Chapitre: 01

Distributions statistiques et représentations associées

CHAPITRE 1 : Distributions statistiques et représentations associées

1. Introduction

En zootechnie comme en épidémiologie, la première étape de toute étude quantitative consiste à recueillir des données sur un ensemble d'individus : animaux, élevages, échantillons biologiques. Ces données, une fois collectées, se présentent sous forme brute : une liste de valeurs, souvent longue et difficile à interpréter directement. La statistique descriptive a précisément pour objet de **résumer**, **organiser** et **représenter** ces données afin d'en dégager les principales caractéristiques.

Ce chapitre pose les fondations de toute analyse statistique. Il définit le vocabulaire de base, distingue les différents types de variables, montre comment construire une distribution d'effectifs et de fréquences, et présente les représentations graphiques adaptées à chaque situation. Les exemples choisis sont issus de situations concrètes en production animale et en santé animale, afin de montrer l'utilité immédiate de ces outils pour l'ingénieur ou le vétérinaire.

Exemple : Un technicien d'élevage ou vétérinaire relève sur 120 vaches la race, le numéro de lactation, la production laitière journalière et le statut mammite (oui/non). Avant de calculer une moyenne ou de tester une relation, il doit savoir de quel type sont ces variables, comment les organiser dans un tableau et comment les représenter graphiquement.

Objectifs du chapitre

À la fin de ce chapitre, l'étudiant sera capable de :

- Maîtriser le vocabulaire de base de la statistique descriptive appliquée à la zootechnie et à l'épidémiologie.
- Distinguer les différents types de variables statistiques.
- Construire un tableau de distribution d'effectifs et de fréquences.
- Choisir la représentation graphique adaptée au type de variable.
- Interpréter une distribution de données animales ou sanitaires.

Préambule méthodologique

En zootechnie comme en épidémiologie, l'observation produit rapidement un grand nombre de données : poids, productions, races, statuts sanitaires, traitements, mortalités, etc. Brutes, ces données ne disent rien. La statistique descriptive est

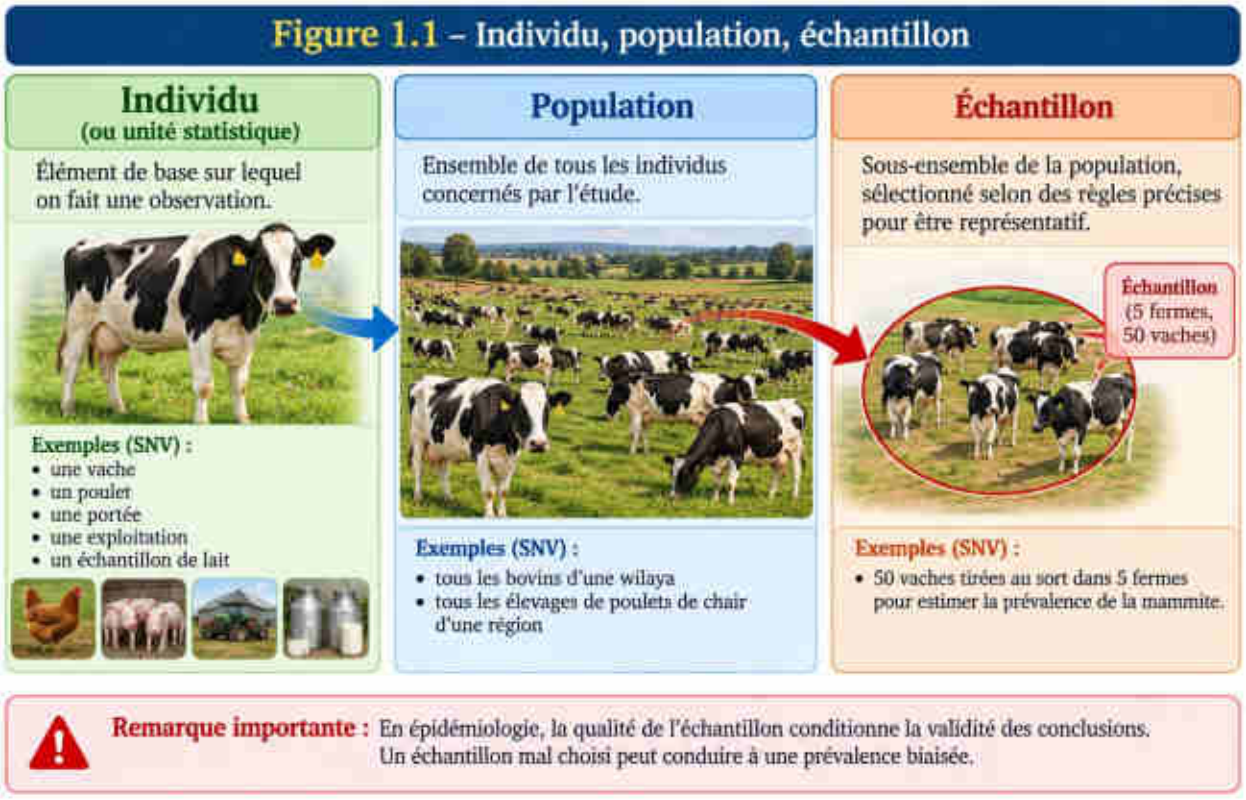
l'ensemble des méthodes qui permettent de **résumer**, **organiser** et **représenter** ces données avant toute analyse plus poussée.

1.1. Définitions de base

1.1.1. Individu, population, échantillon

- **Individu (ou unité statistique)** : élément de base sur lequel on fait une observation.
    - *Exemples SNV* : une vache, un poulet, une portée, une exploitation, un échantillon de lait.
- **Population** : ensemble de tous les individus concernés par l'étude.
    - *Exemples SNV* : tous les bovins d'une wilaya, tous les élevages de poulets de chair d'une région.
- **Échantillon** : sous-ensemble de la population, sélectionné selon des règles précises pour être représentatif.
    - *Exemples SNV* : 50 vaches tirées au sort dans 5 fermes pour estimer la prévalence de la mammite.

Remarque importante : En épidémiologie, la qualité de l'échantillon conditionne la validité des conclusions. Un échantillon mal choisi peut conduire à une prévalence biaisée.



1.1.2. Variable statistique

On appelle **variable** (ou **caractère statistique**) l'aspect que l'on étudie sur chaque individu. Elle prend différentes valeurs appelées **modalités**.

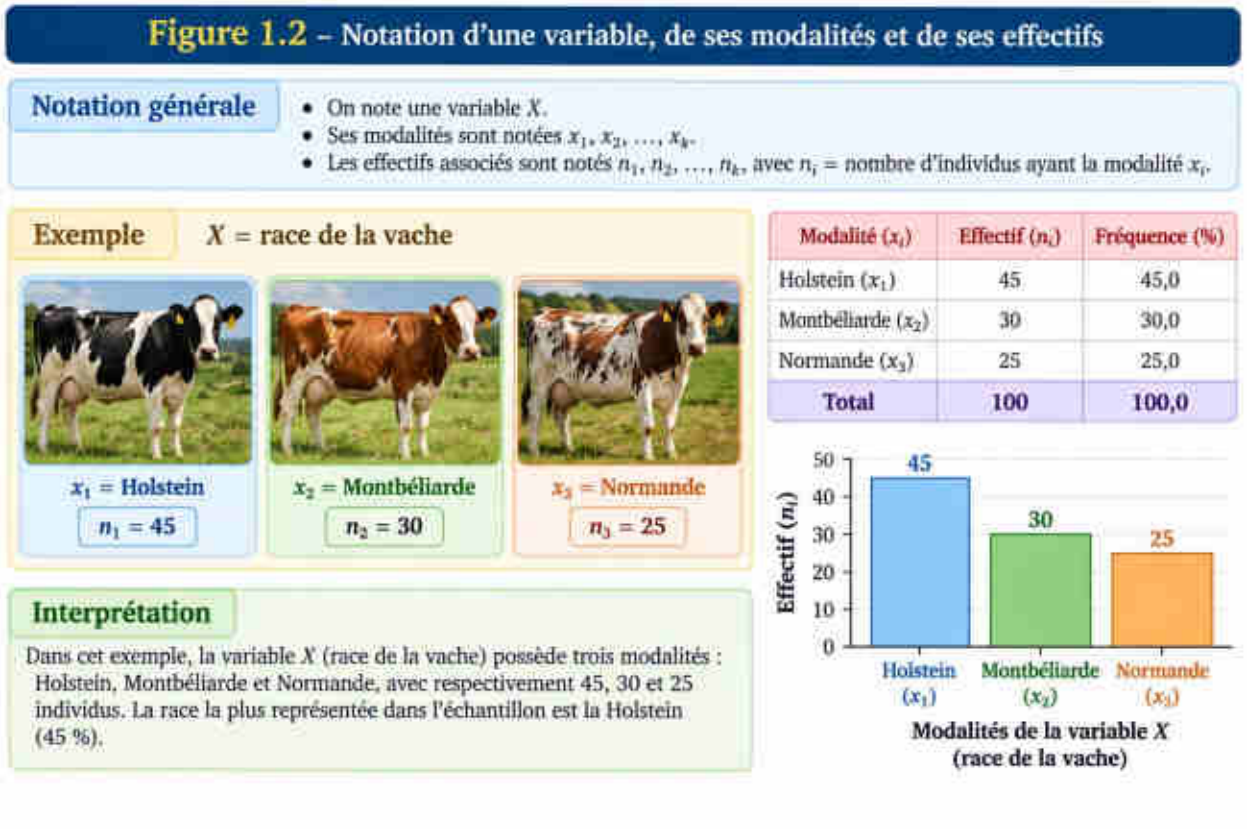
- Notation : on note une variable X, ses modalités $x1,x2,\ldots,xk$, et les effectifs associés $n1,n2,\ldots,nk$.

Exemple :

X = race de la vache ;

$x1$ = Holstein, $x2$ = Montbéliarde, $x3$ = Normande ;

$n1$=45, $n2$=30, $n3$=25.



Présentation du tableau 1.1 : Ce tableau récapitule les notions fondamentales de la statistique descriptive et les traduit immédiatement dans le contexte d'une étude zootechnique ou épidémiologique. Il sert de référentiel avant toute collecte de données.

Tableau 1.1 – Vocabulaire de base appliqué à un élevage

Terme statistique	Définition générale	Exemple en zootechnie / épidémiologie
Individu	Unité d'observation	Une vache de l'élevage
Population	Ensemble des individus	Les 200 vaches de l'exploitation
Échantillon	Sous-ensemble de la population	50 vaches tirées au sort
Variable	Caractère étudié	Statut mammite
Modalité	Valeur possible de la variable	Oui / Non

Effectif	Nombre d'individus par modalité	12 vaches mammiteuses

1.2. Types de variables statistiques

1.2.1. Variables qualitatives

Elles expriment une qualité, une catégorie. On distingue :

- **Variables nominales** : catégories sans ordre.
  - *Exemples* : race, sexe, espèce, statut vaccinal.
- **Variables ordinales** : catégories avec un ordre logique.
  - *Exemples* : note d'état corporel (1 à 5), stade de gestation (début, milieu, fin), score de boiterie.

1.2.2. Variables quantitatives

Elles expriment une quantité mesurable. On distingue :

- **Variables discrètes** : valeurs entières, souvent issues d'un comptage.
  - *Exemples* : nombre de lactations, nombre de traitements antibiotiques.
- **Variables continues** : valeurs réelles, issues d'une mesure.
  - *Exemples* : poids vif (kg), production laitière (L/j), température rectale (°C), taux de matière grasse (%).

Point méthodologique : Le type de variable détermine les calculs autorisés et les graphiques utilisables. Par exemple, on ne calcule pas une moyenne sur une variable nominale.

Présentation du tableau 1.2 : Ce tableau classe les quatre grands types de variables statistiques. Il aide l'étudiant à identifier rapidement la nature d'une variable avant de choisir un indicateur ou un graphique.

Tableau 1.2 – Typologie des variables et exemples SNV

Type de variable	**Nature**	**Exemples en zootechnie / épidémiologie**
Qualitative nominale	Catégories sans ordre	Race, sexe, espèce, statut vaccinal
Qualitative ordinale	Catégories ordonnées	Note d'état corporel, score de boiterie
Quantitative discrète	Nombres entiers	Nombre de porcelets, nombre de lactations

Quantitative continue	Nombres réels	Poids, production laitière, température

Présentation de la figure 1.1 : Avant de construire un tableau ou un graphique, il faut identifier le type de variable. Cette figure propose un arbre de décision simple : on part de la question « la variable est-elle qualitative ou quantitative ? », puis on affine.

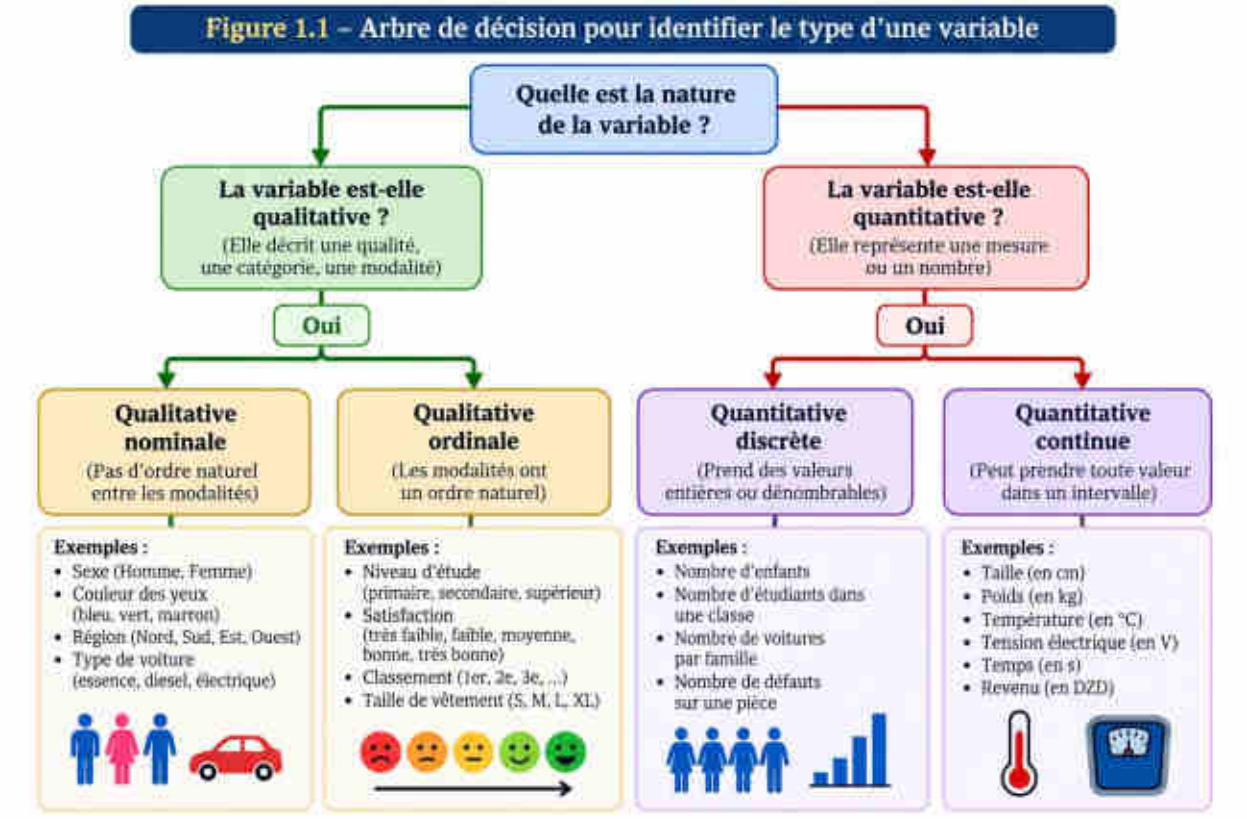


Figure 1.1 – Arbre de décision pour identifier le type d'une variable

Description de la figure 1.1 :

- Si la variable décrit une qualité → qualitative.
  - Si les catégories n'ont pas d'ordre → nominale.
  - Si les catégories ont un ordre → ordinale.
- Si la variable décrit une quantité → quantitative.
  - Si les valeurs sont entières → discrète.
  - Si les valeurs sont mesurées → continue.

Exemple d'application :

- « Race » → qualitative nominale.
- « Note d'état corporel » → qualitative ordinale.
- « Nombre de porcelets » → quantitative discrète.
- « Poids » → quantitative continue.

1.3. Distributions statistiques. Effectifs, fréquences

1.3.1. Principe général

Après la collecte, on organise les données en **distribution statistique**. Pour chaque valeur ou classe, on compte le nombre d'individus : c'est l'**effectif** n_i.

- **Effectif total** : $N=n1+n2+\cdots+n_k$.
- **Fréquence** : $f_i= n_i/N$
- **Pourcentage** : $100\times f_i$.
- **Propriété** : $\sum_{i=1}^{k} f_i = 1$

1.3.2. Cas des variables qualitatives ou discrètes

Pour les variables **qualitatives** (nominales ou ordinales) et les variables **quantitatives discrètes**, le nombre de valeurs possibles est limité et facilement identifiable. On peut donc, sans regroupement en classes, lister toutes les modalités observées, puis compter le nombre d'individus qui prennent chacune de ces modalités. Cette opération conduit à une **distribution statistique d'effectifs**.

Exemple introductif : Dans un élevage de 120 vaches, on étudie la variable « statut mammite » (Oui / Non). Ici, seules deux modalités existent. Il suffit de compter les vaches malades et les vaches saines pour obtenir la distribution.

1.3.3. Cas des variables continues

On regroupe les valeurs en **classes** (intervalles) : $[a0,a1[$, $[a1,a2[$, etc. On compte ensuite les effectifs par classe.

Présentation du tableau 1.3 : Ce tableau montre comment organiser une variable ***qualitative nominale***. Il présente les effectifs, les fréquences et les pourcentages. Il permet de visualiser la race dominante dans l'échantillon.

Tableau 1.3 – Distribution d'une variable qualitative : race de 120 vaches

Race	Effectif *ni*	Fréquence *fi*	Pourcentage (%)
Holstein	54	0,45	45
Montbéliarde	36	0,30	30
Normande	30	0,25	25
Total	**120**	**1,00**	**100**

Interprétation : La race Holstein est la plus représentée (45 %). La fréquence permet de comparer des élevages de tailles différentes.

Figure 1.3 – Organisation des données en distribution statistique

Principe

Après la collecte, on organise les données en **distribution statistique**.
Pour chaque valeur ou classe, on compte le nombre d'individus : c'est l'**effectif** n_i.

Exemple : répartition des vaches selon la race

Modalité (x_i)	Effectif (n_i)	Fréquence ($f_i = n_i/N$)	Pourcentage ($100 \times f_i$)
 Holstein (x_1)	45	0,45	45 %
 Montbéliarde (x_2)	30	0,30	30 %
 Normande (x_3)	25	0,25	25 %
Total	$N = 100$	1,00	100 %

Formules

Effectif total

$$N = n_1 + n_2 + \cdots + n_k$$

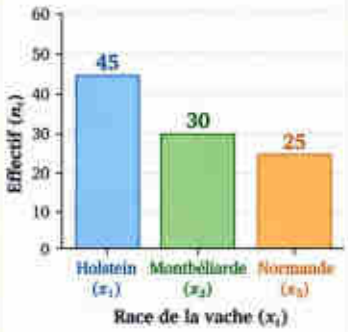
Fréquence

$$f_i = \frac{n_i}{N}$$

Pourcentage

$$\text{Pourcentage}_i = 100 \times f_i$$

Représentation graphique



Propriété

- La somme des effectifs est égale à l'effectif total : $N = \sum_{i=1}^{k} n_i$
- La somme des fréquences est égale à 1 : $\sum_{i=1}^{k} f_i = 1$
- La somme des pourcentages est égale à 100 % : $\sum_{i=1}^{k} 100 \times f_i = 100\ \%$

Présentation du tableau 1.4 : Lorsqu'une ***variable est continue***, on ne peut pas lister toutes les valeurs. On les regroupe donc en classes. Ce tableau présente les classes de poids, les effectifs et les fréquences correspondantes.

Tableau 1.4 – Distribution d'une variable continue : poids de 150 veaux

Classe de poids (kg)	**Effectif *ni***	**Fréquence *fi***
[30 – 35[	9	0,06
[35 – 40[	27	0,18
[40 – 45[	36	0,24
[45 – 50[	45	0,30
[50 – 55[	27	0,18
[55 – 60[	6	0,04
Total	**150**	**1,00**

Interprétation : La classe modale est [45 – 50[kg. La distribution semble relativement centrée autour de 45–50 kg.

Présentation du tableau 1.5 : Ce tableau illustre une variable *quantitative discrète*. Chaque valeur possible est listée avec son effectif et sa fréquence. Il permet de repérer la valeur la plus fréquente.

Tableau 1.5 – Distribution d'une variable discrète : nombre de porcelets par portée

Nombre de porcelets	**Effectif *ni***	**Fréquence *fi***
6	3	0,03
7	8	0,08
8	15	0,15
9	22	0,22
10	28	0,28
11	17	0,17
12	7	0,07
Total	**100**	**1,00**

Interprétation : La valeur modale est 10 porcelets (28 % des portées). La distribution est centrée autour de 9–10.

Présentation du tableau 1.6 : Ce tableau concerne ***une variable qualitative ordinale***. L'ordre des modalités (1 à 5) a une signification. Les effectifs et fréquences permettent de voir si les animaux sont plutôt maigres ou en bon état.

Tableau 1.6 – Distribution d'une variable ordinale : note d'état corporel (NEC) de 80 vaches

Note d'état corporel	**Effectif *ni***	**Fréquence *fi***
1 (très maigre)	4	0,05
2 (maigre)	12	0,15
3 (moyen)	28	0,35
4 (bon)	24	0,30
5 (très bon)	12	0,15
Total	**80**	**1,00**

Interprétation : La majorité des vaches ont une NEC de 3 ou 4. Les extrêmes sont rares.

Présentation du tableau 1.7 : Ce tableau présente ***une variable qualitative nominale*** à deux modalités (oui/non). Il permet de calculer la prévalence de la mammite dans l'échantillon.

Tableau 1.7 – Distribution d'une variable nominale binaire : statut mammite de 120 vaches

Statut mammite	**Effectif *ni***	**Fréquence *fi***	**Pourcentage (%)**
Oui	18	0,15	15

Non	102	0,85	85
Total	**120**	**1,00**	**100**

Interprétation : La prévalence de la mammite est de 15 % dans cet échantillon.

1.4. Représentations graphiques des distributions statistiques

1.4.1. Pourquoi un graphique ?

Un tableau donne des chiffres précis, mais un graphique montre immédiatement la forme de la distribution : symétrique, étalée, avec des valeurs extrêmes, etc. Le choix du graphique dépend du type de variable.

1.4.2. Variables nominales

a) Diagramme en bâtons

Principe : à chaque modalité, on associe un bâton dont la hauteur est proportionnelle à l'effectif ou à la fréquence. Pour une ***variable nominale***, l'ordre des modalités n'a pas d'importance.

Présentation de la figure 1.2 : Cette figure illustre la distribution des races du tableau 1.3. Les bâtons sont disposés sans ordre significatif, car la variable est nominale. La hauteur de chaque bâton représente l'effectif.

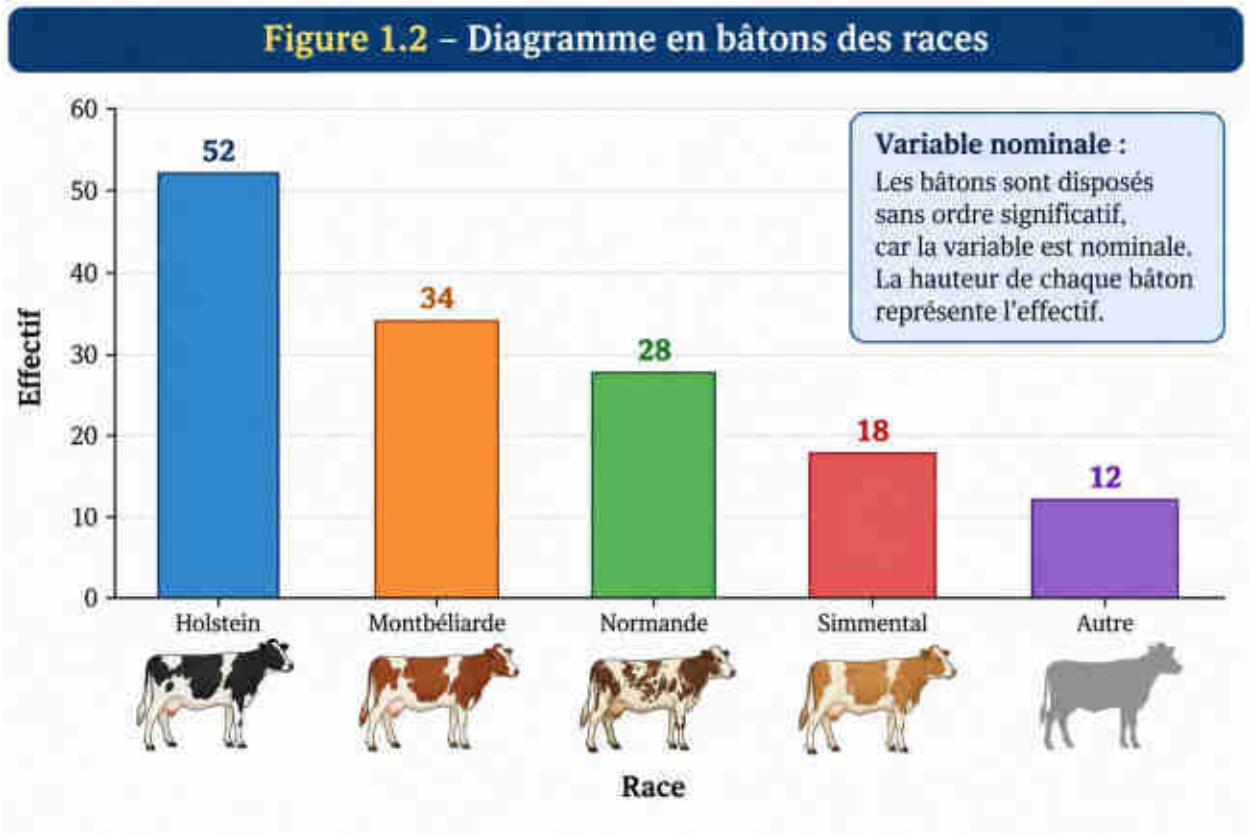


Figure 1.2 – Diagramme en bâtons des races

Description de la figure 1.2 :

- Axe horizontal : races (Holstein, Montbéliarde, Normande).
- Axe vertical : effectifs.
- Bâton Holstein : 54.
- Bâton Montbéliarde : 36.
- Bâton Normande : 30.

Interprétation : Le graphique confirme visuellement la prédominance de la Holstein.

b) Diagramme circulaire

Principe : l'angle de chaque secteur est proportionnel à la fréquence :

$$\alpha = 360° \times f_i$$

Présentation de la figure 1.3 : Ce graphique montre la répartition des causes de mortalité dans un élevage. Chaque ***secteur*** représente une cause. Il est particulièrement adapté aux variables nominales.

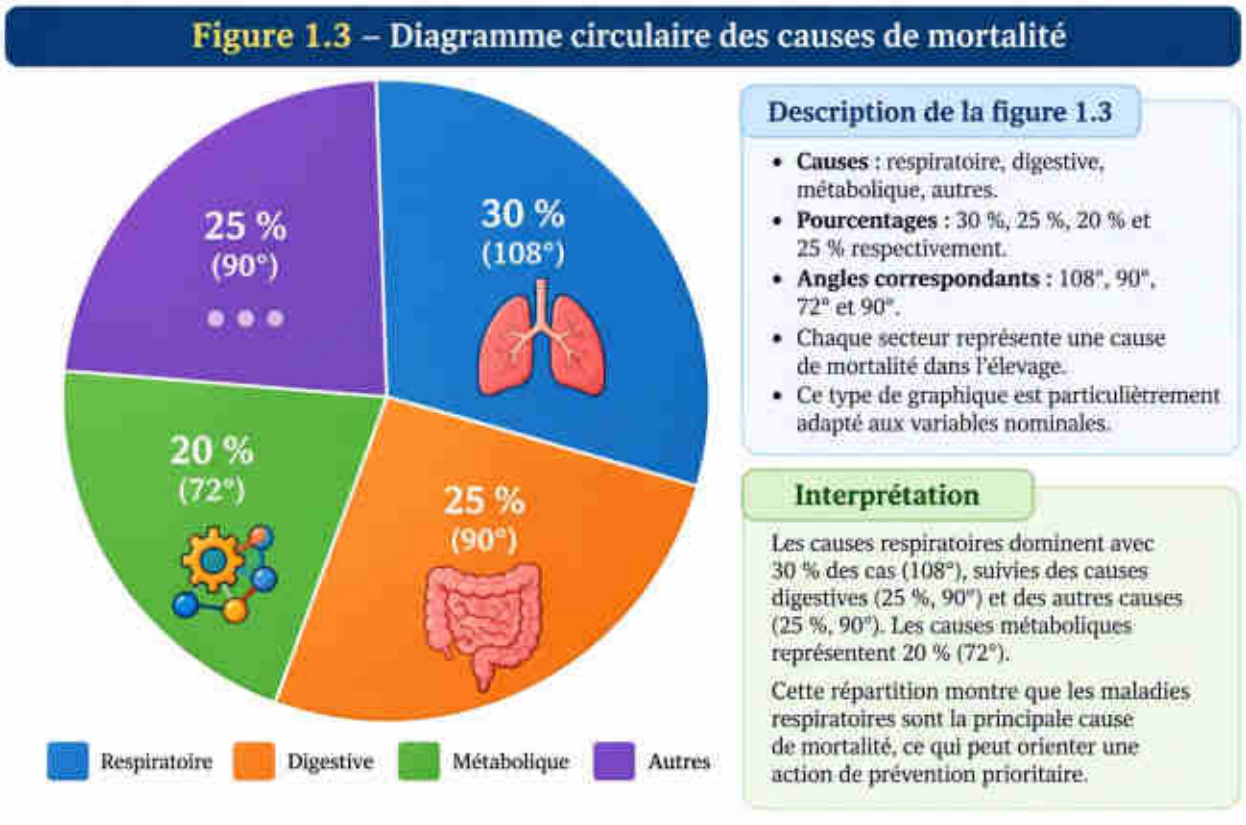


Figure 1.3 – Diagramme circulaire des causes de mortalité

Description de la figure 1.3 :

- Causes : respiratoire (30 %), digestive (25 %), métabolique (20 %), autres (25 %).
- Les angles sont respectivement 108°, 90°, 72° et 90°.

Interprétation : Les causes respiratoires dominent, ce qui peut orienter une action de prévention.

Présentation de la figure 1.4 : Cette figure illustre la répartition des vaches selon leur statut mammite (tableau 1.7). Le secteur « Oui » représente 15 % de l'effectif total.

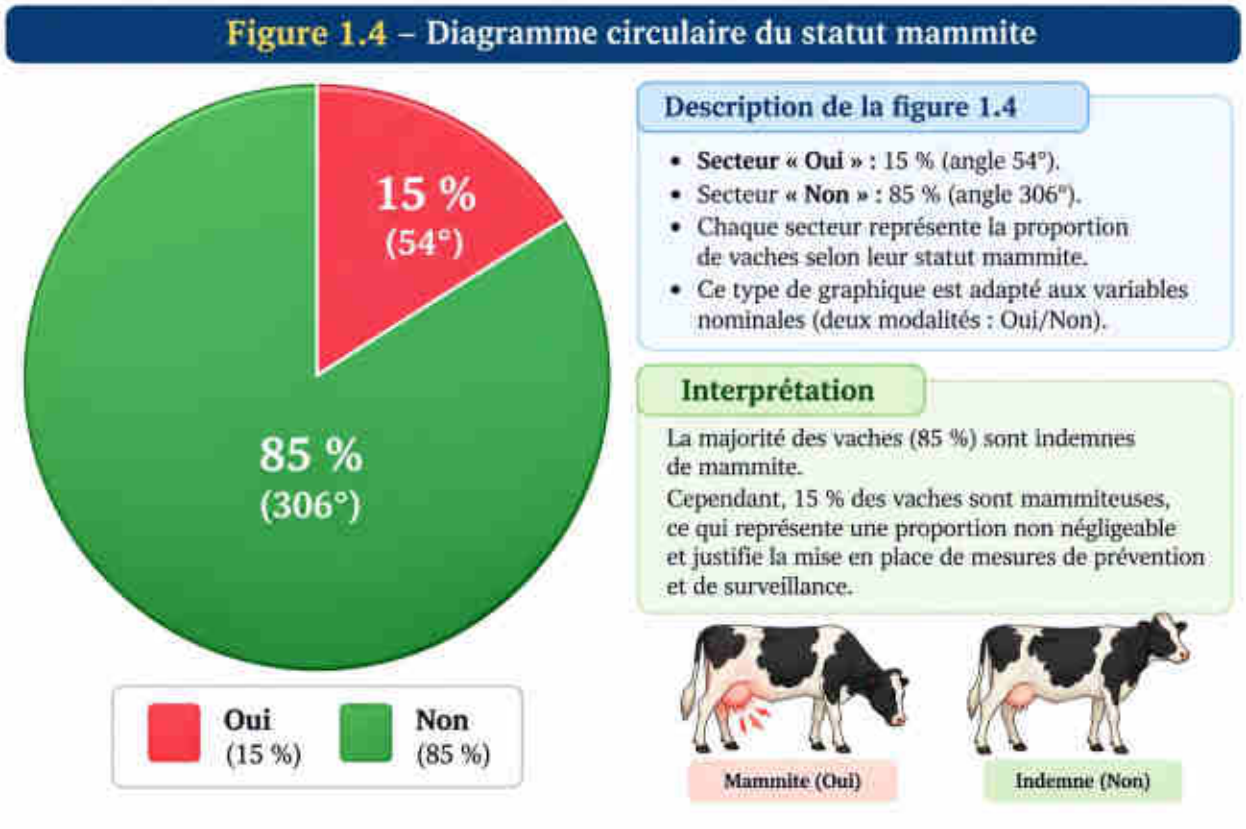


Figure 1.4 – Diagramme circulaire du statut mammite

Description de la figure 1.4 :

- Secteur « Oui » : 15 % (angle 54°).
- Secteur « Non » : 85 % (angle 306°).

Interprétation : La majorité des vaches sont indemnes, mais la proportion de mammiteuses reste non négligeable.

1.4.3. Variables ordinales et discrètes

- **Variables ordinales** : mêmes graphiques que les nominales, mais l'ordre doit être respecté.
- **Variables discrètes** : on privilégie le diagramme en bâtons, car l'écart entre les valeurs a un sens.

Présentation de la figure 1.5 : Cette figure représente **une variable discrète.** L'ordre et l'écart entre les bâtons sont significatifs. On observe la distribution du nombre de porcelets par portée.

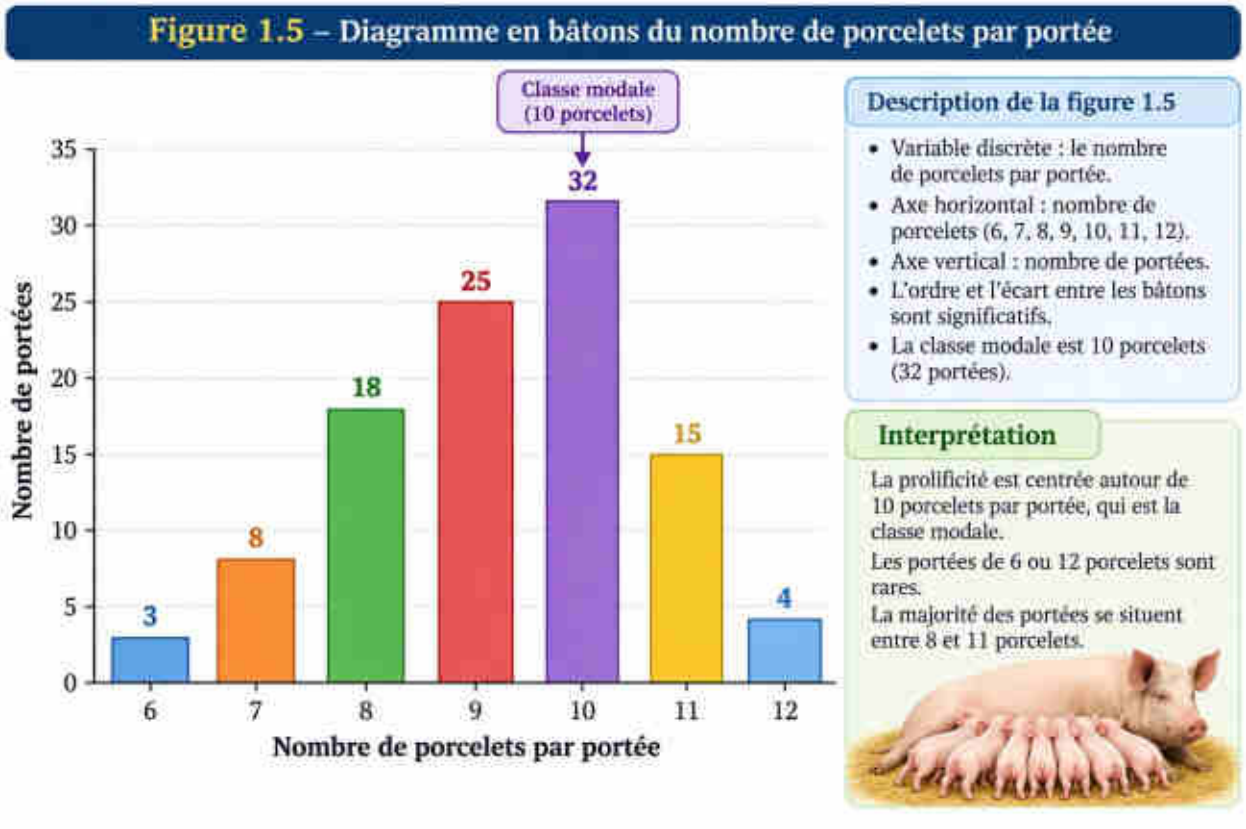


Figure 1.5 – Diagramme en bâtons du nombre de porcelets par portée

Description de la figure 1.5 :

- Axe horizontal : nombre de porcelets (6, 7, 8, 9, 10, 11, 12).
- Axe vertical : nombre de portées.
- La classe modale est 10 porcelets.

Interprétation : La prolificité est centrée autour de 10 porcelets. Les portées de 6 ou 12 sont rares.

Présentation de la figure 1.6 : Cette figure illustre une ***variable ordinale (***un état corporel ***)***. Les barres sont ordonnées de 1 à 5. On visualise la répartition des vaches selon leur état corporel.

Description de la figure 1.6 :

- Axe horizontal : notes de 1 à 5.
- Axe vertical : effectifs.
- La barre la plus haute correspond à la note 3.

Interprétation : La plupart des vaches ont un état corporel moyen à bon.

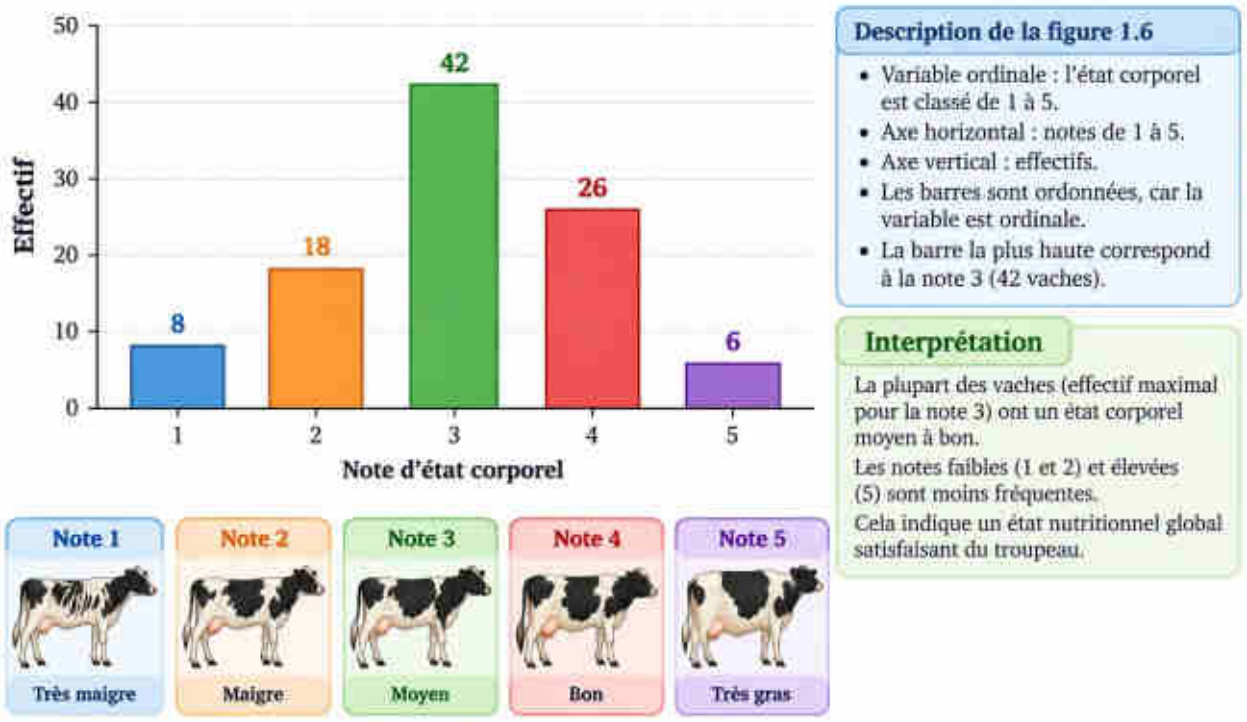


Figure 1.6 – Diagramme en bâtons de la note d'état corporel

1.4.4. Variables continues

a) Histogramme

Principe : la **surface** de chaque rectangle est proportionnelle à l'effectif de la classe. Si les classes ont la même amplitude, la hauteur est proportionnelle à l'effectif. Si les amplitudes diffèrent, on calcule :

$$h_i = \frac{n_i}{amplitude\ de\ la\ classe}$$

Présentation du tableau 1.8 : Ce tableau prépare la construction de l'histogramme. Il indique les classes, les effectifs, les amplitudes et les hauteurs des rectangles. Il est indispensable lorsque les classes n'ont pas la même largeur.

Tableau 1.8 – Effectifs et amplitudes pour l'histogramme des poids

Classe de poids (kg)	**Effectif n*i***	**Amplitude**	**Hauteur *hi***
[20 – 25[	9	5	1,8
[25 – 30[	27	5	5,4
[30 – 35[	36	5	7,2
[35 – 40[	45	5	9,0
[40 – 50[	27	10	2,7
[50 – 60[	6	10	0,6
Total	**150**		

Remarque : Ici, les deux dernières classes ont une amplitude de 10 ; leur hauteur est donc divisée par 10 pour que la surface reste proportionnelle à l'effectif.

Présentation de la figure 1.7 : Cet histogramme représente la distribution des poids de 150 veaux. Chaque rectangle correspond à une classe. La surface du rectangle est proportionnelle à l'effectif. On repère immédiatement la classe modale.

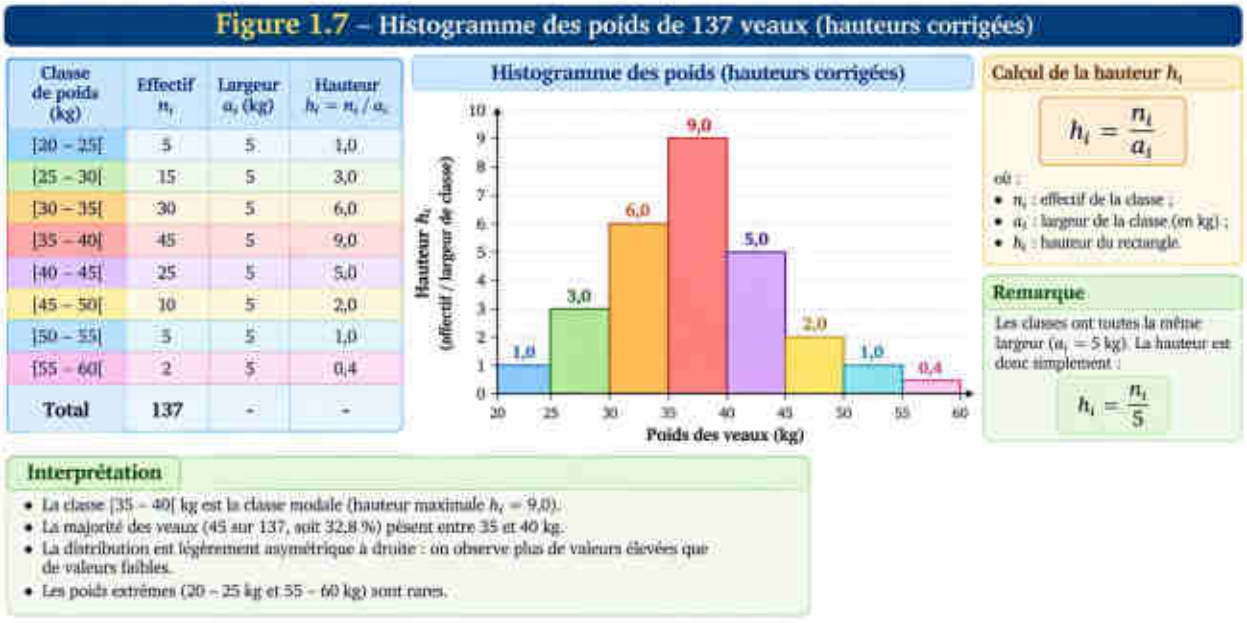


Figure 1.7 – Histogramme des poids

Description de la figure 1.7 :

L'histogramme représente la distribution des poids de 137 veaux répartis en huit classes de largeur égale à 5 kg, allant de ([20,25[) à ([55,60[) kg. L'axe horizontal indique les classes de poids et l'axe vertical les hauteurs (h_i=n_i/a_i), calculées en divisant l'effectif de chaque classe par sa largeur. Le rectangle le plus haut correspond à la classe modale ([35,–,40[) kg, avec un effectif de 45 veaux et une hauteur de 9.

Interprétation : La majorité des veaux pèsent entre 35 et 40 kg. Les poids extrêmes sont rares.

b) Polygone des fréquences

Principe : on joint par des segments de droite les milieux des sommets des rectangles de l'histogramme. Cela donne une courbe de tendance.

Présentation de la figure 1.8 : Ce polygone est construit à partir de l'histogramme précédent. Il permet de visualiser la forme générale de la distribution et de comparer plusieurs distributions entre elles.

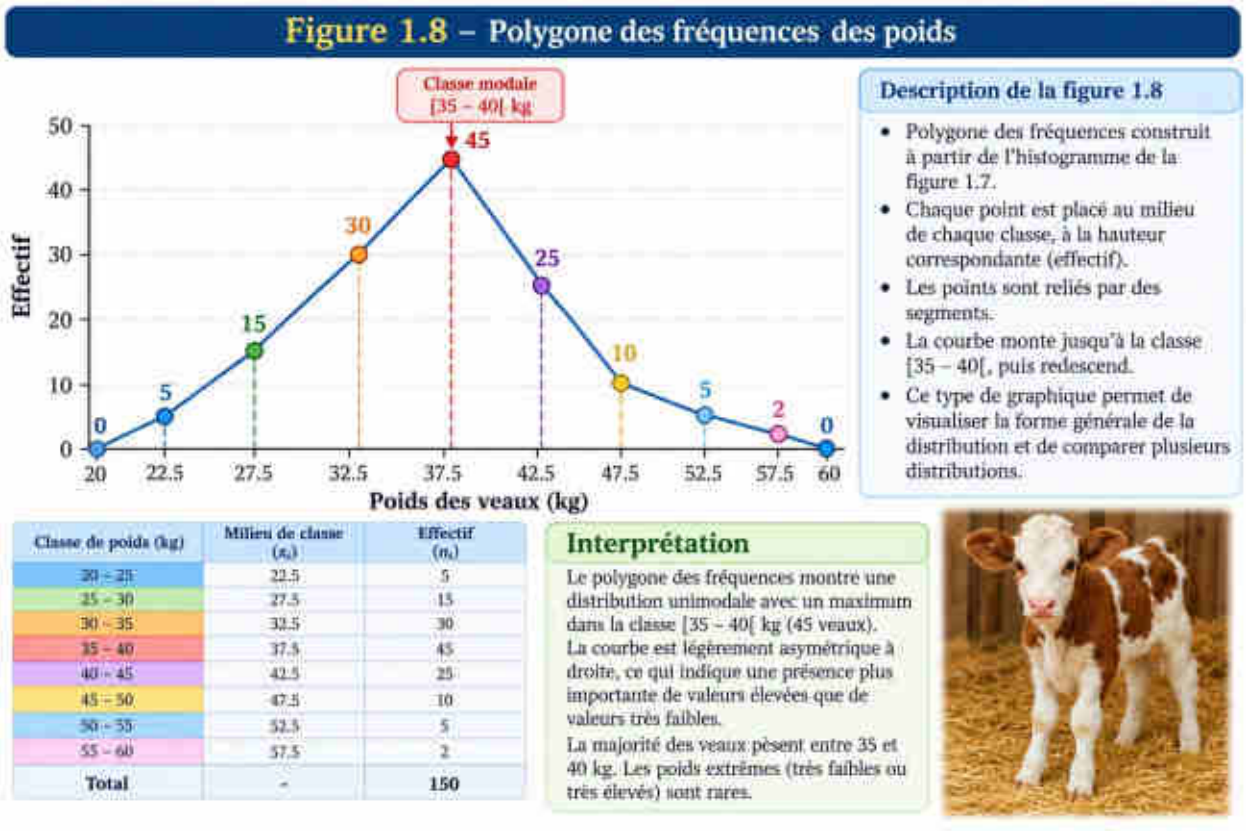


Figure 1.8 – Polygone des fréquences des poids

Description de la figure 1.8 :

- On place un point au milieu de chaque classe, à la hauteur correspondante.
- On relie les points par des segments.
- La courbe monte jusqu'à la classe [35 – 40[, puis redescend.

Interprétation : La distribution est unimodale et légèrement étalée vers les poids élevés.

1.5. Synthèse du chapitre

- La statistique descriptive commence par une **définition claire** de la population, de l'échantillon et de la variable.
- Il existe **quatre types de variables** : nominale, ordinale, discrète, continue.
- Une distribution se résume par un **tableau d'effectifs et de fréquences**.
- Le choix du graphique dépend du type de variable :
    - Nominale → bâtons ou secteurs.
    - Ordinale → bâtons ordonnés.
    - Discrète → bâtons.
    - Continue → histogramme, polygone.
- En zootechnie et en épidémiologie, ces outils permettent de **décrire** une population animale ou un phénomène sanitaire avant toute analyse inférentielle.

1.6. Conclusion

Ce premier chapitre a posé les bases indispensables à toute analyse de données en zootechnie et en épidémiologie. Nous avons appris à définir précisément une population, un échantillon et une variable, à distinguer les quatre grands types de variables, et à organiser les données sous forme de tableaux d'effectifs et de fréquences. Nous avons également vu comment choisir la représentation graphique la plus appropriée selon la nature de la variable : diagramme en bâtons ou circulaire pour les variables qualitatives, diagramme en bâtons ordonné pour les variables ordinales, histogramme et polygone pour les variables continues.

Ces outils permettent déjà de décrire une situation concrète : prévalence d'une maladie, distribution des poids, répartition des races dans un élevage. Mais la description ne s'arrête pas là. Pour comparer des groupes, résumer une distribution par quelques chiffres clés ou étudier la relation entre deux caractères, il est nécessaire de faire appel à des indicateurs quantitatifs. C'est l'objet du chapitre suivant, consacré aux caractéristiques de tendance centrale et de dispersion.

? Questions de révision

1. Définissez : individu, population, échantillon, variable.
2. Classez les variables suivantes : race, poids, nombre de lactations, score de boiterie.
3. Pourquoi ne peut-on pas calculer une moyenne sur une variable nominale ?
4. Construisez un tableau d'effectifs et de fréquences pour la variable « statut mammite » (Oui = 18, Non = 82).
5. Quel graphique choisiriez-vous pour représenter la distribution des poids de 200 bovins ? Justifiez.
6. Interprétez la classe modale dans un histogramme.
7. Construisez un diagramme en bâtons pour la variable « nombre de porcelets » à partir du tableau 1.5.
8. Quelle est la différence entre un histogramme et un diagramme en bâtons ?