26
19	42.26
20	50.09

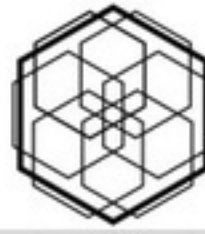




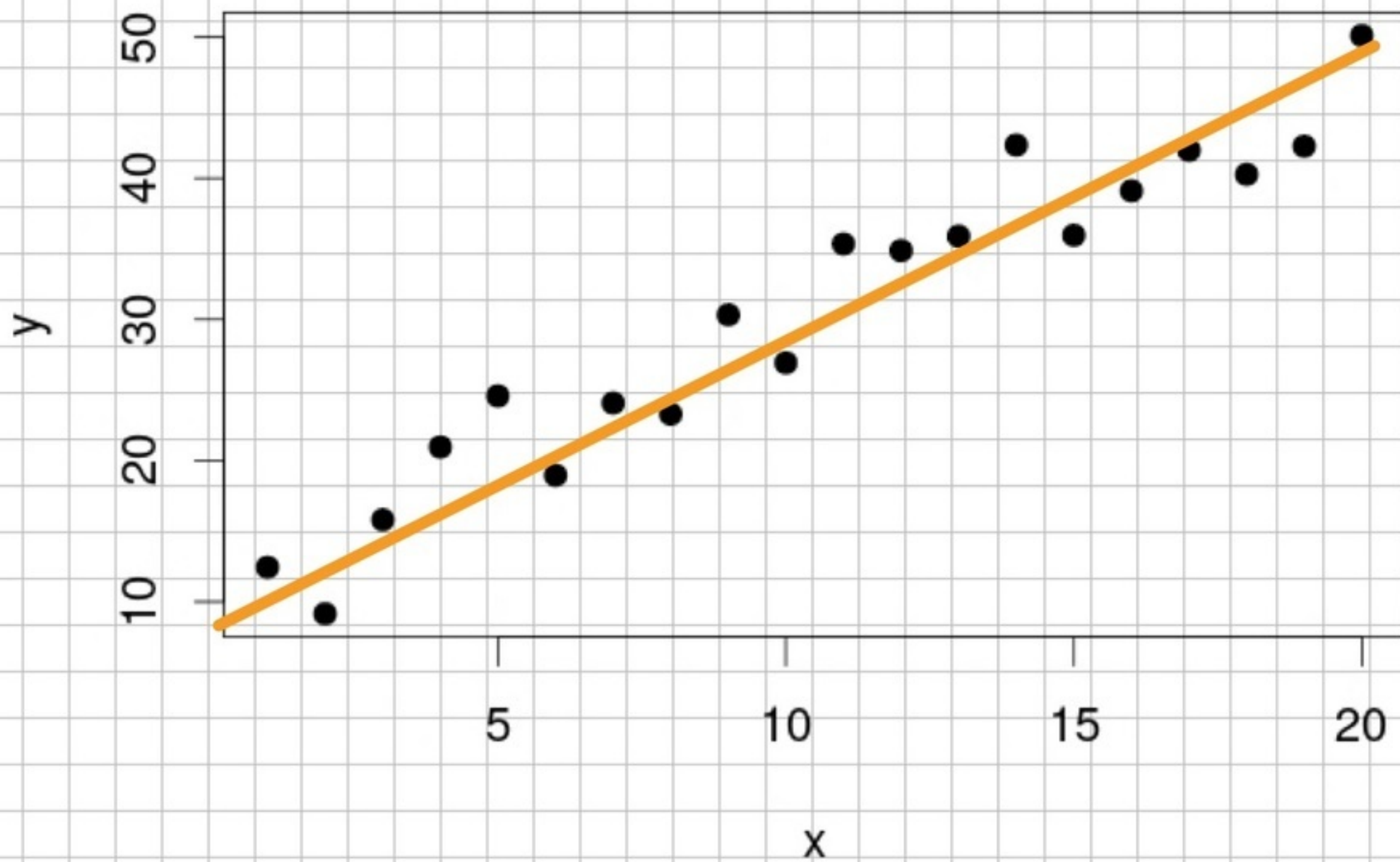
1) la première étape consiste à représenter graphiquement une caractéristique X en fonction d'une caractéristique Y

On réalise cela à l'aide d'un scatterplot





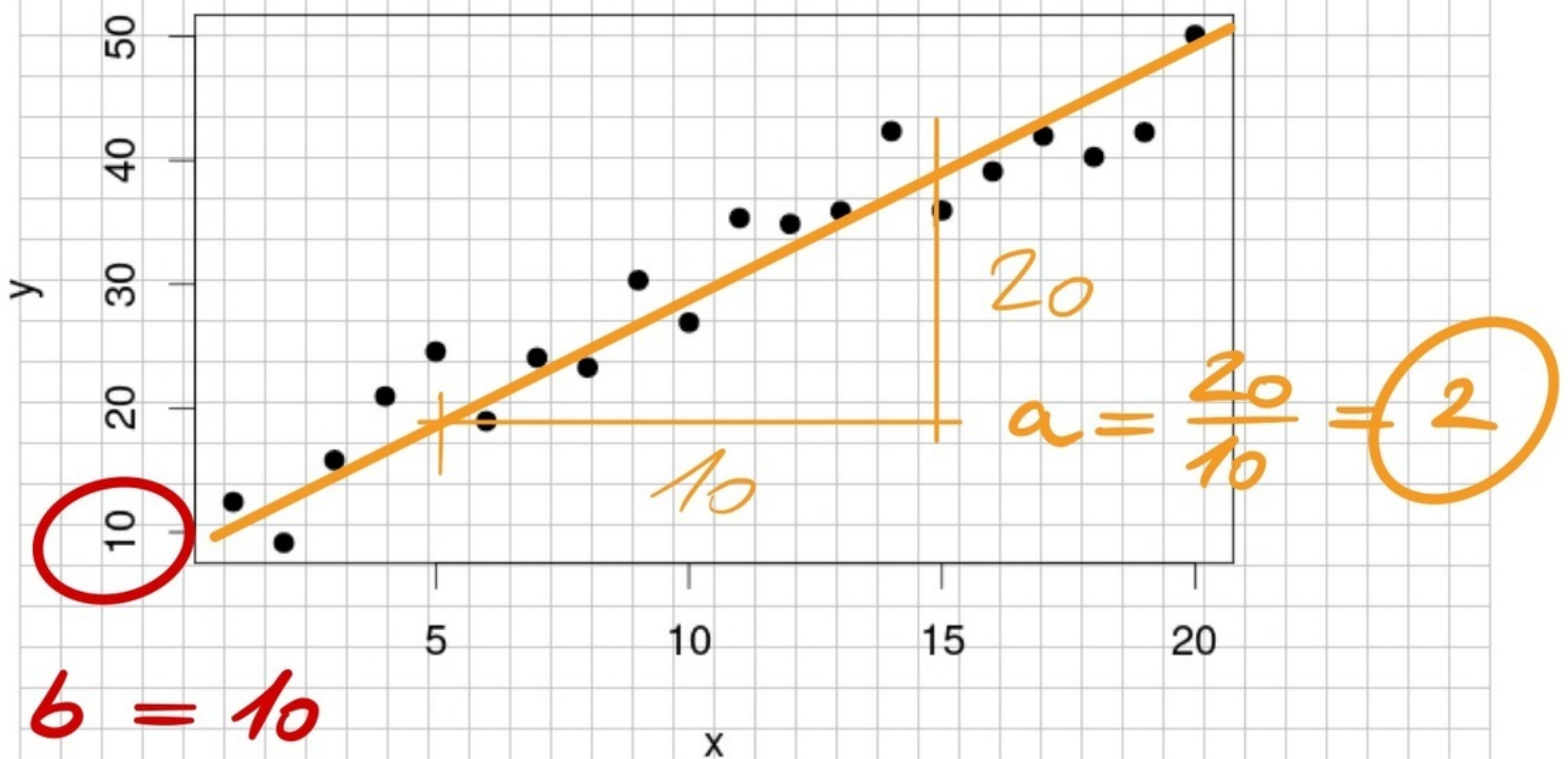
2) On ajuste ensuite une droite à la main.
Le but du jeu est de trouver la
droite qui passe au plus près
de tous les points.





3) On utilise ensuite les outils vu auparavant pour déterminer l'équation de la droite :

$$y = ax + b$$



Il s'agit là de notre modèle :
l'équation de la droite de régression
obtenue par régression linéaire sur
nos observations.



