

Types de variables

Tableaux de distribution

Représentations graphiques

Élaboré par : Afef Ben Zine El Abidine
Janvier 2021

Présentation du fonctionnement général du cours de statistiques

Manuel obligatoire : Christiane Simard (2015), *Notions de statistiques*, 3^e édition, Groupe Modulo.

Code groupe :
83876

Les statistiques se divisent principalement en deux branches : Les **statistiques descriptives** font le sommaire et simplifient l'information dans le but de décrire un phénomène (**Unité 1**). Par exemple, les statistiques publiées quotidiennement concernant les nouveaux cas de COVID- 19 en Ontario.

La **théorie des probabilités** permet de calculer les risques lors d'une prise de décision dans une situation où le hasard intervient (**Unités 2 et 3**). Par exemple, si le pourcentage de tests positifs de la COVID-19 dans la population est de 5%, quelle est la probabilité d'avoir 3 personnes dont le test est positif parmi un échantillon de 10 personnes prélevées au hasard?

Terminologie

Population: ensemble de toutes les personnes ou tous les objets sur lesquels porte une étude.

Unité statistique: chaque élément de la population

Variable : caractéristique que l'on désire étudier

Échantillon : sous ensemble d'unités de la population

Sondage: un enquête menée auprès d'un échantillon de la population

Recensement: une étude réalisée sur toutes les unités statistiques d'une population

EXEMPLE 1

Dans un sondage, on interroge 1 000 électeurs, choisis au hasard parmi tous les électeurs du Québec, afin de connaître leur intention de vote.

Population étudiée : Tous les électeurs du Québec.

Unité statistique : Un électeur.

Échantillon : Les 1 000 électeurs choisis au hasard.

Variable : intention de vote

Exemple pour illustrer le vocabulaire de la statistique

Un sondage a été effectué afin de connaître la couleur des yeux des nouveau-nés journaliers dans un service de maternité. Voici les données collectées.

Bleu	Vert	Marron	Noir	Gris
5	3	11	10	1

- a) Comment appelle-t-on ce tableau? Une **distribution statistique**
- b) Quelle est la **population étudiée** ? Les nouveau-nés dans un service de maternité
- c) Quelle est l'**unité statistique** ? Le nouveau-né
- d) Quel est l'**effectif total de l'échantillon** ? L'effectif total est 30.
- e) Quel est le **caractère** étudié ? La couleur des yeux
- f) Quelles sont les **modalités** du caractère étudié? Ce sont les valeurs (possibles) du caractère: bleu, vert, etc.

D'autres exemples sont proposés p. 2-3 du *manuel*

Les types de variables

Une variable est une caractéristique de l'unité statistique que l'on désire étudier.

On distingue les variables **qualitatives** (modalités = mots) et **quantitatives** (modalités= nombres)

Les variables qualitatives peuvent être: **Nominales ou Ordinales**

Les variables quantitatives peuvent être: **Discrètes ou Continues.**

Variables qualitatives nominales : indiquer (nommer) la catégorie à laquelle chaque observation appartient.

Exemples: Sexe (masculin, féminin), Fumeur (oui, non), Pays (Canada, Australie, Italie, etc), Fréquence d'utilisation (Hebdomadaire, mensuelle, quotidienne), etc.

D'autres exemples sont proposés p. 4 du *manuel*.

Les types de variables - suite

Variables qualitatives ordinales : les modalités peuvent s'ordonner.

Exemples: l'attitude de la personne envers un sujet (pas du tout d'accord/ pas d'accord/ d'accord/ tout à fait d'accord), la satisfaction à l'égard d'une marque (pas satisfait/Satisfait/Très satisfait), etc.

Variables quantitatives discrètes: les valeurs sont des nombres précis, pour la plupart des entiers.

Exemples: le nombre d'enfants par famille, le nombre de films regardés, le nombre de programmes proposés par la Cité, etc.

Variables quantitatives continues : les valeurs sont exprimées par des intervalles de nombres réels.

Exemples: le poids d'une personne, le pourcentage des filles dans la Cité, le temps nécessaire pour effectuer un trajet, etc.

D'autres exemples sont proposés p. 5 du *manuel*.

Exercice 1: identifier le type de chaque variable:

1. Distance parcourue	5. Nombre de bonnes réponses	9. Nombre d'enfants
2. Couleur préférée	6. Nombre moyen d'enfants par famille	10. Nombre de buts de marqués
3. Taille d'un individu	7. Taux de chômage	11. Temps de téléchargement
4. Nationalité d'une personne	8. État civil (marié, veuf, conjoint de fait, divorcé, célibataire)	12. Nombre de sites Web consultés

Qualitative nominale

Qualitative ordinale

Quantitative discrète

Quantitative continue

Les échelles de mesure

L'échelle nominale implique un simple groupement des observations en catégories qualitatives identifiées par un symbole (souvent une étiquette, tel « Homme » et « Femme » pour identifier le sexe).

L'échelle ordinale est similaire à l'échelle nominale exceptée qu'elle permet d'établir une relation d'ordre entre les éléments d'un ensemble, sans toutefois être capable d'évaluer de façon quantitative la distance qui les sépare. Exemple: Aimez-vous les études?

1. Un peu
2. Moyennement
3. Beaucoup

L'échelle d'intervalle. Elle s'applique uniquement aux variables quantitatives. Cette échelle possède une unité de mesure arbitraire mais constante. Un exemple est la température exprimée en celsius. Zéro Celsius est un point arbitraire qui a été choisi par convention. Cette échelle de mesure ne permet pas d'affirmer que de l'eau à 10 Celsius est deux fois plus chaude que de l'eau à 5 Celsius. L'écart entre deux valeurs a un sens mais pas le rapport entre celles-ci.

L'échelle de rapport: en plus de permettre de quantifier la différence entre deux éléments, elle permet aussi de calculer des rapports entre deux mesures. Par exemple, une distance de 4 mètres est bien le double d'une distance de 2 mètres. L'écart et le rapport entre deux valeurs ont un sens.

D'autres exemples sont proposés p. 8 du *manuel*.

Comparaison des échelles de mesure

Échelle	Caractéristiques spécifiques	Opérations permises
Nominale	<ul style="list-style-type: none"> Requiert une variable qualitative nominale. Permet uniquement de différencier les catégories ($=$, \neq). 	Aucune
Ordinale	<ul style="list-style-type: none"> La variable peut être qualitative ou quantitative. Permet de différencier et d'ordonner les catégories ou les valeurs ($=$, \neq, $<$, $>$). 	Aucune
D'intervalle	<ul style="list-style-type: none"> Requiert une variable quantitative. Le zéro est fixé par convention : il n'indique pas l'absence de la caractéristique. Permet de différencier et d'ordonner les valeurs ($=$, \neq, $<$, $>$). L'écart entre deux valeurs a un sens, mais pas le rapport entre celles-ci. 	$+$, $-$
De rapport	<ul style="list-style-type: none"> Requiert une variable quantitative. Le zéro est absolu : il indique l'absence de la caractéristique. Permet de différencier et d'ordonner les valeurs ($=$, \neq, $<$, $>$). L'écart et le rapport entre deux valeurs ont un sens. 	$+$, $-$, \times , \div

Exemples, manuel, p. 9

1. Pour chacune des questions suivantes, donner le type de la variable étudiée et l'échelle de mesure.

a) Êtes-vous marié(e)? 1. Oui 2. Non

Type : qualitative nominale. Échelle nominale.

b) En quelle année vous êtes-vous marié(e)?

Type : quantitative discrète. Échelle d'intervalle.

c) Depuis combien de temps êtes-vous marié(e)?

Type : quantitative continue. Échelle de rapport.

EXERCICE 2: Extrait du chapitre 1 du *manuel* (p.10)

Donner le **type de variable** et de **l'échelle de mesure**

Variable (question)	Échelle de mesure
1. Avez-vous échoué à des cours à votre premier trimestre?	1. Oui 2. Non
2. À combien de cours avez-vous échoué au premier trimestre?	1. 0 2. 1 ou 2 3. 3 ou plus
3. À combien de cours avez-vous échoué au premier trimestre ?	
4. Quel est votre taux d'échec au premier trimestre?	1. 0% 2. 1% à 15,9% 3. 16% à 49,9% 4. 50% et plus
5. Quelle est votre année de naissance?	

Le tableau de distribution ou tableau de fréquence

Exemple :

On a réalisé une enquête sur l'avenue des Érables. Dans chacune des 20 maisons, on a demandé aux gens d'indiquer le **nombre de véhicules** immatriculés dans leur ménage. Voici les résultats enregistrés :

1, 2, 1, 0, 3, 4, 0, 1, 1, 1, 2, 2, 3, 2, 3, 2, 1, 4, 0, 0 (Données brutes= **Série statistique**)

Répartition des ménages habitant sur l'avenue des Érables selon le nombre de véhicules

Nombre de véhicules	Nombre de ménage (effectif)
0	4
1	6
2	5
3	3
4	2

Tableau tiré du site Statistique Canada

Fréquence absolue ou l'effectif : Nombre de fois qu'un événement se produit ou qu'un élément se présente dans un ensemble de données.

Exercice (p. 20)

EXERCICE DE COMPRÉHENSION | 1.2

Une étude indique que 200 accidents sont survenus sur une autoroute durant l'année : 160 n'ont fait aucune victime, 32 ont fait une victime, 6 ont fait deux victimes et 2 ont fait trois victimes. Parmi les 50 victimes, 30 ont eu des blessures légères, 16 ont eu des blessures graves et 4 sont décédées.

a) Les tableaux suivants présentent les résultats de l'étude. Donner un titre aux tableaux.

Tableau 1

Titre : _____

Nombre de victimes	Nombre d'accidents	Pourcentage d'accidents
0	160	80 %
1	32	16 %
2	6	3 %
3	2	1 %
Total	200	100 %

Le tableau de distribution de fréquence relative

La **fréquence relative** d'une catégorie (ou d'une valeur) particulière est la proportion de la catégorie. Ainsi :

Fréquence relative = fréquence absolue ÷ nombre total d'observations

Répartition des ménages habitant sur l'avenue des Érables selon le nombre de véhicules

Nombre de véhicules	Fréquence absolue	Fréquence relative
0	4	4/20=20%
1	6	6/20=30%
2	5	25%
3	3	15%
4	2	10%

La structure du titre du tableau de distribution :

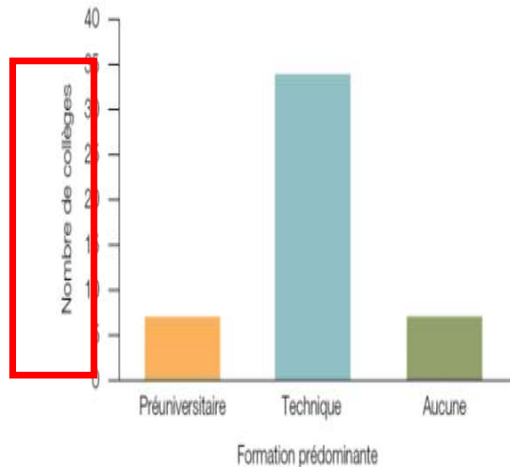
Répartition de(Unité statistique) selon(Caractère étudié)

Représentations graphiques –variables qualitatives

Lorsque le caractère statistique est une variable **qualitative** : on construit un **diagramme circulaire**, le **digramme linéaire** ou le **diagramme à rectangles verticaux ou horizontaux** (explications détaillées p. 14-16)

Diagramme à rectangles verticaux

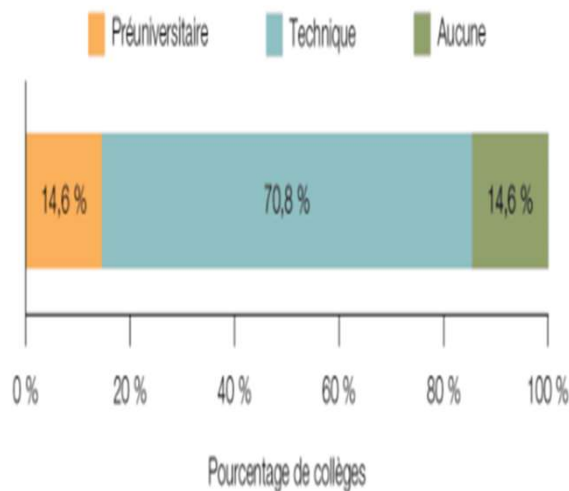
Répartition des 48 collèges publics selon la formation prédominante, Québec, 2010



Source: Ministère de l'Éducation, du Loisir et du Sport. Réseaux, DSID Portail Informationnel, système Socrate, données au 25 décembre 2012.

Diagramme linéaire

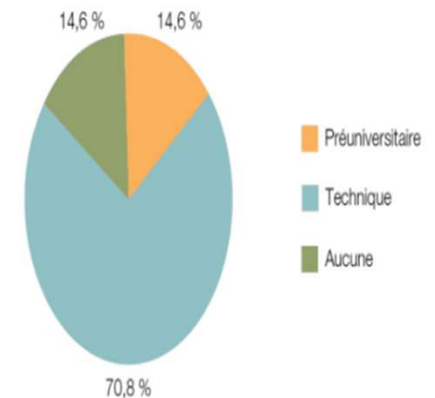
Répartition des 48 collèges publics selon la formation prédominante, Québec, 2010



Source: Ministère de l'Éducation, du Loisir et du Sport. Réseaux, DSID Portail Informationnel, système Socrate, données au 25 décembre 2012.

Diagramme circulaire

Répartition des 48 collèges publics selon la formation prédominante, Québec, 2010

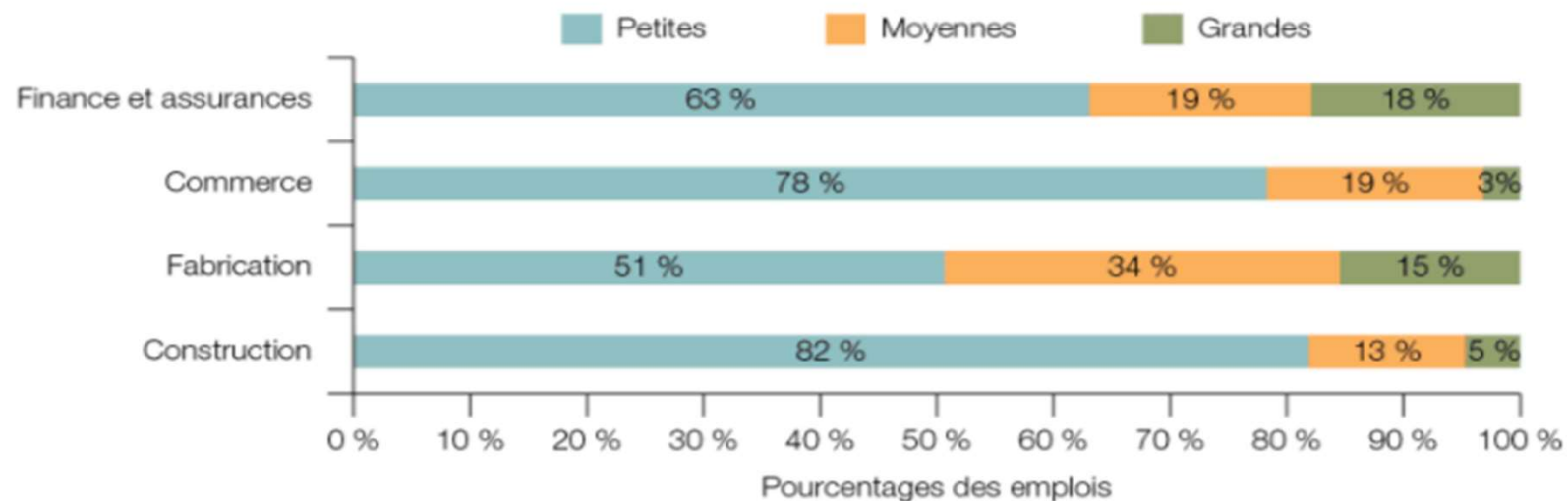


Source: Ministère de l'Éducation, du Loisir et du Sport. Réseaux, DSID Portail Informationnel, système Socrate, données au 25 décembre 2012.

Exemple (p. 17)

On divise les entreprises en trois groupes: les petites entreprises (moins de 100 employés), les moyennes entreprises (de 101 à 499 employés) et les grandes entreprises (500 employés et plus). On dit que ce sont les petites entreprises qui créent le plus d'emplois. Est-ce vrai? Pour répondre à cette question, analyser les statistiques suivantes.

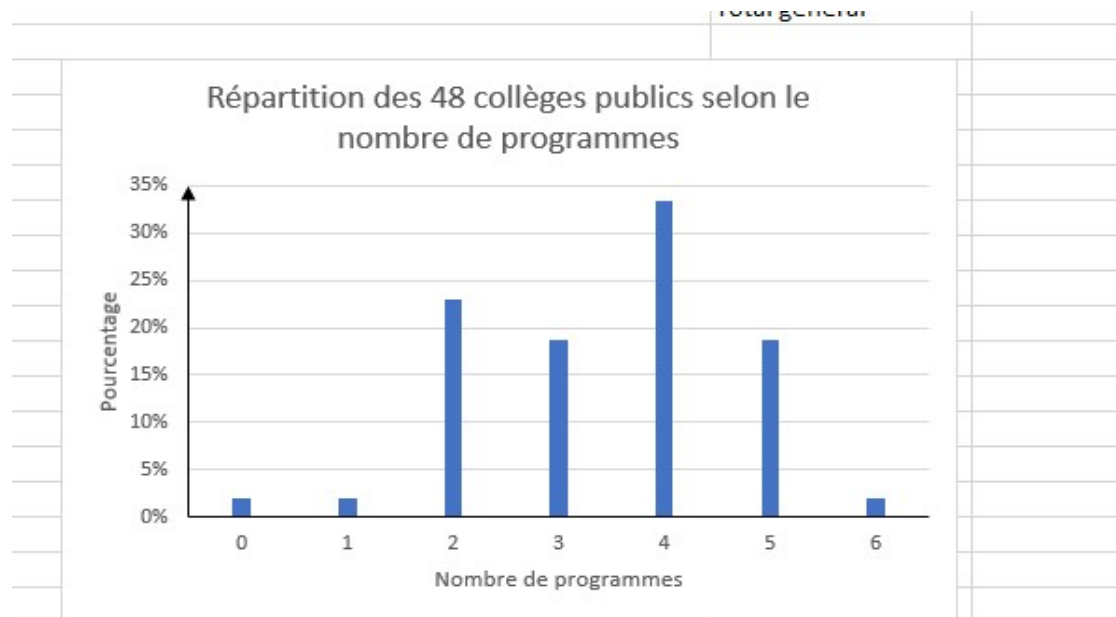
Répartition des emplois selon la taille des entreprises pour certains secteurs d'activité, Québec, 2010



Source: Industrie Canada. *Principales statistiques relatives aux petites entreprises* – Août 2013.

Représentations graphiques-variables quantitatives discrète

Le **diagramme en bâtons** est utilisé pour représenter les séries statistiques correspondant à un caractère **discret** (discontinu). La longueur de chaque bâton est proportionnelle **au pourcentage ou la fréquence relative** de la valeur de la variable.



Représentations graphiques-variable quantitative continue

Lorsque le caractère statistique est une variable quantitative **continue** et les valeurs sont regroupées en **classe** : on construit un **histogramme**. On accompagne le graphique de l'histogramme d'un **polygone de fréquence** qui relie les centres de classe des colonnes de l'histogramme (utile pour trouver le modèle mathématique de la distribution).

Important : L'histogramme est composé de **colonnes accolées** de hauteurs variables. On ne laisse pas d'espace entre les colonnes d'un histogramme afin d'indiquer qu'il s'agit de l'étude d'une variable continue. À ne pas confondre avec un diagramme à rectangles verticaux !

Procédure de construction de l'histogramme (manuel p. 24)

- Fixer le nombre de classes avec la table de Sturges. Elle repose sur la formule $1+3,322 \log n$ (n: nombre de données)
- Calculer **l'étendue** de la série ($E = X_{\max} - X_{\min}$)
- Calculer **l'amplitude** de la classe, avec $\text{amplitude} = \text{étendue} / \text{nombre de classes}$

Procédure de construction de l'histogramme-mise en situation (p.21 du manuel)

MISE EN SITUATION

Un café offre sur demande un code d'accès à un réseau Wi-Fi à ses clients. Toutefois, afin de s'assurer qu'il y ait toujours des tables disponibles pour la clientèle, la durée du branchement au réseau est limitée à 60 minutes. Pour vérifier si, malgré cette contrainte, le temps alloué répond aux besoins des clients, on note la durée (en minutes) du branchement au réseau d'un échantillon de 40 clients. Voici la série statistique obtenue :

48	35	29	44	42	52	43	38	40	47
30	56	32	49	40	<u>59</u>	37	39	40	46
53	37	48	45	46	42	43	35	33	51
<u>26</u>	45	41	41	34	38	43	41	38	35

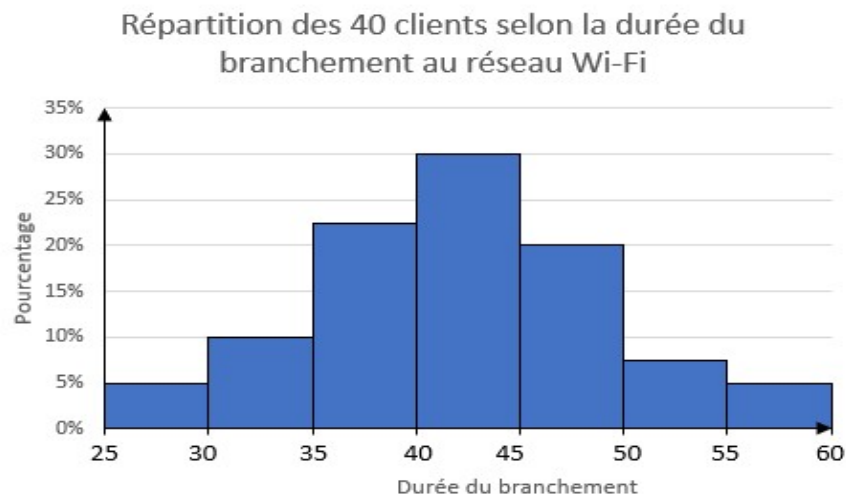
À partir de la série statistique suivante, détermine :

1. Le nombre de classes selon la table de Sturges
2. L'étendue de la série.
3. L'amplitude des classes

Représentations graphiques-variable quantitative continue: Histogramme

Répartition des 40 clients selon la durée du branchement au réseau Wi-Fi

Durée du branchement	Nombre de clients	Pourcentage
25-29	2	5,0%
30-34	4	10,0%
35-39	9	22,5%
40-44	12	30,0%
45-49	8	20,0%
50-54	3	7,5%
55-60	2	5,0%
Total général	40	



Exercice 2: indique la représentation graphique appropriée

Âge	Fréquence absolue
[25-35]	7
]35-45]	5
]45-55]	6
]55-65]	3
]65-75]	3
]75-85]	4
]85-95]	2

Films regardés par semaine	Fréquence relative
0	20%
1	10%
2	30%
3	15%
4 et plus	25%

Films regardés	Taux de visionnement
Le jeu	40%
Le voyage au Groenland	15%
Jour J	25%
Le chant du loup	20%

Important : il faut calculer la fréquence **rectifiée** (ou l'effectif rectifié) si les classes ont des amplitudes inégales avant de construire l'histogramme.

Exemple d'une distribution avec des classes à amplitudes inégales: répartition des élèves d'une garderie selon leur poids

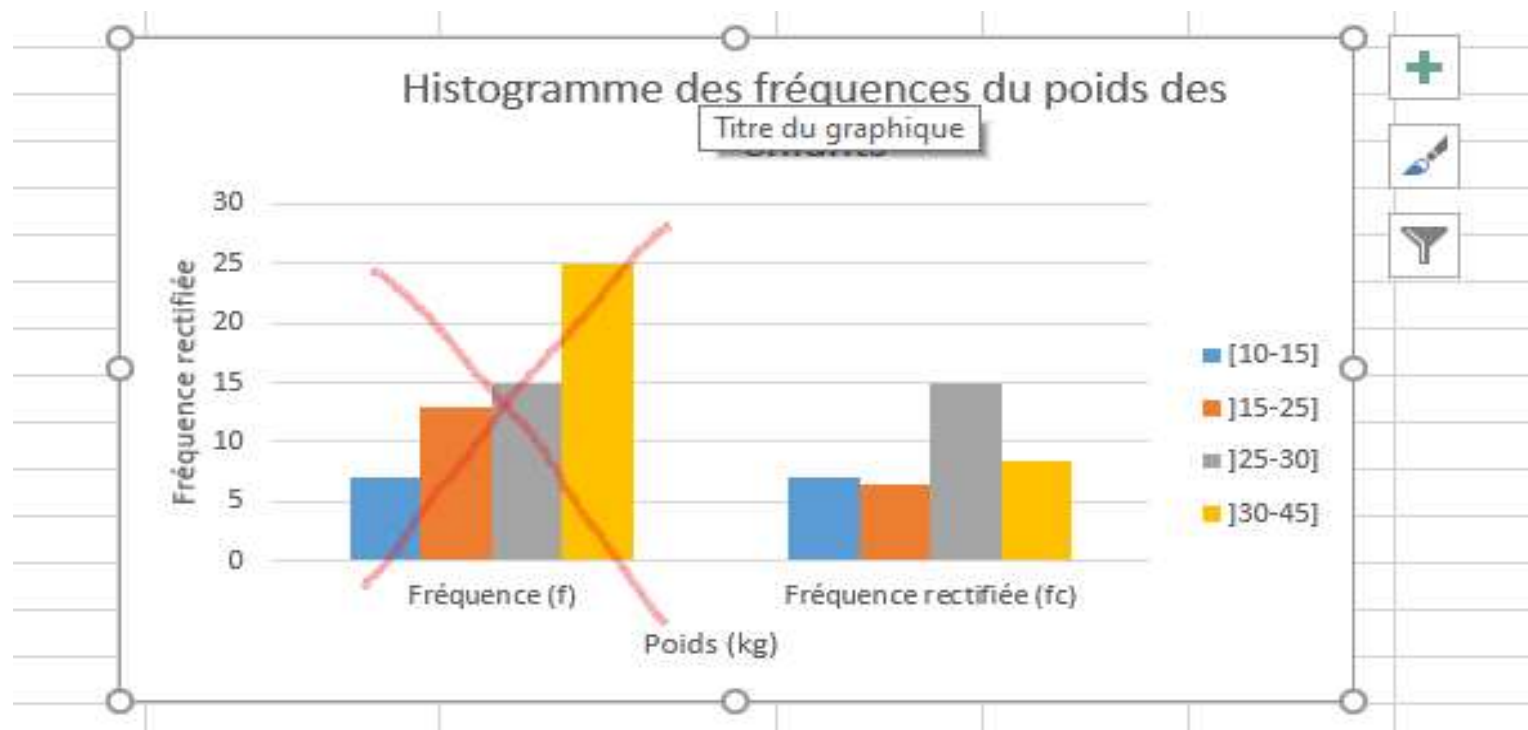
Pour construire un histogramme lorsque les amplitudes de classes sont inégales, il faut:

1. Déterminer l'amplitude de chaque classe avec **amplitude** = $X_{\max} - X_{\min}$
2. Calculer la densité de chaque classe avec **densité** = **fréquence absolue/amplitude**
3. Calculer la fréquence rectifiée pour l'amplitude choisie a^* avec **fréquence rectifiée** = $a^* \times$ **densité de la classe**

Répartition des élèves d'une garderie selon leur poids

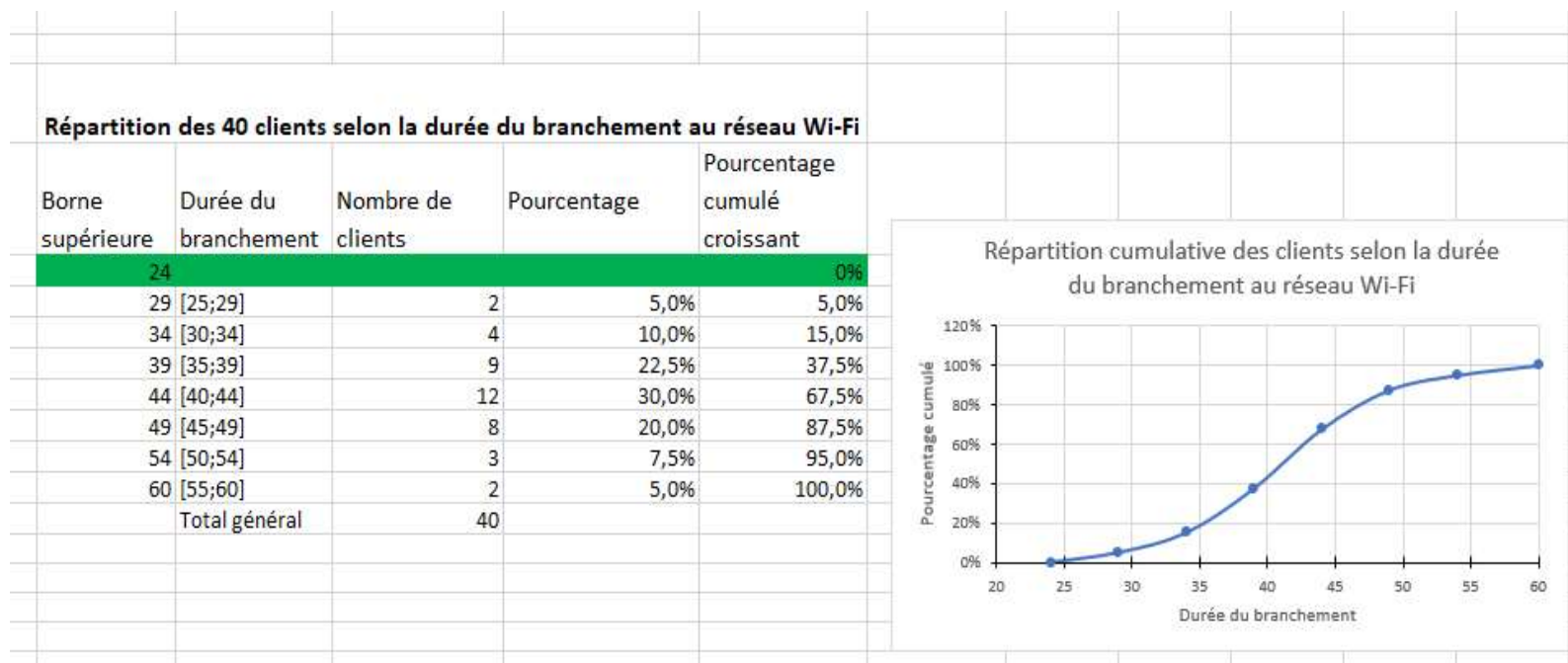
Poids	Fréquence absolue	Amplitude	Densité (d)	Fréquence rectifiée (fc) $a^*=5$
[10-15]	7	15-10= 5	7/5 = 1,4	1,4 x 5=7
]15-25]	13	25-15= 10	13/10 = 1,3	1,3 x 5 =6,5
]25-30]	15	5	15/5 = 3	3 x 5 = 15
]30-45]	25	15	25/15 = 1,66	1,66 x 5= 8,33
Total	60			

Histogramme avec fréquences absolue non rectifiées versus fréquences rectifiées en cas de classes d'amplitudes inégales



Construire la courbe de fréquences cumulée croissante (appelée aussi Ogive) pour une variable quantitative continue

Pour chaque classe, on détermine un point (x,y) dans le plan cartésien où x est la **limite supérieure de la classe** et y la **fréquence cumulée de cette classe**.



Interprétation du tableau de distribution des fréquences cumulées croissantes (manuel, p.32)

Un tableau de distribution de fréquences cumulées nous permet de connaître le pourcentage de données d'une distribution inférieure à une certaine valeur. Elle est la somme des fréquences relatives des valeurs inférieures ou égales à une classe x_i exprimée en **pourcentage**.

Répartition des infractions pour excès de vitesse selon l'âge du conducteur, Québec, 2011

Âge du conducteur	Pourcentage d'infractions	Pourcentage cumulé
Moins de 25 ans	16,6 %	16,6 %
[25 ans ; 35 ans[21,3 %	37,9 %
[35 ans ; 45 ans[21,0 %	58,9 %
[45 ans ; 55 ans[21,0 %	79,9 %
[55 ans ; 65 ans[13,1 %	93,0 %
65 ans et plus	7,0 %	100,0 %
Total	100,0 %	

Source: Société de l'assurance automobile du Québec. Les infractions et les sanctions reliées à la conduite d'un véhicule routier, 2002-2011, 2012.

Analyse des données

En 2011, près de 60 % des infractions pour excès de vitesse sont commises par des conducteurs de moins de 45 ans. Si l'on ajoute celles commises par les conducteurs de 45 à 55 ans, on atteint 80 % des infractions. Seulement 7 % des infractions pour excès de vitesse sont commises par des conducteurs de 65 ans et plus.

Travail à faire

Travail obligatoire :

- Exercice 1, 2, 3, 4, 7 et 8, p. 35-37 du manuel.

Remarques générales pour les exercices

1. Pour les calculs effectués manuellement dans les tableaux de distribution (fréquence relative, fréquences cumulées et autres), utiliser la recommandation du *Manuel* p.13. Sur Excel, spécifier uniquement le nombre de décimal à 1.

NOTES

- Dans un tableau de distribution, la deuxième ou la troisième colonne peut être absente.
- Dans cet ouvrage, nous conviendrons de conserver une seule décimale dans le calcul des pourcentages du tableau de distribution. Nous nous permettrons toutefois d'arrondir les pourcentages à l'entier dans le texte d'analyse des données.
- Pour arrondir les pourcentages, nous appliquerons la règle suivante : si la deuxième décimale est 5 ou plus, on ajoute 1 à la première décimale. Par exemple, 6,38 % sera arrondi à 6,4 %, alors que 10,63 % sera arrondi à 10,6 %.

2. Les critères importants à considérer dans la réalisation des exercices et des travaux :

- **Complétude des graphiques : présence des axes et leur spécification**
- **Formulation exacte des titres des tableaux et des graphiques**
- **Utilisation des termes justes pour la formulation des phrases dans l'analyse et l'interprétation des données**

Sur Excel :

- ✓ Réalisation d'un tableau de distribution à partir d'une série statistique.
- ✓ Construction du diagramme approprié. Par exemple, un histogramme et le polygone de fréquences, un diagramme en bâton, etc.
- ✓ Construction de la courbe de fréquences cumulées croissantes.

Annexe: table de Sturges pour déterminer le nombre de classes d'un histogramme (manuel, p. 23)

Table de Sturges

Nombre de données	Nombre approximatif de classes ¹
Entre 10 et 22	5
Entre 23 et 44	6
Entre 45 et 90	7
Entre 91 et 180	8
Entre 181 et 360	9
Entre 361 et 720	10

1. Ce nombre est déterminé par la formule $1 + 3,322 \log n$, où n est le nombre de données.