



Les statistiques descriptives

Terminologie, tableaux et
graphiques

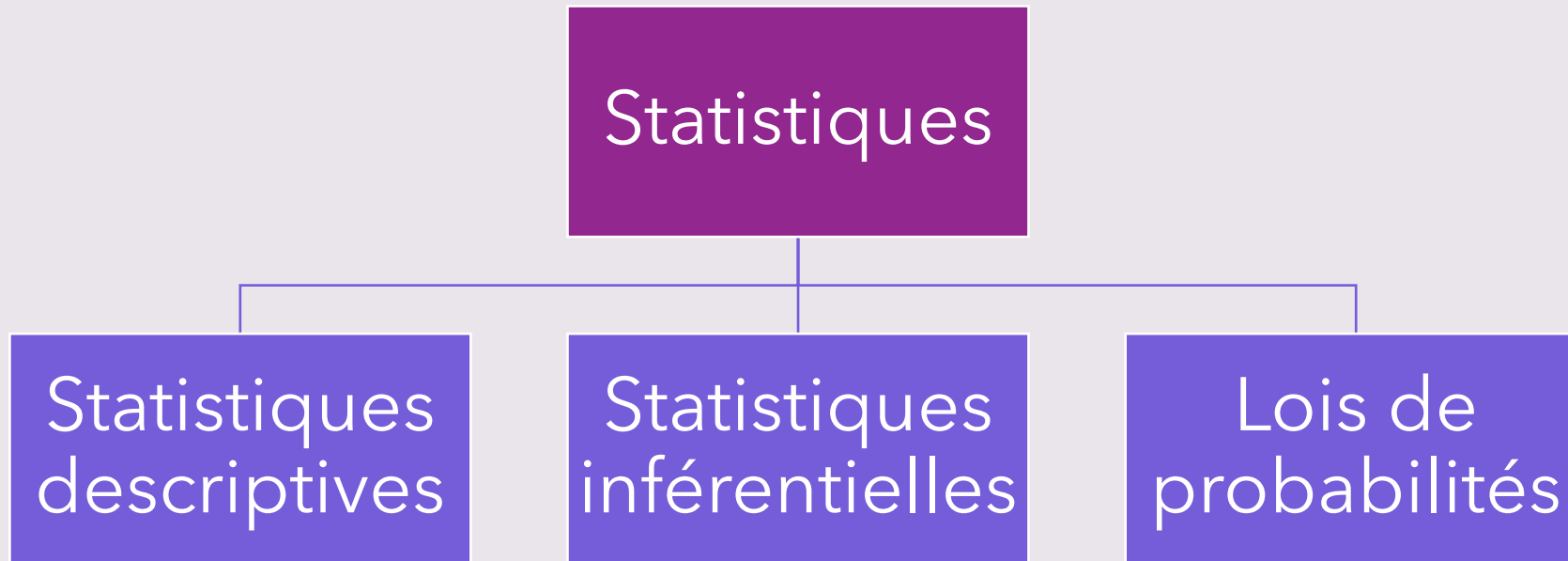
Introduction

❖ Pourquoi? Utilité des statistiques?...

Les statistiques servent à :

- *Amasser, classer, résumer, présenter et interpréter des données.*
 - *Elles peuvent aussi servir à :*
-

Vue d'ensemble



Terminologie

Population :

Ensemble de tous les individus ou de tous les objets, c'est-à-dire de toutes les unités statistiques sur lesquelles porte une recherche.

Exemple : On veut en savoir plus sur les habitudes de travail des étudiants du cégep de Shawinigan.

Population: Tous les étudiants du cégep

Le nombre d'individus d'une population est noté par la lettre majuscule « N ».

Exemple précédent : $N =$ _____

Unité statistique :



Etc.

C'est l'objet ou celui dont on étudie les caractéristiques dans une recherche.

Exemple précédent :

Unité statistique: un étudiant du cégep

Échantillon :

C'est une partie, un sous-ensemble de la population.

Exemple : On choisit 20 étudiants au hasard parmi les 1250 étudiants du cégep.

La taille d'un échantillon est notée par la lettre minuscule « **n** »

Exemple précédent :

L'échantillon est de taille $n = 20$

- Si une recherche porte sur une population, elle porte le nom de

.....

- Si une recherche porte sur un échantillon, elle porte le nom de

.....

Pourquoi échantillonner (travailler à partir d'un échantillon) ?

☑ pour diminuer les coûts et le temps que prendra la recherche car on a un nombre moindre d'individus à observer ou interroger

Comment choisir un échantillon ?

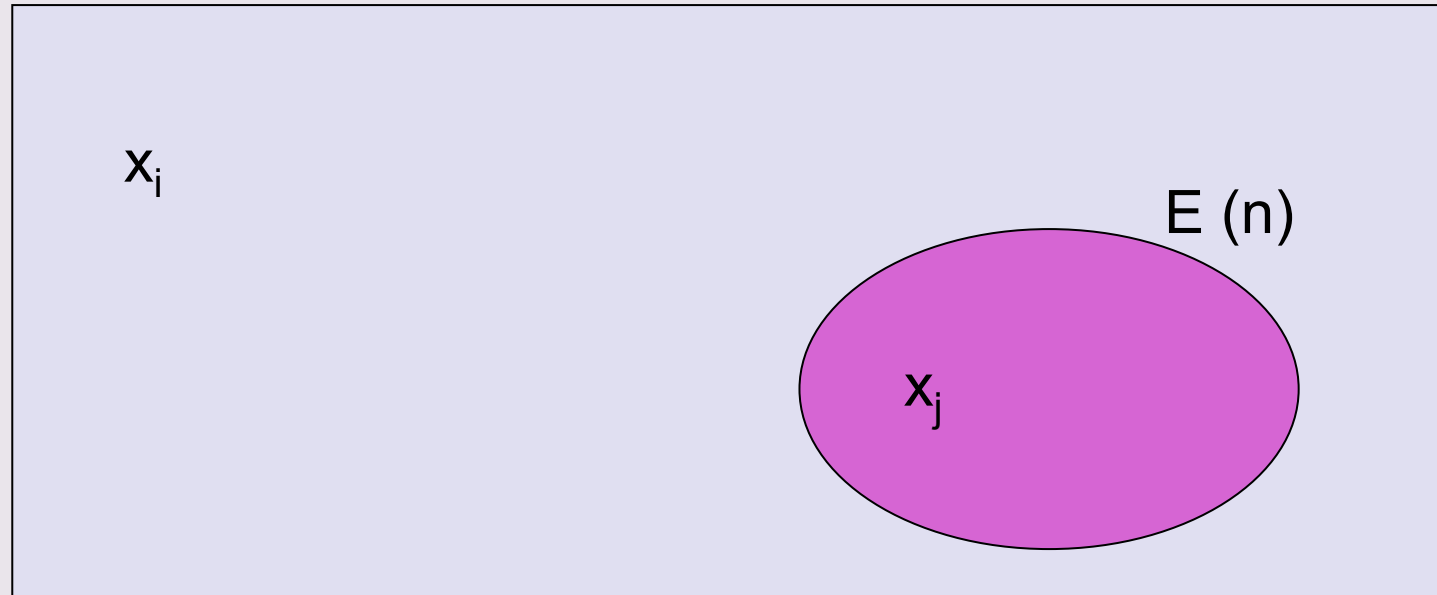
☑ il doit être représentatif de la population

☑ il doit être choisi au hasard

☑ on doit respecter les procédures d'échantillonnage

Contexte statistique

$P(N)$



On dit que les x_i et x_j sont des

Variables

Lors d'une étude, on choisit tout d'abord les unités statistiques (êtres humains, animaux, machinerie, etc.), on détermine ensuite quelles seront les caractéristiques d'intérêt liées à ces unités statistiques (âge, sexe, poids, etc.). Dans le langage statistique, on appelle ces caractéristiques, les variables.

Exemple : On s'intéresse aux habitudes de travail des étudiants du cégep.

On décide de prendre (afin de les compiler) les informations suivantes :

Âge ; Sexe ; programme d'étude ; Nb d'heures travaillées par semaine ;

Salaire hebdomadaire ; Nb heures d'étude par semaine ; etc.

Nature d'une variable

1) Variable qualitative : la valeur prise par la variable est une _____

Variable qualitative	
Nominale	Ordinale
Les modalités ne possèdent pas d'ordre	Les modalités possèdent un ordre
Exemples : _____ _____ _____	Exemples : _____ _____ _____

Nature d'une variable

2) Variable quantitative : la valeur prise par la variable est un _____

Variable quantitative	
Discrète	Continue
La valeur est un nombre entier obtenue par dénombrement	La valeur peut être décimale, elle est obtenue suite à une « mesure »
Exemples : _____ _____ _____	Exemples : _____ _____ _____

Exemples :

1) X : Les employés consultent-ils leurs courriels ?

- ☐ Jamais
- ☐ Rarement
- ☐ Souvent
- ☐ Toujours

2) Y : Nb de fautes d'orthographe dans un texte de cinq pages

3) Z : Âge à laquelle vous avez terminé vos études

4) W : Temps consacré à la préparation d'un examen

5) A : Profession d'un individu

Remarque : Certaines variables quantitatives comme « Âge » ou « Revenu » peuvent être considérées pour des fins d'analyse comme des ***variables qualitatives*** si on crée des regroupements.

Ex. : 1) Groupes d'âge pour des enfants

0-5 ans

6-12 ans

13-17 ans

2) Catégories de revenu pour des profs

30 000 à 45 000

45 000 à 60 000

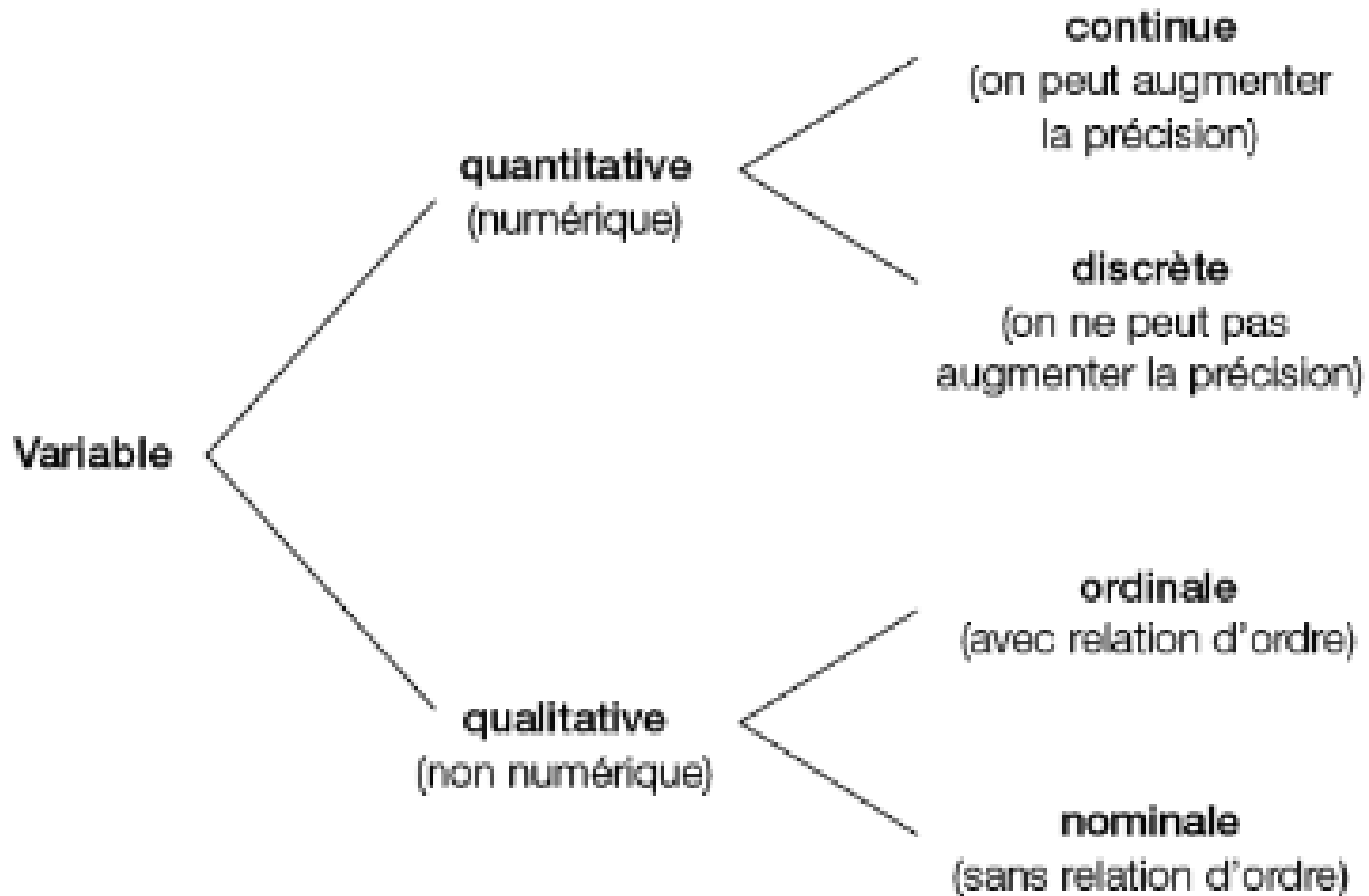
60 000 à 75 000

Exercice:

Dites si la variable est qualitative, quantitative discrète ou quantitative continue.

- Le temps écoulé depuis le début de l'écoulement d'un bassin
- Les nombre d'employés dans l'usine
- Le salaire annuel de chaque employé
- Le module d'assurance choisit par un employé

Types de variables



Statistiques descriptives

A) Présentation tabulaire : Le tableau de distribution de fréquences

C'est un tableau où on présente une compilation des données recueillies lors d'une enquête auprès d'unités statistiques. Il s'agit donc d'un mode d'organisation des données.

Il est intéressant de l'utiliser si la taille de l'échantillon ou de la population étudiée est d'au moins 20 unités statistiques.

Ce type de tableau présente des fréquences, c'est-à-dire le nombre d'unités statistiques correspondant à une modalité de réponse, à une valeur ou à une classe de valeur.

Que doit contenir un tableau
de distribution de fréquences ???



Titre approprié : <i>Répartition des unités statistiques selon la variable d'intérêt</i>		
Nom de la variable	Fréquence absolue	Fréquence relative (%)
Total	N ou n	(100%)

1)

↓

RÉPARTITION EN POURCENTAGE DES JEUNES QUÉBÉCOIS ÂGÉS DE 18 À 35 ANS,



Degré de confiance

Pourcentage des jeunes Québécois (%)

Très confiance

16

Assez confiance

60

Peu confiance

19

Pas du tout confiance

5

Total

100

← *Fréquences relatives*

← Total

2) On a interrogé 30 enseignants de cégep. On leur a demandé le nombre de cours différents qu'ils ont donné durant l'année scolaire 2013-2014.

Série statistique

Nombre de cours différents donnés pendant l'année scolaire 2013-2014									
5	4	5	5	1	2	2	1	5	2
1	2	3	5	4	3	3	5	3	2
2	5	4	3	4	4	3	5	5	2

Tableau de dénombrement

Nombre de cours différents	Dénombrement
1	
2	
3	
4	
5	

Exemple 2) suite...

Tableau de distribution de fréquences

Répartition des enseignants selon le nombre de cours différents donnés en 2013-2014				
Nombre de cours différents	Fréquences	Fréquences cumulées	Fréquences relatives	Fréquences relatives cumulées
1	3	3	10,0%	10,0%
2	7	10	23,3%	33,3%
3	6	16	20,0%	53,3%
4	5	21	16,7%	70,0%
5	9	30	30,0%	100,0%
Total	30		100,0%	

Exercices

Construire un tableau de distribution de fréquences pour représenter les données suivantes :

Variable : Émission de télé préférée

Population : Étudiants du Cégep de Shawinigan

Compilation des données:

Émission A: 20

Émission D: 36

Émission B: 16

Émission E: 46

Émission C: 32

Émission F: 10

Exercices suite...

Titre : _____

Remarque : Pour un %, conserver une seule décimale...

Cas des variables quantitatives discrètes avec 10 valeurs différentes ou plus ou pour les variables quantitatives continues !

On doit construire des **classes** pour regrouper les valeurs sous forme d'intervalles.

+ Pour décrire la première classe, on utilise la notation « $25 \leq X < 30$ », qui signifie que la durée du branchement, variable symbolisée par la lettre X , des clients dans cette classe va de 25 minutes à moins de 30 minutes. On peut aussi utiliser la notation « $[25, 30[$ ». On dit alors que 25 est la limite inférieure de la classe et que 30 en est la limite supérieure.

Comment construire ces classes ???

Répartition des clients selon leur durée de branchement au Wi-Fi		
Durée du branchement (minutes)	Nombre de clients	Pourcentage de clients
$25 \leq X < 30$	2	5,0%
$30 \leq X < 35$	4	10,0%
$35 \leq X < 40$	9	22,5%
$40 \leq X < 45$	12	30,0%
$45 \leq X < 50$	8	20,0%
$50 \leq X < 55$	3	7,5%
$55 \leq X < 60$	2	5,0%
Total	40	100,0%

Comment construire ces classes ???

1. Choisir le nombre de classes (pas trop mais ni trop peu !?)

Règle de Sturges

Nombre observations (n)	Nombre de classes (k)
< 10	4
$10 < n \leq 22$	5
$22 < n \leq 44$	6
$44 < n \leq 90$	7
$90 < n \leq 180$	8
$180 < n \leq 360$	9
$360 < n \leq 720$	10
> 720	11

2. calculer l'étendue « E » des données.

$$E = \text{valeur max} - \text{valeur min}$$

3. on calcule la largeur ou l'amplitude « A » des classes .

$$A \approx \frac{E}{\text{nombre de classes}}$$

On doit ajuster le nombre obtenu pour qu'il soit entier. Idéalement, l'amplitude sera un multiple de 10 ou de 5. La seconde option, est de prendre un chiffre pair et puis, si ce n'est pas possible, on prend un chiffre impair. On choisit la première classe en respectant des règles de cosmétiques semblables

Exemple

Voici les revenus de travail d'été (en \$) de 50 étudiants du Collège

Revenu (\$) du travail d'été de 50 étudiants									
4250	3660	3467	4823	2900	2987	5450	2567	3548	4896
4147	3456	4327	4398	3564	3987	4321	4112	3765	3256
3478	3987	4065	2050	3098	5007	2609	4987	3645	3874
4654	3541	3076	2806	4123	4008	3976	3467	4651	2341
3334	4655	3075	3275	3864	3598	2964	4632	2964	4123

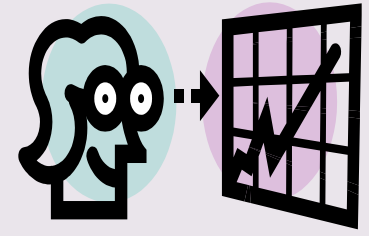
1. Nombre de classes?
2. Étendue?
3. Amplitude?

Exemple suite...

Tableau de distribution de fréquences obtenu...

Répartition des étudiants selon le revenu du travail d'été		
Revenu (\$)	Nombre d'étudiants	Pourcentage d'étudiants
[2000-2500[2	4%
[2500-3000[7	14%
[3000-3500[10	20%
[3500-4000[12	24%
[4000-4500[10	20%
[4500-5000[7	14%
[5000-5500[2	4%
Total	50	100%

B) Présentation des données à l'aide de Graphiques

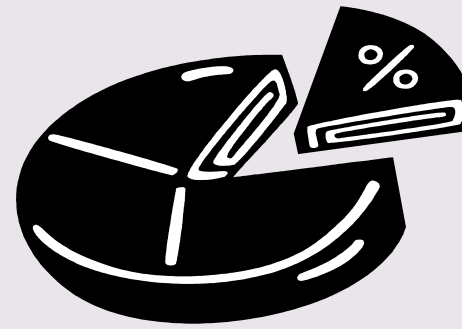


Pour représenter un tableau de distribution de fréquences de façon plus visuelle ...

Types de graphiques utilisés en statistique :

- ❖ diagramme circulaire ou à secteurs
- ❖ diagramme à bandes rectangulaires
- ❖ diagramme à bâtons
- ❖ histogramme
- ❖ polygone de fréquences
- ❖ courbe de fréquences cumulées (ogives)

1) Diagramme circulaire ou à secteurs



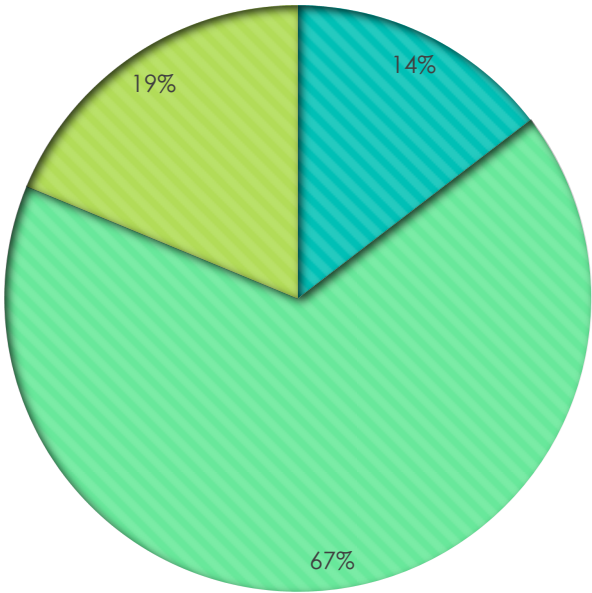
- pour variable qualitative
- fait de secteurs circulaires qui ont une aire proportionnelle à la fréquence des différentes modalités ou valeurs
- identifier chaque secteur et indiquer la fréquence ou la fréquence relative de ce secteur (légende ...)
- donner un titre au diagramme : **le même que celui du tableau à partir duquel on construit le graphique**

Répartition des 48 collèges publics selon la formation prédominante

Formation	Nombre de collèges	Pourcentage de collèges
Préuniversitaire	7	14,6%
Technique	32	66,7%
Aucune	9	18,8%
Total	48	100,0%

Titre

Répartition des 48 collèges publics selon la formation prédominante

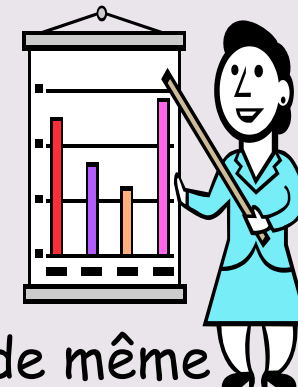


Secteurs

Légende

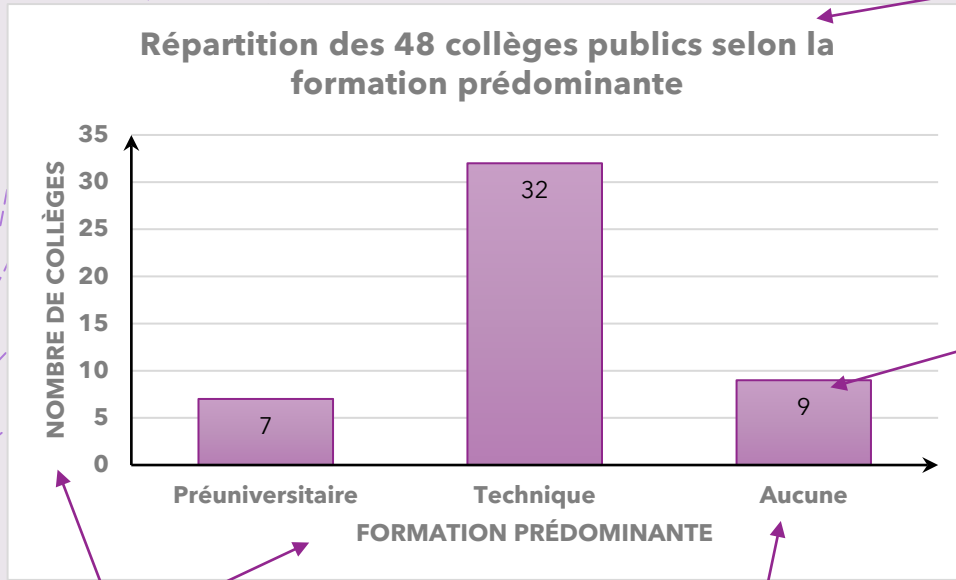
■ Préuniversitaire ■ Technique ■ Aucune

2) Diagramme à bandes rectangulaires



- pour variable qualitative (nominale ou ordinale)
- fait de bandes verticales (ou horizontales) espacées et parallèles de même largeur dont la hauteur est proportionnelle à la fréquence absolue ou relative de chaque modalité
- donner un titre au diagramme: le même que...
- bien identifier les 2 axes
- graduer l'axe des fréquences absolues ou relatives avec le même intervalle (coupure d'axe si nécessaire...)
- identifier sur l'autre axe chacune des modalités de la variable qualitative

Diagramme à bandes verticales



Titre

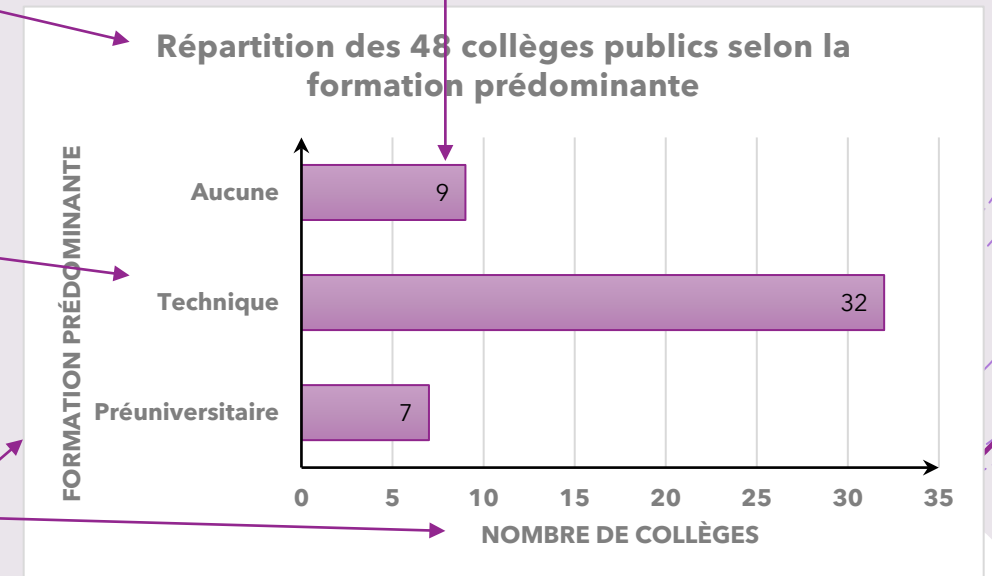
Fréquences / étiquettes de données

Titre des axes

Modalités

Titre

Diagramme à bandes horizontales



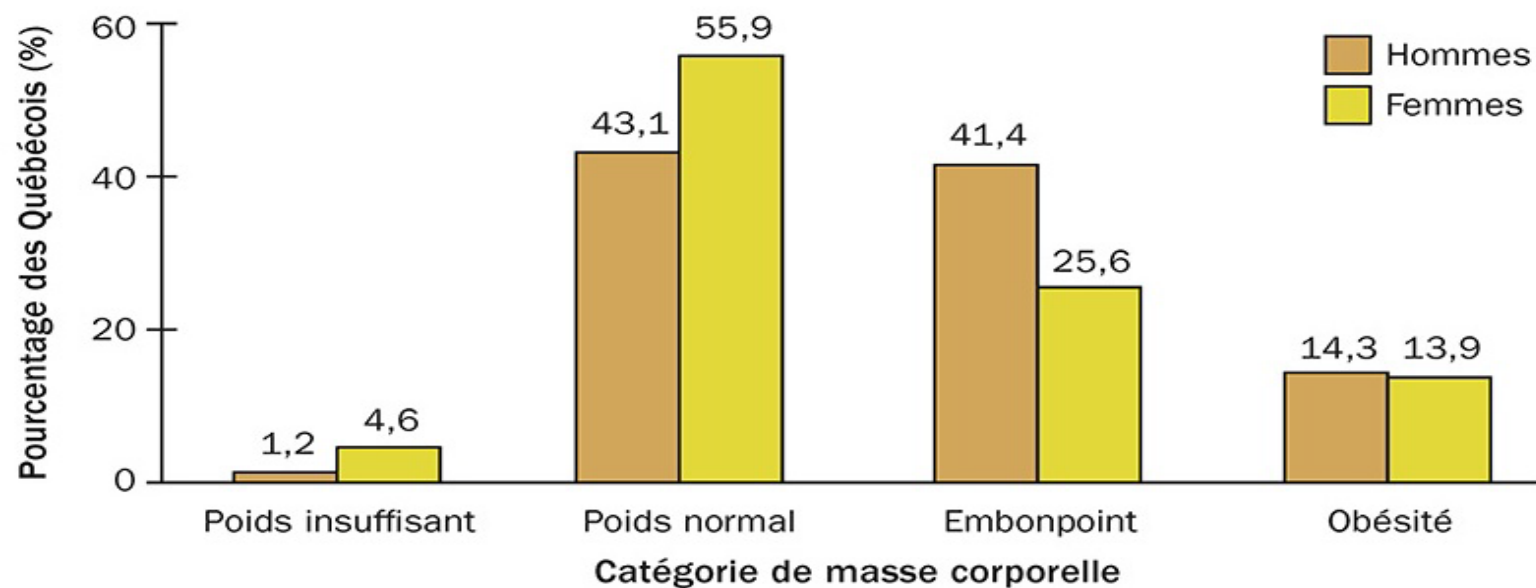
Titre des axes

Exemple 3) à bandes chevauchées

RÉPARTITION EN POURCENTAGE DE LA POPULATION DE 18 ANS ET PLUS, PAR SEXE, SELON LA CATÉGORIE DE MASSE CORPORELLE, QUÉBEC, 2003

Catégorie de masse corporelle	Sexe	
	Hommes (%)	Femmes (%)
Poids insuffisant	1,2	4,6
Poids normal	43,1	55,9
Embonpoint	41,4	25,6
Obésité	14,3	13,9
Total	100,0	100,0

RÉPARTITION EN POURCENTAGE DE LA POPULATION DE 18 ANS ET PLUS, PAR SEXE, SELON LA CATÉGORIE DE MASSE CORPORELLE, QUÉBEC, 2003



3) Diagramme à bâtons

- pour variable quantitative discrète avec moins de 10 valeurs différentes
- fait de bâtons verticaux parallèles dont la hauteur est proportionnelle à la fréquence absolue ou relative de chaque valeur
- donner un titre au diagramme: le même que...
- bien identifier les 2 axes
- identifier toutes les valeurs possibles de la variable sur l'axe des x
- graduer l'axe des y (coupure d'axe si nécessaire)

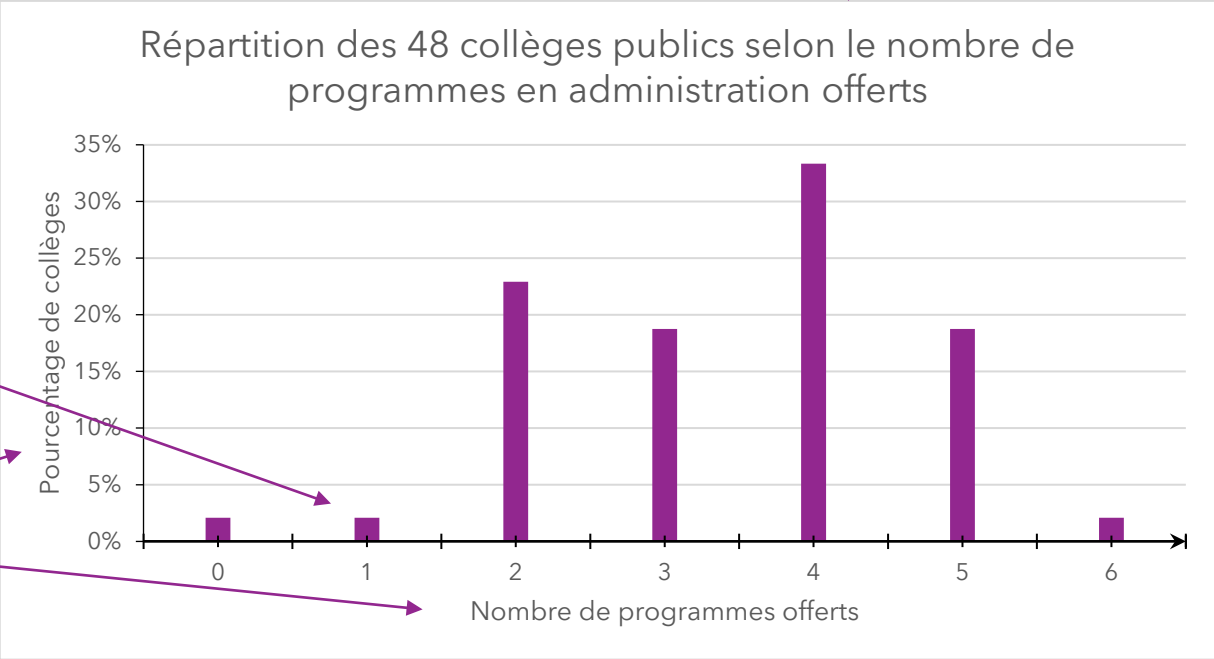
Répartition des 48 collèges publics selon le nombre de programmes en administration offerts

Nombre de programmes offerts	Nombre de collèges	Pourcentage de collèges
0	1	2%
1	1	2%
2	11	23%
3	9	19%
4	16	33%
5	9	19%
6	1	2%
Total	48	100%

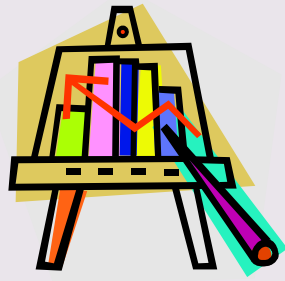
Titre

Bâtons!
Pas des bandes

Titre des axes



4) Histogramme



- pour variable quantitative continue ou variable discrète avec 10 valeurs différentes ou plus
- en résumé, on utilise ce graphique pour les variables qui doivent être groupées en classes
- fait de rectangles verticaux adjacents qui représentent la fréquence absolue (ou relative) de chaque classe
- l'aire de chaque rectangle est proportionnelle à la fréquence de chaque classe

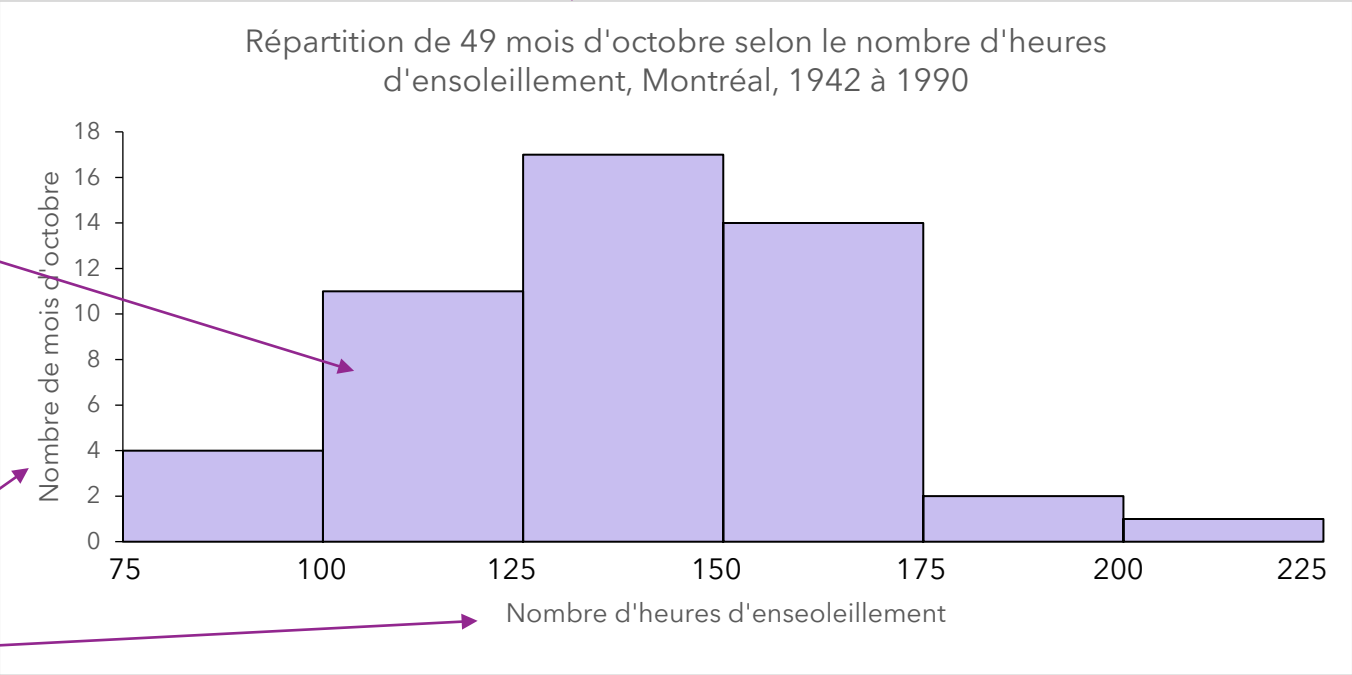
Répartition de 49 mois d'octobre selon le nombre d'heures d'ensoleillement, Montréal, 1942 à 1990

Nombre d'heures d'ensoleillement	Nombre de mois d'octobre
[75, 100[4
[100, 125[11
[125, 150[17
[150, 175[14
[175, 200[2
[200, 225[1
Total	49

Titre

Bandes larges

Titre des axes



5) Polygone de fréquences

- même types de variables que l'histogramme
- ajouter une classe de fréquence « 0 » à chaque extrémité de la distribution
- placer dans un graphique cartésien les points dont les coordonnées sont :
 $(x, y) = (\text{milieu de la classe}, \text{fréquence de la classe})$
- relier tous ces points en prenant bien soin de fermer le polygone de fréquences sur l'axe des x

Répartition de 49 mois d'octobre selon le nombre d'heures d'ensoleillement, Montréal, 1942 à 1990

Nombre d'heures d'ensoleillement	Nombre de mois d'octobre
[75, 100[4
[100, 125[11
[125, 150[17
[150, 175[14
[175, 200[2
[200, 225[1
Total	49

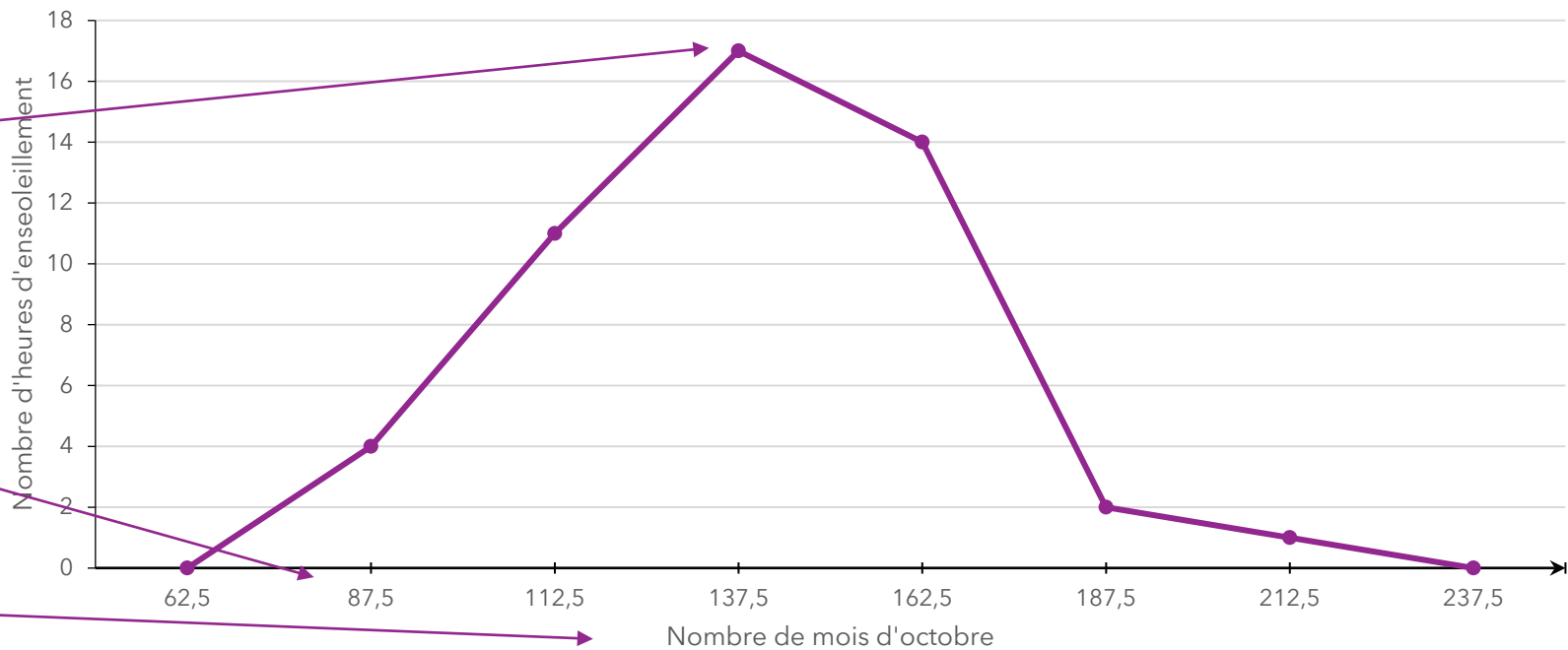
Titre

Coordonnées (x,y)
x : milieu de la classe associée à la variable
y : fréquence de cette classe

Milieu de classe

Titre des axes

Répartition de 49 mois d'octobre selon le nombre d'heures d'ensoleillement, Montréal, 1942 à 1990



6) Courbe de fréquences cumulées (ogive)



→ pour les variables quantitatives continue

→ la courbe part de la limite inférieure de la 1^e classe (x) avec une fréquence de « 0 » (y).

→ on place ensuite dans un graphique cartésien les points dont les coordonnées sont : $(x, y) = (\text{limite supérieure de la classe}, \text{fréquence cumulée})$

→ on relie tous ces points au moyen d'une ligne brisée.

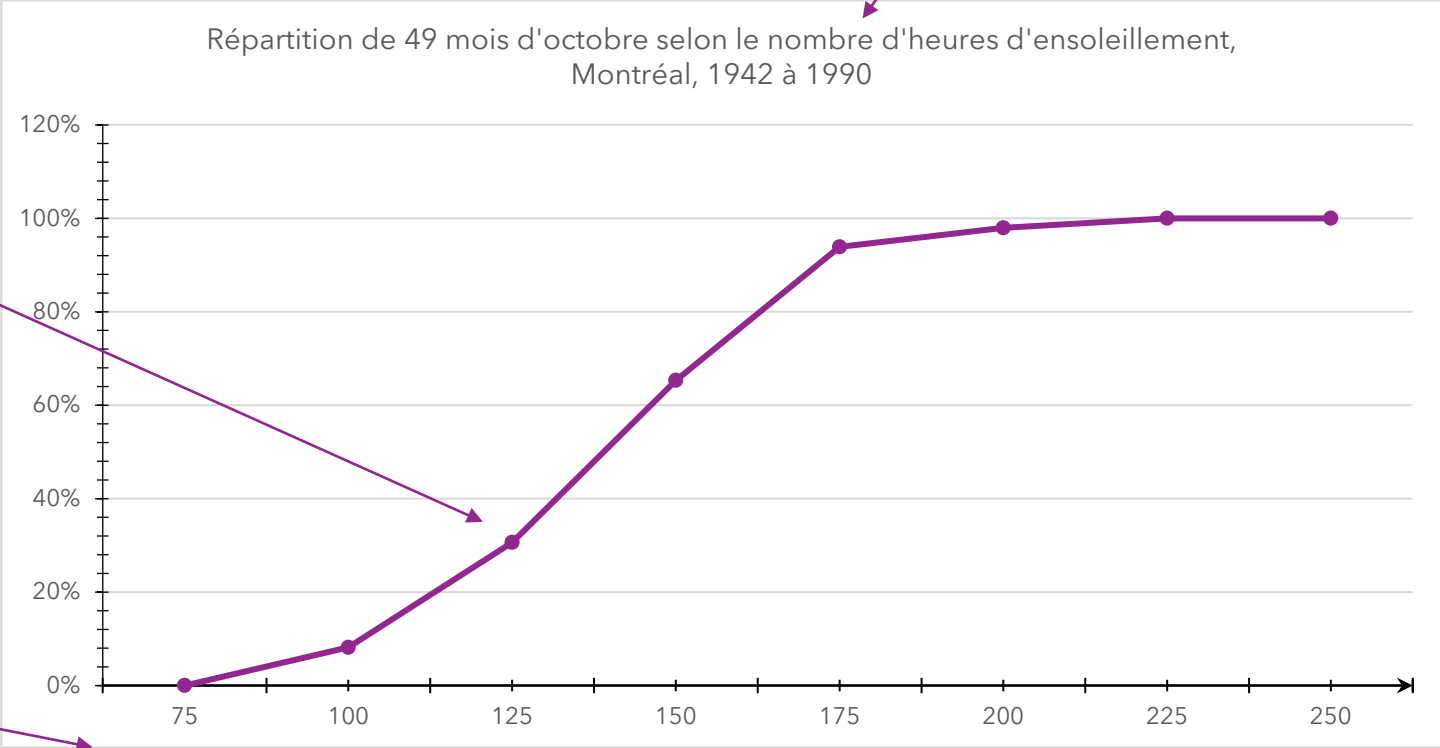
Répartition de 49 mois d'octobre selon le nombre d'heures d'ensoleillement, Montréal, 1942 à 1990

Nombre d'heures d'ensoleillement	Nombre de mois d'octobre	Fréquences relatives	Fréquences relatives cumulées
[75, 100[4	8,2 %	8,2 %
[100, 125[11	22,4 %	30,6 %
[125, 150[17	34,7 %	65,3 %
[150, 175[14	28,6 %	93,9 %
[175, 200[2	4,1 %	98 %
[200, 225[1	2 %	100%
Total	49	100 %	

Titre

Fréquences relatives cumulées associées à la fin des classes

Titre des axes



Résumé: diagramme à utiliser selon la nature de la variable

Variable qualitative: → diagramme circulaire
→ diagramme à bandes

Variable quantitative discrète avec 9 valeurs ou moins:
→ diagramme à bâtons

Variable quantitative discrète avec 10 valeurs et plus (données groupées en classes) et variable quantitative continue:

- histogramme
- polygone de fréquences
- Courbe de fréquences cumulées