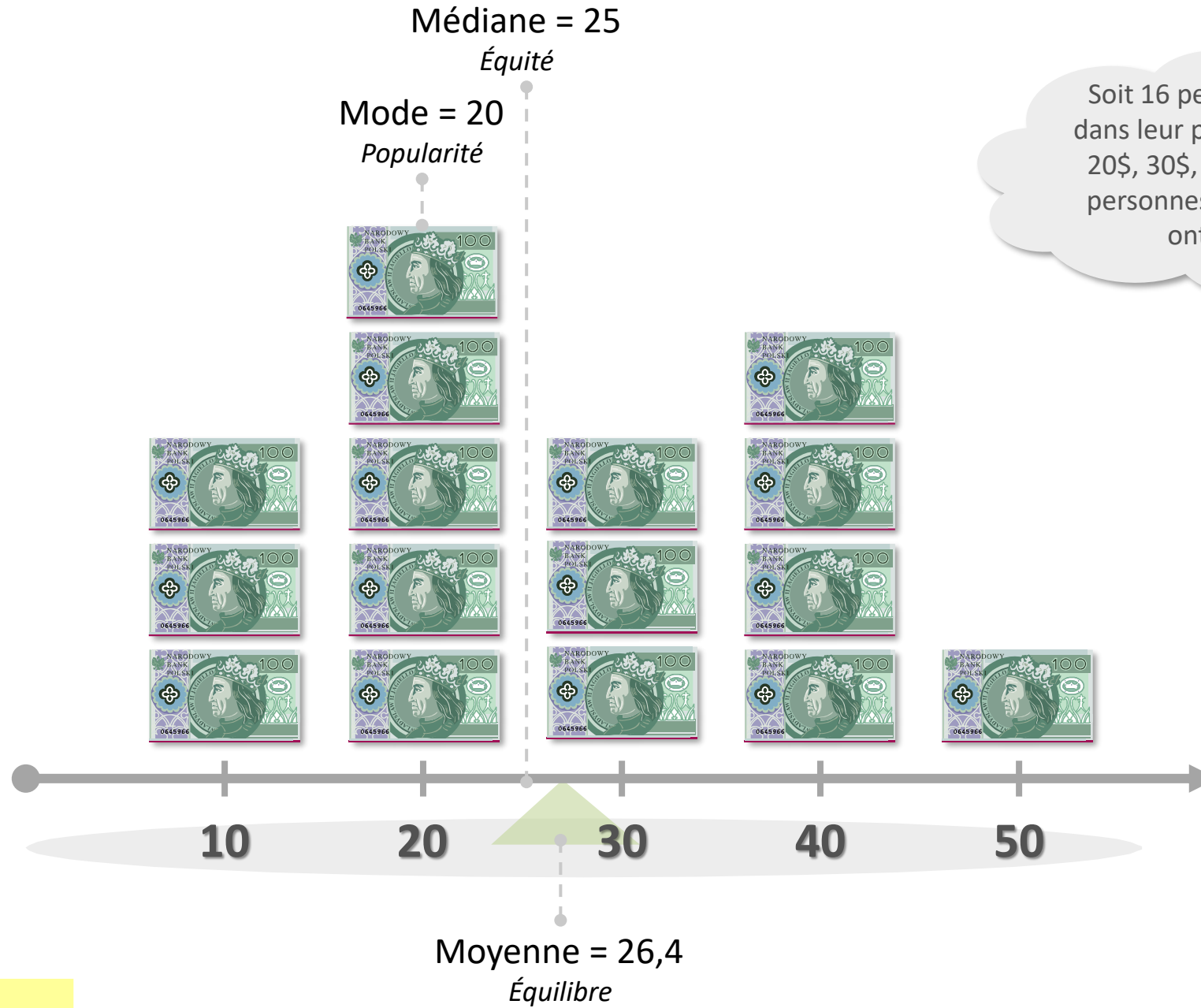


SCI6007 Méthodes de recherche en sciences de l'information

Statistiques descriptives (partie 2 de 2)

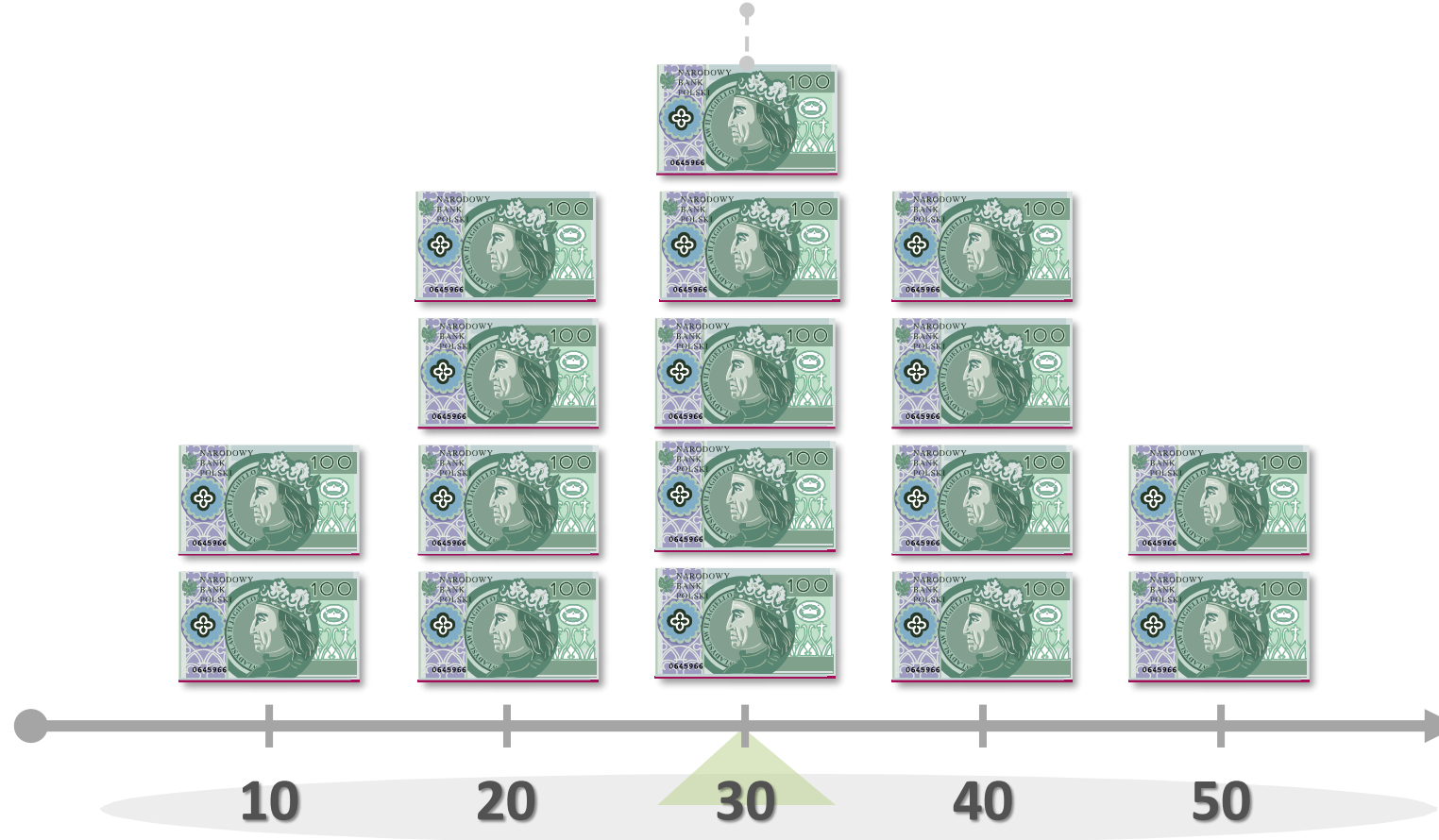
Exemples



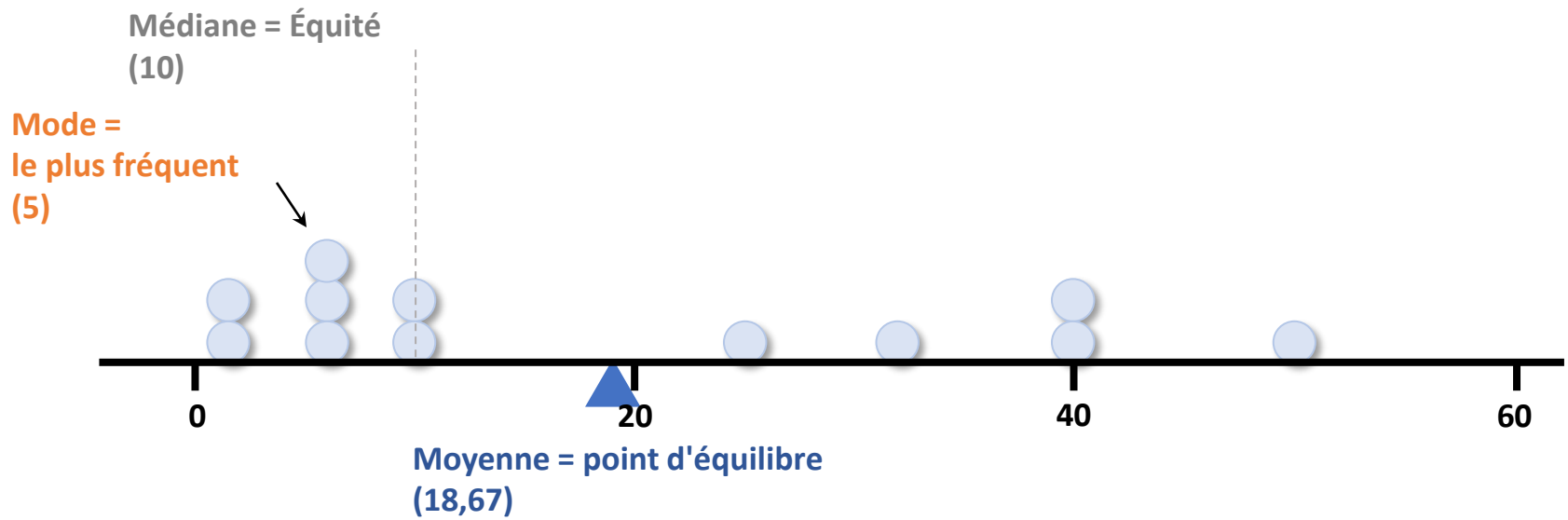
Soit 16 personnes ayant dans leur portefeuille 10\$, 20\$, 30\$, 40\$ ou 50\$: 3 personnes ont 10\$, 5 en ont 20\$, ...

Une distribution **unimodale symétrique** aura la même valeur pour sa médiane, son mode et sa moyenne. La distribution symétrique est parfaitement équilibrée!

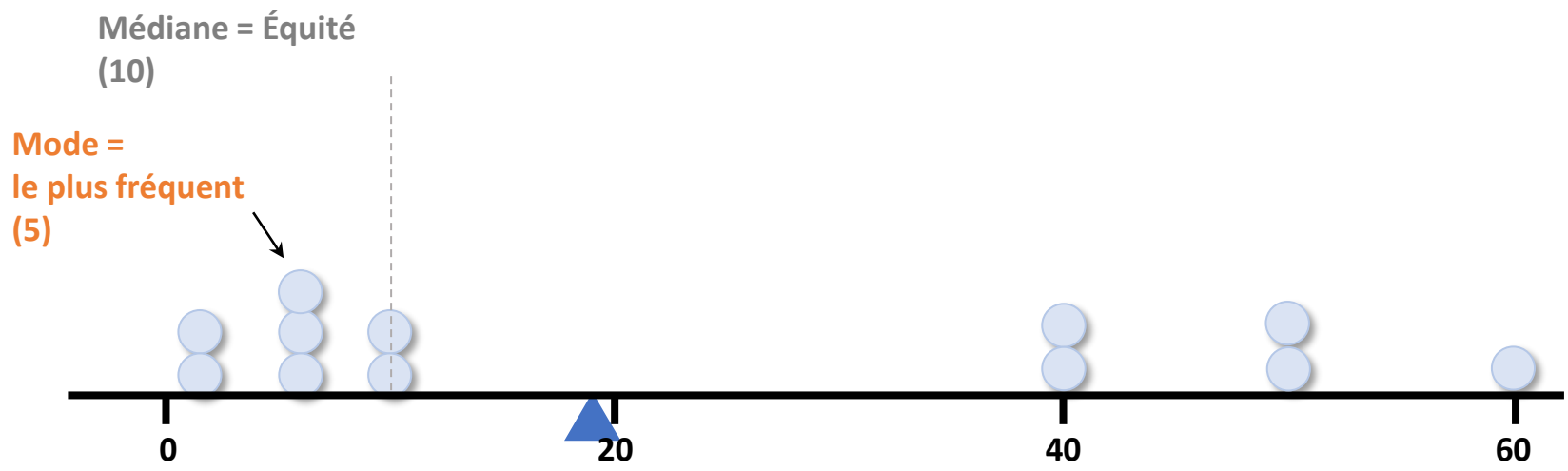
Médiane = Mode = Moyenne



Tendances centrales : Impact des valeurs aux extrémités

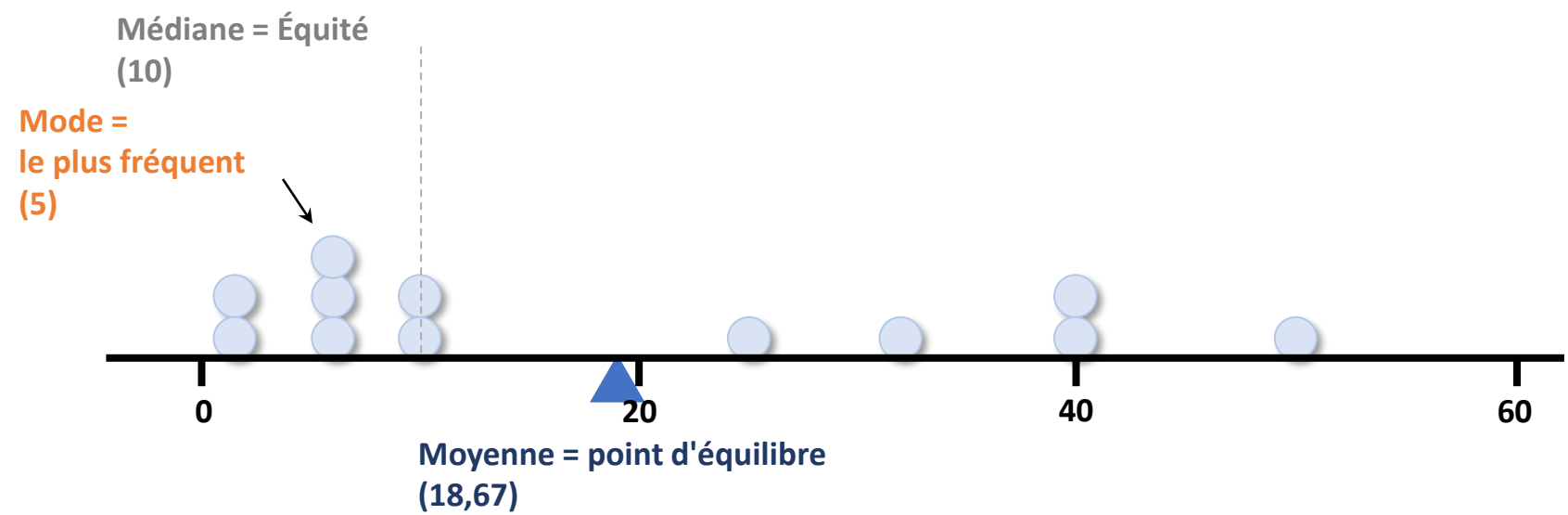


1
1
5
5
5
10
10
25
32
40
40
50

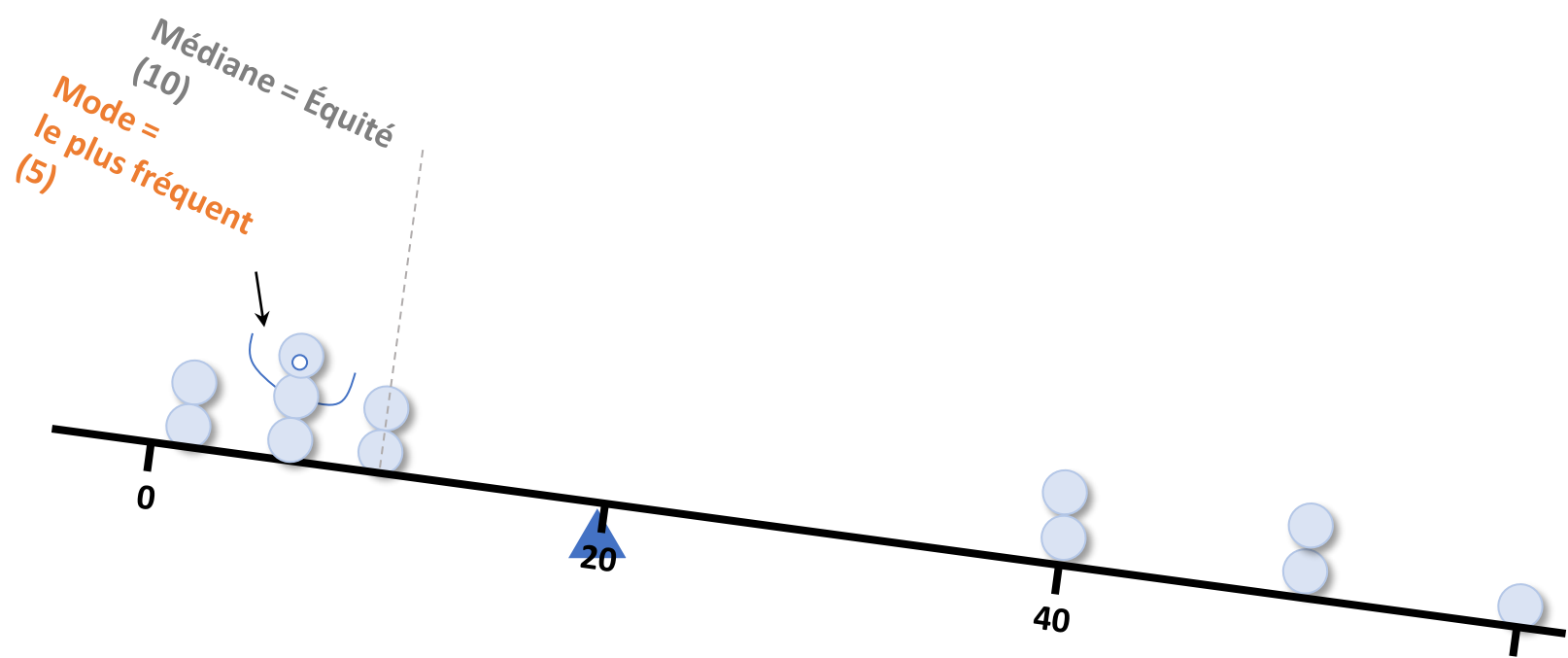


1
1
5
5
5
10
10
40
40
50
50
60

Tendances centrales : Impact des valeurs aux extrémités

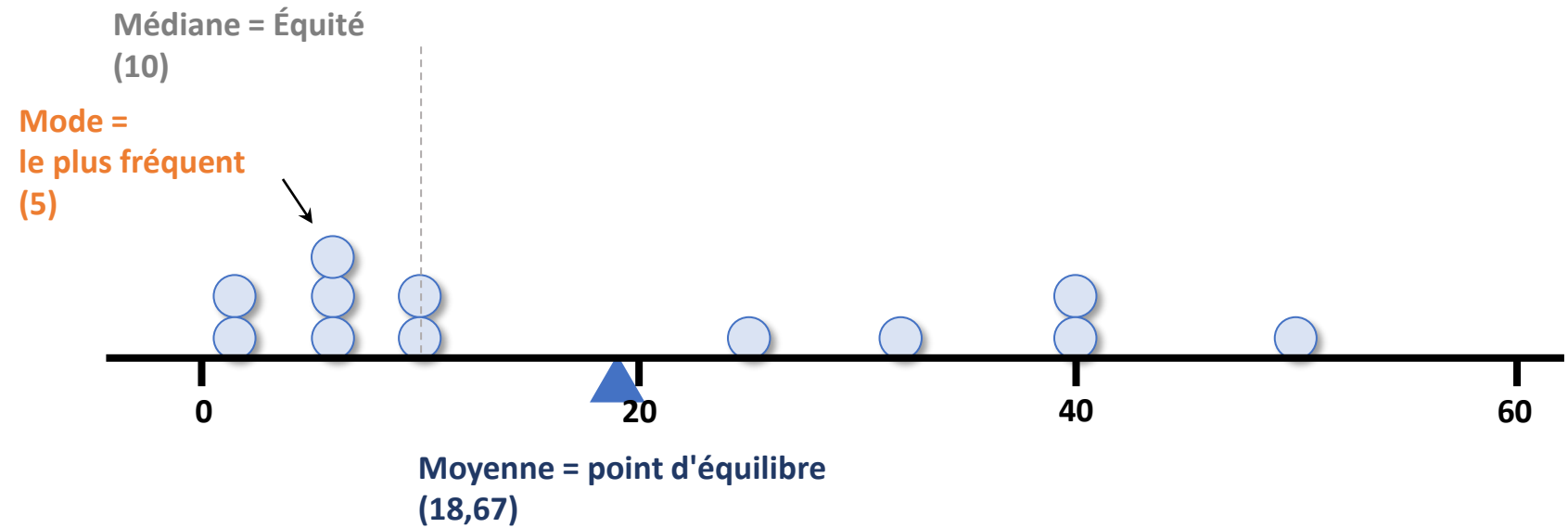


1
1
5
5
5
10
10
25
32
40
40
50

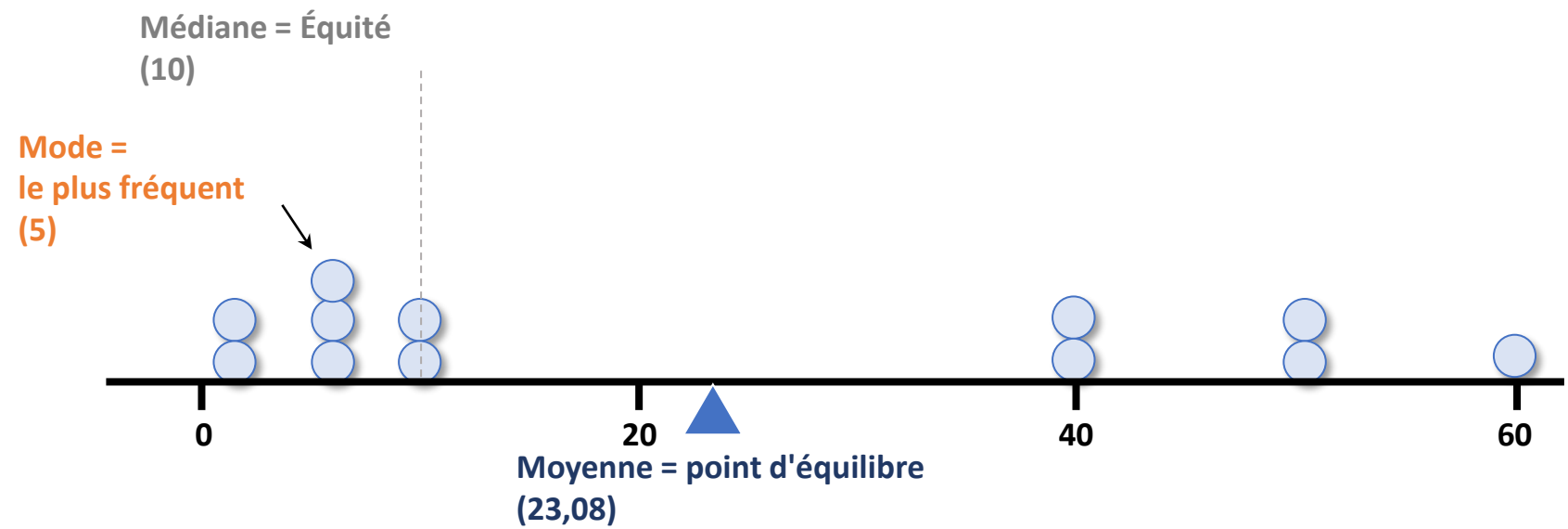


1
1
5
5
5
10
10
40
40
50
50
60

Tendances centrales : Impact des valeurs aux extrémités



1
1
5
5
5
10
10
25
32
40
40
50



1
1
5
5
5
10
10
40
40
50
50
60

Impact de l'asymétrie sur la moyenne et la médiane

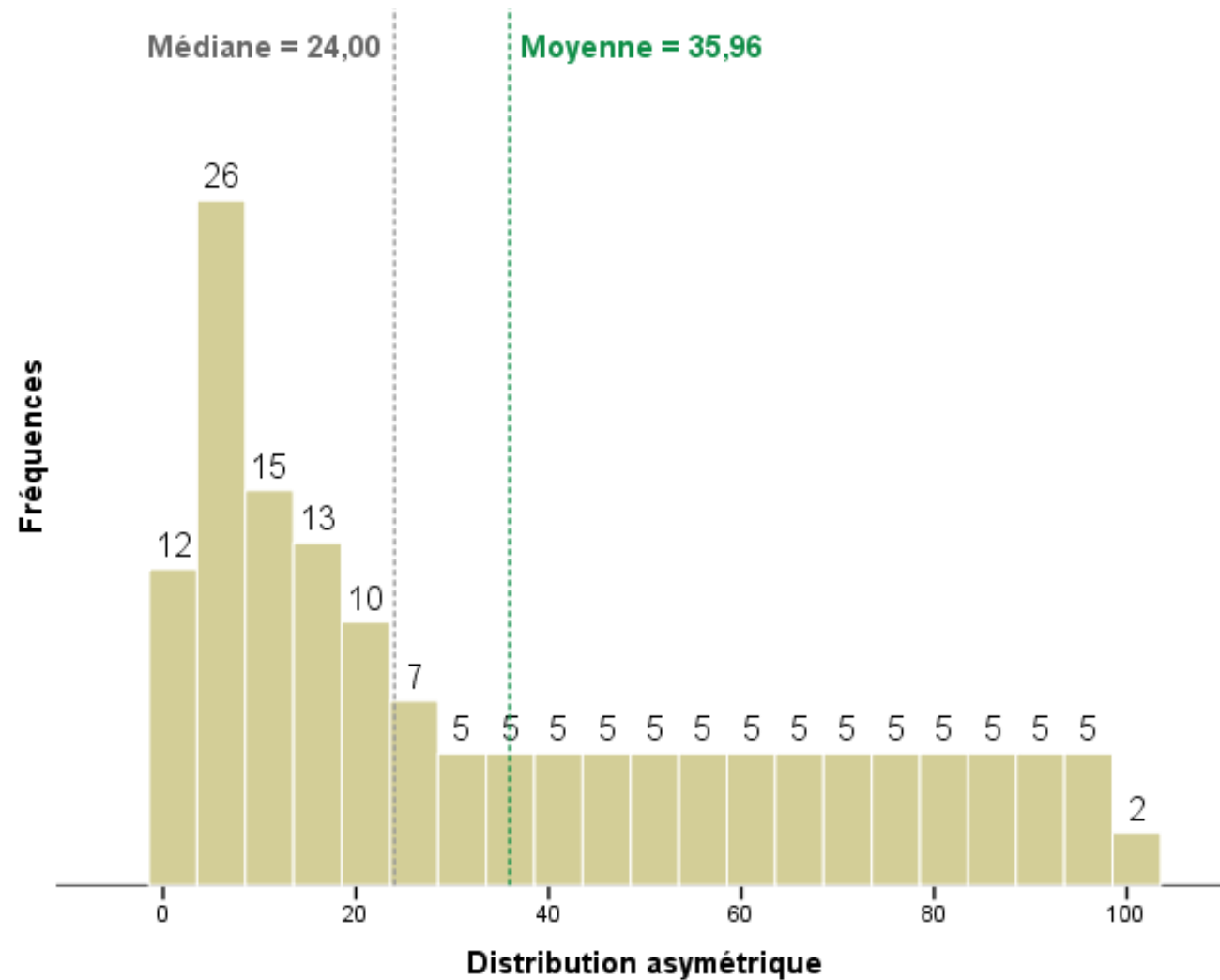
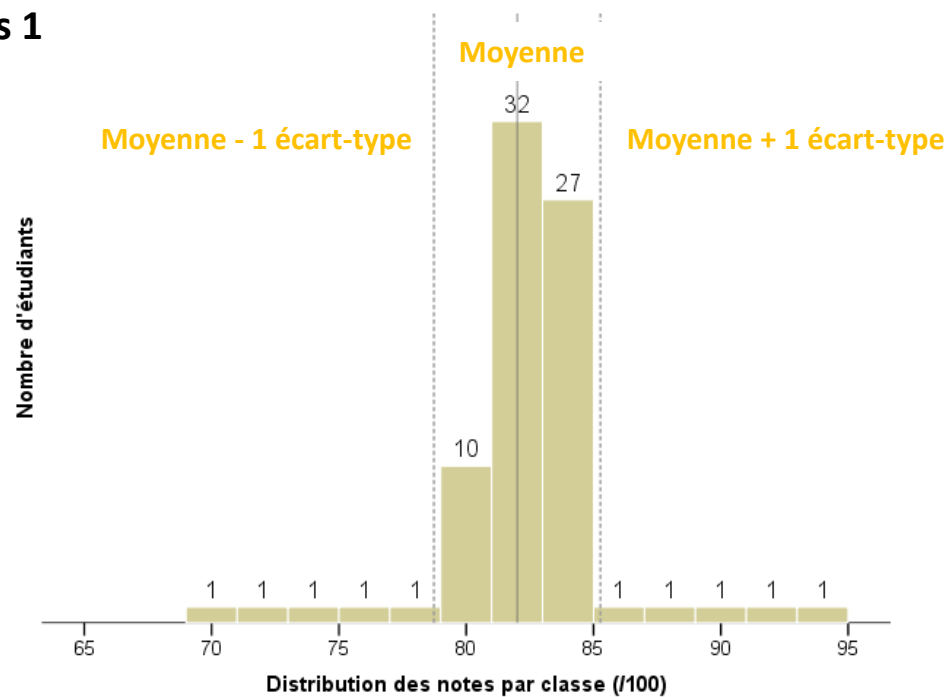


Illustration du lien entre la dispersion des données et l'écart-type

Cas 1



Exemples de distribution des notes pour 2 examens différents dans un cours

Cas 1 (examen fictif) :

Moyenne = 82

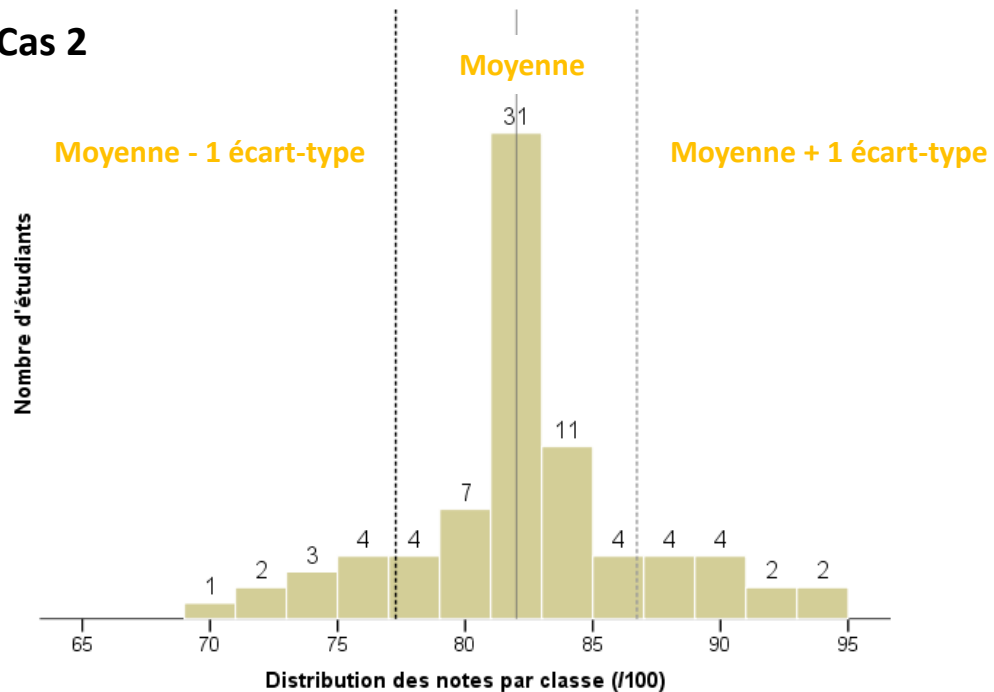
Nbre étudiants = 79

Presque tous les étudiants entre 80 et 85

Écart-type = 3,27

Coefficient de variation = 3,99%

Cas 2



Cas 2 (final) :

Moyenne = 82

Nbre étudiants = 79

Notes sont plus étendues entre 77 et 86

Écart-type = 4,72

Coefficient de variation = 5,76%

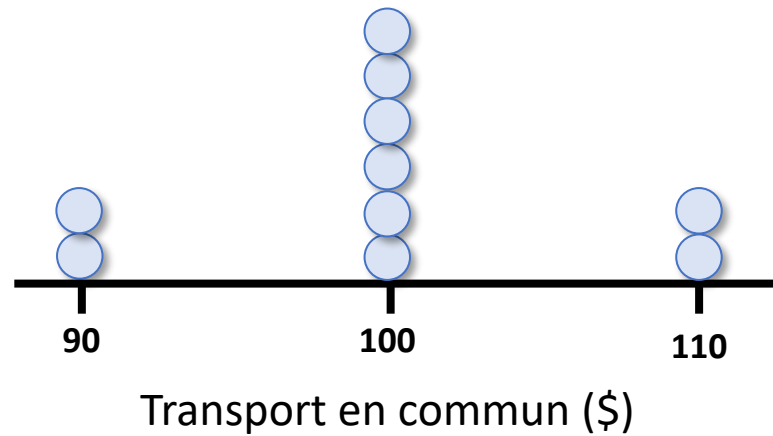
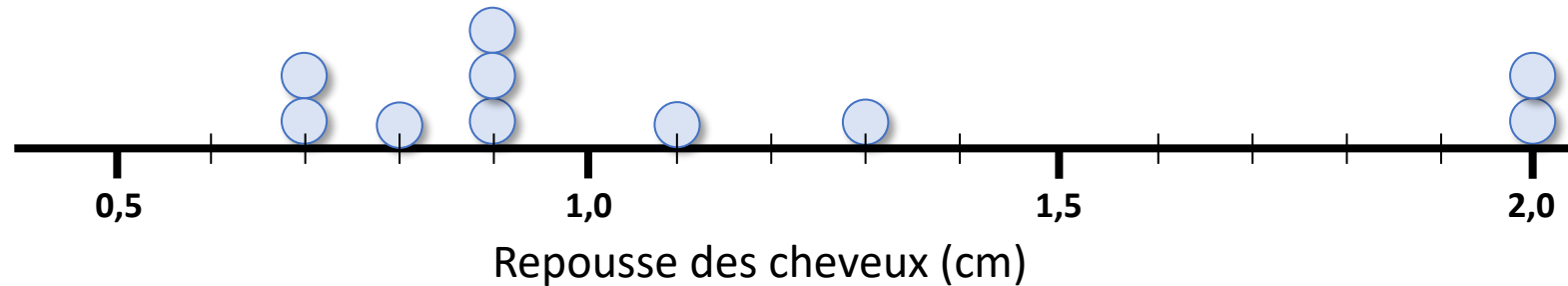
Interprétation de l'écart-type [1/2]

- ➡ Soit les deux indicateurs suivants :
 - A) Somme mensuelle dépensée pour les transports en commun (\$)
 - B) Longueur de la repousse des cheveux par mois (cm)
- ➡ 10 personnes vous fournissent leurs données pour ces deux indicateurs et vous calculez leur moyenne ainsi que leur écart-type

	Transports (\$)	Cheveux (cm)
Moyenne	100	1,13
Écart-type	6,67	0,49

- ➡ Selon vous, pour lequel de ces indicateurs vos répondants et répondantes diffèrent-ils/elles le plus (c'est-à-dire que les données sont les plus dispersées)?

Interprétation de l'écart-type [2/2]



Rép.	Transports (\$)	Cheveux (cm)
1	90	0,7
2	90	0,8
3	100	0,9
4	100	1,1
5	100	2,0
6	100	0,9
7	100	1,3
8	100	0,7
9	110	2,0
10	110	0,9
Moyenne	100	1,13
Écart-type	6,67	0,49

L'écart-type ne s'interprète pas uniquement sur la grandeur de son chiffre, mais doit s'interpréter par rapport à la moyenne et en fonction de son unité de mesure. Ici, il s'avère que c'est pour la repousse des cheveux que l'écart-type est le plus grand!

Coefficient de variation (COV)

- ➔ Aide à interpréter l'écart-type en le ramenant en %
- ➔ **COV = écart-type / moyenne**

Rép.	Transports (\$)	Cheveux (cm)
1	90	0,7
2	90	0,8
3	100	0,9
4	100	1,1
5	100	2,0
6	100	0,9
7	100	1,3
8	100	0,7
9	110	2,0
10	110	0,9
Moyenne	100	1,13
Écart-type	6,67	0,49
COV	6,67%	43,56%

Plus le COV est grand, plus l'écart-type représente une grande proportion de la moyenne, plus les données sont dispersées autour de la moyenne.

Ici, le COV nous permet de facilement voir que c'est l'indicateur sur la repousse des cheveux qui a la plus grande dispersion des données.

Si **COV < 15%**, les données sont considérées comme **homogènes** par rapport à l'indicateur.

Si **COV > 15%**, les données sont **hétérogènes** par rapport à l'indicateur.